Finding spectral features for the early identification of biotic stress in plants

Till Rumpf

Institut für Geodäsie und Geoinformation

Till Rumpf

Finding spectral features for the early identification of biotic stress in plants

Inaugural-Dissertation
zur Erlangung des Grades Doktor-Ingenieur (Dr.-Ing.)
der Landwirtschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn
von Till Rumpf aus Münster in Westfalen

vorgelegt am 02.08.2012

Referent	Prof. Dr. rer. nat. Lutz Plümer Institut für Geodäsie und Geoinformation, Universität Bonn
Korreferenten	PD Dr. Erich-Christian Oerke Institut für Nutzpflanzenwissenschaften und Ressourcenschutz, Universität Bonn
	Prof. Dr. Björn Waske Institut für Geodäsie und Geoinformation, Universität Bonn

Tag der mündlichen Prüfung: 23.11.2012 Erscheinungsjahr: 2013
Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn `http://hss.ulb.uni-bonn.de/diss_online` elektronisch publiziert.

Für Nadine

Kurzfassung

Früherkennung von biotischem Pflanzenstress ist für den Präzisionspflanzenschutz wesentlich, aber schwierig zu erreichen. Die Vorhersage von Pflanzenkrankheiten und Unkräutern in einem frühen Entwicklungsstadium hat signifikanten Einfluss auf das Ausmaß und die Effektivität einer Pflanzenschutzmaßnahme. Aufgrund der Abhängigkeit einer Maßnahme von der Art der Pflanzenkrankheit oder des Unkrauts und ihrer ökonomischer Schadschwelle ist eine präzise Identifizierung der Schadursache essentiell, aber gerade im Frühstadium durch die Ähnlichkeit der Schadbilder problematisch.
Nicht-invasive optische Sensoren mit hoher Auflösung sind vielversprechend für eine Früherkennung von biotischem Pflanzenstress. Daten dieser Sensoren, beispielsweise Hyperspektral- oder Fluoreszenzspektren, enthalten relevante Informationen über das Auftreten von Pathogenen; Formparameter, abgeleitet aus bispektralen Bildern, zeigen großes Potential für die Früherkennung von Unkräutern in Kulturpflanzen.
Die Analyse dieser hochdimensionalen Sensordaten unter Berücksichtigung vielfältiger Faktoren ist eine anspruchsvolle Herausforderung. Moderne Methoden des maschinellen Lernens bieten hier zielführende Möglichkeiten. Während die traditionelle Statistik die a-posteriori Wahrscheinlichkeit der Klasse basierend auf Wahrscheinlichkeitsverteilungen schätzt, verwenden maschinelle Lernverfahren Algorithmen für eine Optimierung der Vorhersagegenauigkeit auf Basis diskriminierender Funktionen. Grundlage zur Bearbeitung dieser nicht-linearen Klassifikationsprobleme sind robuste maschinelle Lernverfahren.
Die vorliegende Dissertationsschrift zeigt, dass die Integration moderner Sensortechnik mit fortgeschrittenen Methoden des maschinellen Lernens eine Erkennung und Differenzierung von Pflanzenkrankheiten und Unkräutern ermöglicht. Einen wesentlichen Beitrag für eine effektive und robuste Klassifikation leisten Support Vektor Maschinen (SVMs) mit nicht-linearen Kernels. Weiterhin wird gezeigt, dass SVMs auf Basis spektraler Vegetationsindizes die Detektion von Pflanzenkrankheiten noch vor Auftreten visuell wahrnehmbarer Symptome ermöglichen. Dies wurde mit bekannten Verfahren noch nicht erreicht.
Zur Identifikation krankheitsspezifischer Merkmale aus den zugrunde liegenden originären hochdimensionalen Sensordaten wurden Merkmale konstruiert und selektiert. Die Selektion ist sowohl vom Klassifikationsproblem als auch von den Eigenschaften der Merkmale abhängig. Im Fall von Fluoreszenzspektren war eine Extraktion von neuen Merkmalen notwendig. In diesem Zusammenhang trägt die Modellierung des Signalrauschens durch eine analytische Beschreibung der spektralen Signatur zur deutlichen Verbesserung der Klassifikationsgenauigkeit bei. Im Fall der Differenzierung von unterschiedlichen Unkräutern erhöht die Ausnutzung der Hierarchie der Unkrautarten die Genauigkeit signifikant.
Diese Arbeit zeigt das Potential von Support Vektor Maschinen, Merkmalskonstruktion und Selektion für den Präzisionspflanzenschutz. Eine problemspezifische Extraktion und Selektion relevanter Merkmale in Verbindung mit sachbezogenen Klassifikationsmethoden ermöglichen eine robuste Identifikation von Pathogenen und Unkräutern zu einem sehr frühen Zeitpunkt.

Abstract

Early detection of biotic stress in plants is vital for precision crop protection, but hard to achieve. Prediction of plant diseases or weeds at an early stage has significant influence on the extent and effectiveness of crop protection measures. The precise measure depends on specific weeds and plant diseases and their economic thresholds. Weeds and plant diseases at an early stage, however, are difficult to identify. Non-invasive optical sensors with high resolution are promising for early detection of biotic stress. The data of these sensors, e.g. hyperspectral or fluorescence signatures, contain relevant information about the occurrence of pathogens. Shape parameters, derived from bispectral images, have enormous potential for an early identification of weeds in crops.

The analysis of this high dimensional data for an identification of weeds and pathogens as early as possible is demanding as the sensor signal is affected by many influencing factors. Nevertheless, advanced methods of machine learning facilitate the interpretation of these signals. Whereas traditional statistics estimate the posterior probability of the class by probability distribution, machine learning methods provide algorithms for optimising prediction accuracy by the discriminant function. Machine learning methods with robust training algorithms play a key role in handling non-linear classification problems.

This thesis presents an approach which integrates modern sensor techniques and advanced machine learning methods for an early detection and differentiation of plant diseases and weeds. Support vector machines (SVMs) equipped with non-linear kernels prove as effective and robust classifiers. Furthermore, it is shown that even a presymptomatic identification based on the combination of spectral vegetation indices is realised. Using well-established data analysis methods of this scientific field, this has not achieved so far.

Identifying disease specific features from the underlying original high dimensional sensor data selection is conducted. The high dimensionality of data affords a careful selection of relevant and non-redundant features depending on classification problem and feature properties. In the case of fluorescence signatures an extraction of new features is necessary. In this context modelling of signal noise by an analytical description of the spectral signature improves the accuracy of classification substantially. In the case of weed discrimination accuracy is improved by exploiting the hierarchy of weed species.

This thesis outlines the potential of SVMs, feature construction and feature selection for precision crop protection. A problem-specific extraction and selection of relevant features, in combination with task-oriented classification methods, is essential for robust identification of pathogens and weeds as early as possible.

Contents

1 Introduction

Production of food for a growing population and substitution of fossil fuels by renewable energy sources pose the most demanding challenges for modern agriculture. As expected by the Food and Agriculture Organisation of the United Nations (FAO), agricultural output needs to increase by 70% until 2050 (FAO, 2009). Limitation of arable land demands a sustainable intensification, i.e. a production of more food from the same area of land while reducing the environmental impacts. Precision agriculture strategies, in terms of an integration of modern technologies like sensors and information management systems, have the potential to increase the yield of crops (Gebbers and Adamchuk, 2010).
The yield of crops is affected by different plant stresses. Plant stress has been defined as impairment that influences plant growth, productivity and reproductive capacity in a negative way (Gaspar et al., 2002). During the vegetation period crop plants are exposed to different kinds of stress. Most *biotic* stress factors like plant pathogens and weeds as well as *abiotic* stress factors like water deficiency, nutrient deficiency and temperature affect the photosynthetic apparatus observably (Carter and Knapp, 2001; Stafford, 2000).

Figure 1.1: Crop stand of *Hordeum vulgare* and the weed *Galium aparine* at a later stage. In this study *Galium aparine* was of high relevance caused by a low economic threshold and high similarity to other weeds.

Focusing on biotic stress, quantitative and qualitative losses in crop production are caused by a broad spectrum of pathogens and weeds (Oerke and Dehne, 2004). Weeds compete with crop plants for the environmental resources such as light, water, nutrients and space (Figure 1.1) and serve as hosts for pests and diseases (Patterson, 1995). Plant pathogen interactions and the resulting disease symptoms are influenced by various external factors and variable in physiological and morphological changes (Figure 1.2). Before characteristic symptoms are visible, several putative modifications in cellular leaf structure occur, for example changes in water content at infection sites, initiating cell death caused by fungal toxins or resistance reactions of plant tissue (Jones and Dangl, 2006). In order to achieve high yields in agricultural crop systems, the control of biotic stress is of high relevance.

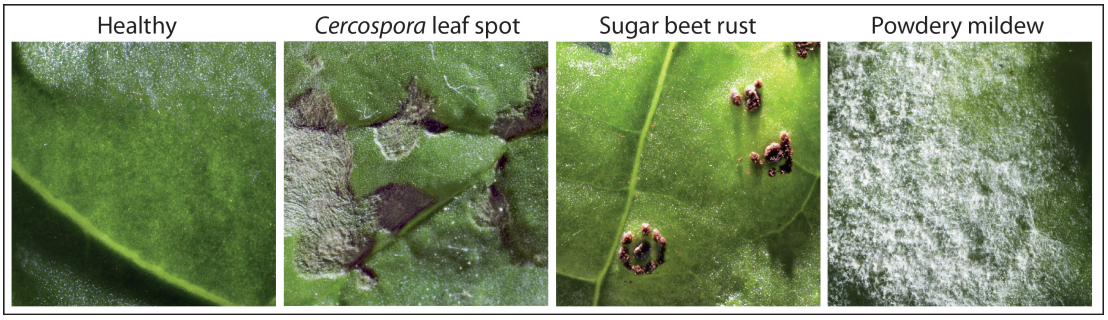


Figure 1.2: Healthy sugar beet leaf and leaves with characteristic symptoms of *Cercospora* leaf spot, sugar beet rust and powdery mildew. These foliar pathogens are exemplary used to prove the potential for disease detection of machine learning methods.

Therefore yield limiting factors have to be detected as early as possible in order to apply appropriate counter measures.

A better understanding of these complex processes can be achieved by the use of non-invasive sensors (Finkel, 2009). Optical sensors and imaging techniques like hyperspectral sensing and chlorophyll fluorescence have demonstrated high potential for an early detection and monitoring of plant diseases or weed populations (Gerhards and Sökefeld, 2003; Mahlein et al., 2012b; West et al., 2003). Further progress in sensor technology offers new opportunities for precision agriculture. Innovative sensor systems can provide detailed and highly resolved information on agricultural crop stands; an overview is given in Mahlein et al. (2012a). An identification of pathogens and weeds as early as possible necessitates the integration of sensor data of high spectral, spatial and temporal resolution with advanced methods of data analysis.

This challenge is further increased by many factors which affect signal acquisition. Leaf orientation, spatial arrangements of the elements involved in the interaction between radiation and vegetation, biologic variability between plants, different illumination and changing environmental conditions influence sensor signals (Jensen, 2007).
Changes in the reflectance signal caused by early and presymptomatic stress symptoms are often subtle, especially in regard to the biologic variability. To overcome this problem, several assumptions have to be taken. Measurement conditions have to be modelled and removed from sensor signals, otherwise the conditions have to be constant in order to be neglectable. The remaining signal has to underlie proper model assumptions which include all target variables. Most approaches for data analysis in precision crop protection are based on models which assume a linear correlation between observed data and target variables. In order to analyse these linear models statistically, the underlying distribution of observations has to be known. The unmodelled part of observation, denoted as noise, is usually assumed to follow a multivariate Gaussian distribution which allows to use least squares methods for optimisation of the model parameters. Linear models have been used both for regression and classification. In the case of regression, the target variable is continuous, whereas classification assumes discrete class labels. If the involved classes happen to be linearly

separable, linear models are appropriate.
Unfortunately, in many relevant cases the assumption of linearity is violated. To circumvent the resulting problem, non-linear wavelength combinations were developed. In turn these combinations, however, can be used in order to detect correlation of spectral characteristics to plant parameters by linear regression (Carter and Knapp, 2001; Delalieux et al., 2009). Since nearly 40 years task-specific combinations are published as spectral vegetation indices (VIs). Several research groups analysed spectral data using VIs with quantitative correlation to biophysical and biochemical traits (Gitelson et al., 2002). Thus, VIs are feasible to differentiate healthy and diseased plants (Delalieux et al., 2009; Steddom et al., 2005). Mahlein et al. (2010) found out that the correlation of VIs to plant diseases depends on the kind of disease and disease severity. As a rule, we found that a single VI is not sufficient for the identification and differentiation of diseases. Many researchers were able to deduce changes in plant health using VIs, but a specification of the individual disease using VIs was not feasible so far. Moreover, a desirable presymptomatic identification of plant diseases could not be realised.

This thesis has been conducted within the Research Training Group 722 'Information Techniques for Precision Crop Protection', funded by the German Research Foundation (DFG). The results benefit from the interdisciplinary collaboration with the Institute of Crop Science and Resource Conservation (INRES – Phytomedicine) of Bonn University and the Institute of Phytomedicine, Department of Weed Science, of Hohenheim University.
In this thesis, an early detection and differentiation of plant diseases in sugar beets is considered. For the first time this is facilitated based on the combination of nine spectral vegetation indices (VIs) by using machine learning methods. In comparison to data analysis methods, well-established in this scientific field, Support Vector Machines (SVMs) are proved to be superior. The identification of specific diseases, namely *Cercospora* leaf spot, sugar beet rust and powdery mildew, is realised before symptoms become visible. Furthermore, an early detection and differentiation of weeds based on image series from different vegetation periods is achieved with high accuracy.

This thesis is structured as follows:

- In Chapter 2 an introduction in classification for precision crop protection is given. After a comparison of generative and discriminative models in statistical modelling, a general description of machine learning, especially Support Vector Machines, is presented.

- Chapter 3 'Early Detection of Biotic Stress Using SVMs' describes the results of the appended papers using machine learning methods. All these publications are based on the interdisciplinary collaboration within the Research Training Group 722.

- Chapter 4 summarises the main results of this thesis and gives an outlook.

Substantial parts of this work are integrated into the article 'Advanced machine learning methods for early detection of biotic stress in precision crop protection' submitted to the journal Precision Agriculture[1]. Chapter 1 contains a modified section of data evaluation and Chapter 2 is strongly extended in comparison to the article in Precision Agriculture. Chapter 3 and 4 are similar related to the submitted article.

[1] Rumpf, T., Römer, C., Mahlein, A.-K., Behmann, J., Plümer,L.,2012. Advanced machine learning methods for early detection of biotic stress in precision crop protection. Precision Agriculture (under Review)

2 Classification for Precision Crop Protection

Linear models optimised by least squares methods are able to classify multi-dimensional data under the assumption of a Gaussian distribution. In this case the results are statistically optimal. If the observations do not follow the underlying assumptions the derived models and their parameters are not optimal and inappropriate in many cases.
In this context Breiman (2001) has introduced the metaphor of two cultures to draw conclusions from data. One culture uses generative models, based on traditional statistics, with the assumptions of linearity and Gaussian distribution in most cases, leading to probabilities. The other culture derives and optimises discriminative functions, which are often used in machine learning and lead to predictions in form of class labels. Machine learning provides more flexible discriminative models with the promise to improve the detection and diagnosis of plant diseases and weeds (Sajda, 2006). These methods play a key role in handling complex non-linear classification problems.

2.1 Generative vs. discriminative models

Statistical modelling can generally be divided into generative and discriminative models. Generative models are full probabilistic models of all variables, whereas discriminative models only provide a model for the target variables depending on the observations. Both generative and discriminative models determine the maximal posterior probability $p(\mathbf{y}|\mathbf{x})$ given the observations $\mathbf{x}$ in order to assign one of the classes y to each new $\mathbf{x}$. This classification problem can be broken down into two stages, the *inference stage* in which observations $\mathbf{x}$ are used to learn a model for $p(\mathbf{y}|\mathbf{x})$, and the subsequent *decision* stage to make optimal class assignments based on these posterior probabilities (Bishop, 2006).
Generative models arise from assumptions about the distribution of the data. The generative models assign a joint probability to paired observation and classes $p(\mathbf{y}, \mathbf{x})$, which involves implicit modelling of the data generated by a stochastic data model $p(\mathbf{x})$ (Kumar and Hebert, 2003). Thus, a generative model can be used to generate values of any variable in the model. This model represents the distribution of the observations resulting in the demand of exhaustive data amount or in simplifying model assumptions. Errors in these model assumptions are causing unrealistic results if taken wrong. By limited knowledge of influencing factors or the generation process the estimation of $p(\mathbf{x})$ is very problematic. Generative models estimate likelihoods $p(\mathbf{x}|\mathbf{y})$ (Figure 2.1)) and priors $p(\mathbf{y})$, but connected over the Bayes theorem $p(\mathbf{y}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{y})p(\mathbf{y})}{p(\mathbf{x})}$ it becomes obvious that $p(\mathbf{x})$ is implicitly included,

$$p(\mathbf{y}, \mathbf{x}) = p(\mathbf{x}|\mathbf{y})\, p(\mathbf{y}) = p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})\,. \tag{2.1}$$

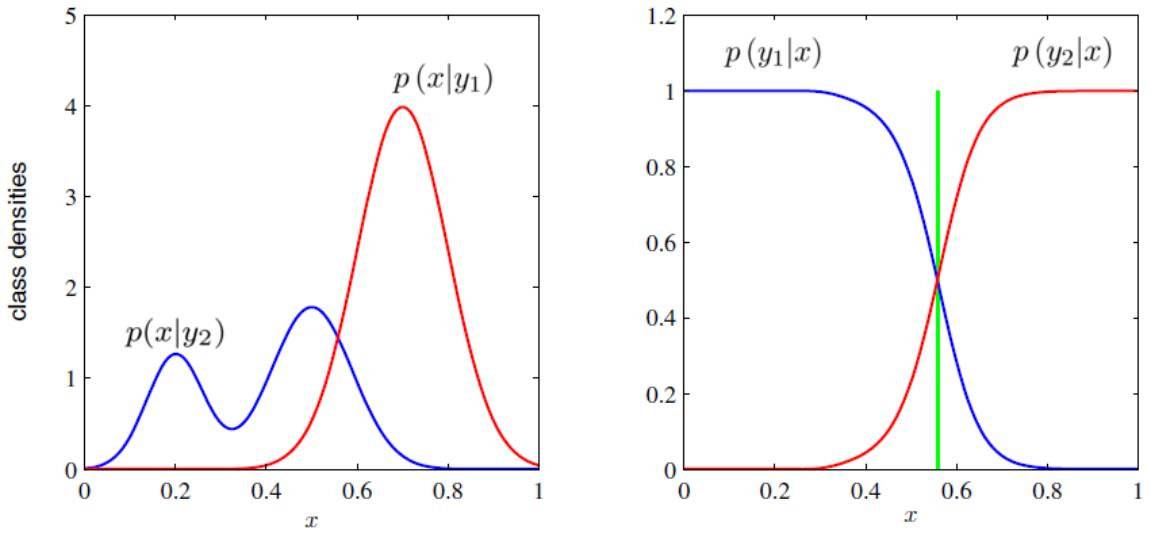


Figure 2.1: Example of the likelihood densities from generative models for two classes having a single input variable x (left plot) together with the corresponding discriminative models figured as posterior probabilities (right plot). The vertical green line in the right plot shows the decision boundary in x that gives the minimum misclassification rate (Bishop, 2006, modified).

This approach explicitly models the distribution of inputs as well as outputs.

Discriminative models, however, directly focus on the class posterior $p(\mathbf{y}|\mathbf{x})$ without explicitly modelling the marginal $p(\mathbf{x})$ (Lafferty, 2001). Therefore the discriminative approach[1] defines only the first term on the right side of equation (2.1) by a learning algorithm (Rubinstein and Hastie, 1997). It is much easier to assign class labels to observations based on maximal posterior $p(\mathbf{y}|\mathbf{x})$ than to model the distribution $p(\mathbf{x})$ which has generated the available observations $\mathbf{x}$. Thus, a discriminative model allows only sampling of the target variables conditional on the observed quantities.

Comparing these two approaches, generative models are most demanding because they involve finding the joint distribution over both $\mathbf{x}$ and the classes $\mathbf{y}$. For many applications, $\mathbf{x}$ will have high dimensionality, and consequently a large set of observations is needed in order to determine the likelihood with reasonable accuracy. One advantage, however, is that a qualitative statement for new observations is given based on the marginal density of data $p(\mathbf{x})$. Thus, the classical approach inappropriately uses a generative joint model when in fact only the posterior probabilities are needed (McCallum et al., 2000). Certainly, if only a conditional problem (in which the observations are given) has to be solved, discriminative models without the need to explicitly model the distribution of the underlying observations are faster, more robust and more accurate. When limited information about the underlying data model exists, which is mostly the case in precision crop protection, this simplification

[1] Some people distinguish between conditional and discriminative models. For sake of simplicity, these two terms are used interchangeably.

is definitely beneficial. Actually, in this case, we do not even require the true posterior as long as we can use the training data to find a discriminative function that maps each $\mathbf{x}$ directly onto a class label. Thereby the inference and decision stages are combined into a single learning problem. In the example of Figure 2.1, this corresponds to finding the value of x shown by the vertical green line, because this is the decision boundary giving the minimum probability of misclassification. Due to the unknown data distribution the validation of the derived models is conducted by estimating the unpredicted class values. As a rule, a hold-out test data set is used for an unbiased accuracy estimation of the current model.

2.2 Machine learning

Machine learning, a sub-discipline of artificial intelligence, is concerned with automatically learning regular patterns from data. In the case of a dichotomous classification task the training data consists of observations $\mathbf{x} \in \mathbb{R}^n$ and labels $\mathbf{y} \in \{+1, -1\}$. The aim is to conclude from training data to new unlabelled data, i.e. to classify the unlabelled observations. Since only a subset of all possible inputs is given by training data, the learner must be able to generalise. The result is a discriminative function which can be used to interpret new data.

The main aim of using machine learning methods in precision crop protection is to detect variability and heterogeneity within crop stands caused by biotic stresses like diseases or weeds. Great influence on data from all different scales is given by biologic variability of plants which represents a major challenge in early detection of biotic stress by using machine learning methods. In precision agriculture, machine learning methods and pattern recognition are in the early stages of development (Mucherino et al., 2009; Huang et al., 2010). Recently, these methods have also found attention in precision crop protection and plant sciences, promising to cope with the challenging boundary conditions of early detection of biotic stress.
Different research groups applied machine learning to precision crop protection applications. De Wolf and Francl (2000) used neural network classification for a forecasting of infection periods of tan spot and *Stagonospora* blotch in wheat. A prediction of favorable conditions was realized based on environmental features as temperature, dew period, relative humidity and precipitation. One of the first approaches to detect or to differentiate biotic stress directly by using machine learning was realised by Moshou et al. (2004). They automatically detected yellow rust in wheat on the basis of reflectance measurements using neural networks. Wang et al. (2008) used artificial neural networks (ANNs) to predict *Phytophthora infestans* infections on tomato plants by relevant regions of the hyperspectral signatures. Wu et al. (2008) have recently shown that an early detection of *Botryis cinerea* on eggplant leaves is possible applying back-propagating neural networks and principle component analysis (PCA) to hyperspectral signatures. For weed species discrimination Burks et al. (2005) evaluated different neural-network classifiers.
In the last years, especially in genetics or remote sensing, support vector machines (SVMs) have proven to be very effective. Golub et al. (1999) successfully used unsupervised and supervised learning methods for cancer classification. Gene selection for cancer classification

using SVMs was evaluated by Brown et al. (2000), Furey et al. (2000) and Guyon et al. (2002). Biological and bioinformatics applications of SVMs have been reviewed in Byvatov and Schneider (2003). Melgani and Bruzzone (2004) showed a high potential of SVMs for classifying remote sensing data. In Mountrakis et al. (2011) remote sensing applications of SVMs, as promising machine learning method, are reviewed. Furthermore, remote sensing has been widely explored as a possibility for detection and mapping of weeds and plant diseases in agricultural crops (Lamb and Brown, 2001; Mewes et al., 2011; Thorp and Tian, 2004).
Compared to remote sensing in precision agriculture differences in the observation scale exist. The measuring scale ranges from contact measurement over canopy to field measurements. In precision crop protection, however, the conditions are comparative to remote sensing: the mis-balance between the number of observations and feature dimension or the signal noise are common. Karimi et al. (2006) applied support vector machines for weed and nitrogen stress detection in corn. The application of SVMs in the field of precision crop protection seems suitable provided by the good generalisation ability of SVMs which is closely examined below.

The most common discrimination function is linear and can be defined as

$$y_j = \text{sign}\left(\mathbf{w}^T\mathbf{x}_j + \omega_0\right), \tag{2.2}$$

assigning the class label y_j to the vector valued sample $\mathbf{x}_j$ (Schölkopf and Smola, 2002). This equation can be seen as a definition of a hyperplane that maps from the data space $\mathbb{D}^n$ to two classes y. An observed sample $\mathbf{x}_j$ is mapped into the two classes via the weight vector $\mathbf{w}$ and bias ω_0. The task is to learn $\mathbf{w}$ and ω_0, using supervised methods, which result in maximising prediction performance conforming to a good generalisation ability.

Specifically, SVMs have outstanding generalisation ability by maximising the margin around the hyperplane using structural risk minimisation (SRM) (Boser, 1992; Vapnik, 2000). Instead of the empirical risk minimization which minimizes the empirical model errors – in other words finds the best fitted model for a given dataset – SRM minimizes the error in prediction of unseen test data provided by Vapnik Chervonenkis theory (VC theory) (Vapnik, 2000). The VC theory can be seen as implementation of Occam's Razor principle which selects from among competing solutions the one making the fewest assumptions and thereby offers the simplest model. As a consequence SRM punishes the overfitting to training data. Accordingly, SRM addressed the vast excess in measured features (in our case shape parameters or hyperspectral data) over the number of samples, known as the 'large-p, small-n' problem. A support vector machine (SVM) is a linear discriminant that separates data into classes using a hyperplane with maximum margin. In the case of a small set of training samples this property is particularly important. The model of the SVM is defined by the training samples nearest to the decision boundary, called support vectors (SVs) $\mathbf{x}_i$ and the attached Lagrange multipliers α_i combined to $\mathbf{w}^T = \sum_i \alpha_i y_i \mathbf{x}_i$ and the bias ω_0. As described thus far, the SVM assumes linearly separable data. Cortes and Vapnik (1995) considered the case which allows some misclassification by defining a 'soft margin'. The 'soft margin' classification finds a hyperplane that splits the training

data as best as possible while maximising the distance to the nearest cleanly split examples.

Formally, the demand for a maximal margin defines an optimisation problem under constraints where the training data only appears as a scalar product $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$. Conducting now a non-linear separation using SVMs, a non-linear transformation into high-dimensional space is necessary. This requires the computation of scalar products in high-dimensional space. These expensive calculations are reduced significantly by using the 'kernel trick' (Schölkopf and Smola, 2002). The scalar product can be replaced by a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ in the input space which enables the computation of affinity in high-dimensional space by using the untransformed feature vector. Based on the 'kernel trick' an explicit transformation is not needed. The non-linear discrimination function is now defined as (Schölkopf and Smola, 2002)

$$y_j = \operatorname{sign}\left(\sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}_j) + \omega_0\right). \tag{2.3}$$

The most common non-linear kernel function is the rbf-kernel which uses radial basis functions as affinity measure in order to enable non-linear discrimination. The parameter σ determines the width of the Gaussian kernel with the following calculation formula (Schölkopf and Smola, 2002)

$$k_{\text{rbf}}(\mathbf{x}_i, \mathbf{x}_j) = \exp -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}. \tag{2.4}$$

These properties, viz. excellent generalisation performance, building sparse models and the possibility to discriminate with non-linear decision boundaries, characterise SVMs as suitable for the analysis of high dimensional data for precision crop protection. The application of SVMs to the early detection of weeds and plant diseases is discussed in the following examples.

3 Early Detection of Biotic Stress Using SVMs

In this section an approach which integrates modern sensor techniques and advanced machine learning methods is presented. The potential and the challenges of this interdisciplinary approach are discussed. It turns out that using non-invasive observations, in combination with non-linear SVMs, detect and classify plant diseases and weeds at a very early stage. In the case of plant diseases even a presymptomatic identification based on spectral vegetation indices (VIs) is realised. Regarding the original spectral signature, the question of an optimal ratio between classification accuracy and model complexity arises. A careful selection of relevant and non-redundant features depending on classification problem and feature properties is necessary. Instead of assuming the number of features as given, specific features are extracted out of the spectral signature. Moreover, the modelling of signal noise by an analytical description of the spectral signature attains the objective of early detection of plant diseases even before symptoms become visible. Robust fitting of fluorescence spectra is demonstrated using a piecewise fitting by polynomials of low order. Furthermore, various weeds of different growth stages were discriminated by exploiting the hierarchy of weed species. Sequential classification was adjusted to image series from different vegetation periods provided by a database. A problem-specific extraction and selection of relevant features, together with task-oriented classification using SVMs, is necessary for robust identification of pathogens and weeds as early as possible.

3.1 Combination of partly redundant features with non-linear classifier

VIs are related to specific physiological parameters. They are therefore feasible to differentiate healthy and diseased plants (Delalieux et al., 2009; Steddom et al., 2005; Mahlein et al., 2010). All these researchers were able to deduce changes in plant health using VIs, but a specification of the individual disease using VIs was not feasible so far. This problem was analysed based on data from healthy and diseased sugar beet leaves, viz. *Cercospora* leaf spot, sugar beet rust and powdery mildew (Figure 1.2). Figure 3.1 shows that normalised difference vegetation index (NDVI) values of the two classes 'healthy' and 'diseased leaves' are highly overlapping, visualised by two histograms as a discrete description of the data distribution.

A classifier which simultaneously regards various features in order to recognize latent patterns has been designed. In Rumpf et al. (2010) a method is developed which uses SVMs to fully exploit the combined information of nine VIs derived from hyperspectral data. This

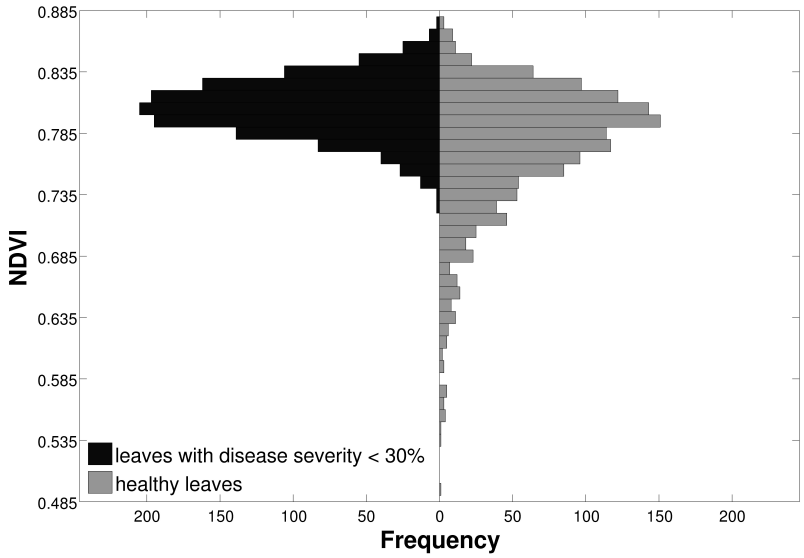
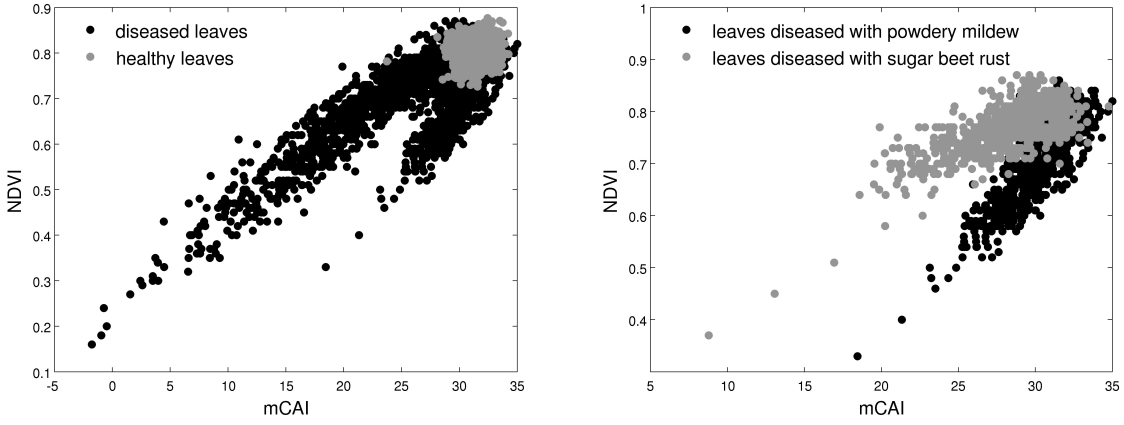


Figure 3.1: Frequencies of normalised difference vegetation index (NDVI) values from healthy and diseased sugar beet leaves. Class differentiation is not feasible.

study aimed for the differentiation between diseased and non-diseased sugar beet leaves. The discrimination between healthy and diseased sugar beet leaves resulted in classification accuracies up to 97% (Rumpf et al., 2010). A further aim was the separation between the three diseases *Cercospora* leaf spot, sugar beet rust and powdery mildew.



(a) Healthy and diseased leaves

(b) Powdery mildew and sugar beet rust

Figure 3.2: Separation between healthy and diseased leaves (3.2a) and between sugar beet rust and powdery mildew (3.2b) using a non-linear discrimination function projected to the two vegetation indices NDVI and mCAI.

Figure 3.2 shows the separation of healthy and diseased leaves (3.2a) and powdery mildew

and sugar beet rust diseased leaves (3.2b) by the two vegetation indices NDVI and modified chlorophyll absorption integral (mCAI). For separation a non-linear function is needed. However, a more specific and sensitive detection and differentiation of plant diseases can be obtained using more than two VIs, correlated to different biological traits as features. In this complex task a non-linear separation also seems superior to a linear separation. Accordingly, a SVM with rbf-kernel was used. Classification accuracy by SVMs was even between nearly one and two percent higher compared to classification methods like ANNs or decision trees (Rumpf et al., 2010). When visual symptoms appeared, a differentiation of the three diseases *Cercospora* leaf spot, sugar beet rust and powdery mildew, was possible with a main accuracy over 88% (Table 3.1).

Prediction	**Ground truth**				
	Healthy	Cercospora leaf spot	Sugar beet rust	Powdery mildew	*Class precision*
Healthy	942	32	47	69	86.42%
Cercospora leaf spot	12	748	61	13	89.69%
Sugar beet rust	20	88	622	14	83.60%
Powdery mildew	46	12	10	834	92.46%
Class recall	92.35%	85.00%	84.05%	89.68%	**88.12%**

Table 3.1: Results of the Support Vector Machines multi-class classification based on spectral vegetation indices (Rumpf et al., 2010).

In addition the classification result improved with increasing disease severity (Rumpf et al., 2010). With only $1-2\%$ diseased leaf area, the classification accuracy was about 65% for all diseases. Starting with a disease severity of $10-15\%$ powdery mildew was differentiated from healthy leaves with high accuracy of about 95% or higher. Concerning sugar beet rust this high accuracy was already reached when $6-9\%$ of the leaf area was diseased and *Cercospora* leaf spot even needs a disease severity of $3-5\%$. Consequently, using vegetation indices as features SVMs were able to handle multi-class classification problems. Contrary simple threshold separations only evaluate the quality of the correlation of parameters.

A major challenge for precision plant protection and phenotyping is the early detection of plant diseases before visual symptoms appear. For plant sciences the effect of presymptomatic processes on the spectral signature is mostly unknown so far. In the following the potential of SVMs for the challenging boundary conditions of presymptomatic stress detection of plant diseases will be described.
Separation between healthy leaves and leaves inoculated with fungal pathogens at early stages of pathogenesis is possible with high accuracy (Rumpf et al., 2010). In all discriminations between healthy leaves and leaves inoculated with *Cercospora beticola*, *Uromyces betae* and *Erysiphe betae*, respectively, a high level of accuracy was achieved (Figure 3.3). The classification results obtained by SVMs are comparative to visually assessed with disease severity. Minor variations between automatic and visual disease assessment occur for powdery mildew. In this case the visual separation was superior between 6–9 days after

inoculation (dai) and after 15 dai. Additionally, highly specific and reliable results for the early detection of plant diseases were achieved already three days after inoculation. An identification of powdery mildew was feasible two days before symptoms become visible; *Cercospora* leaf spot and sugar beet rust were already detected three and five days before symptoms appeared, respectively.

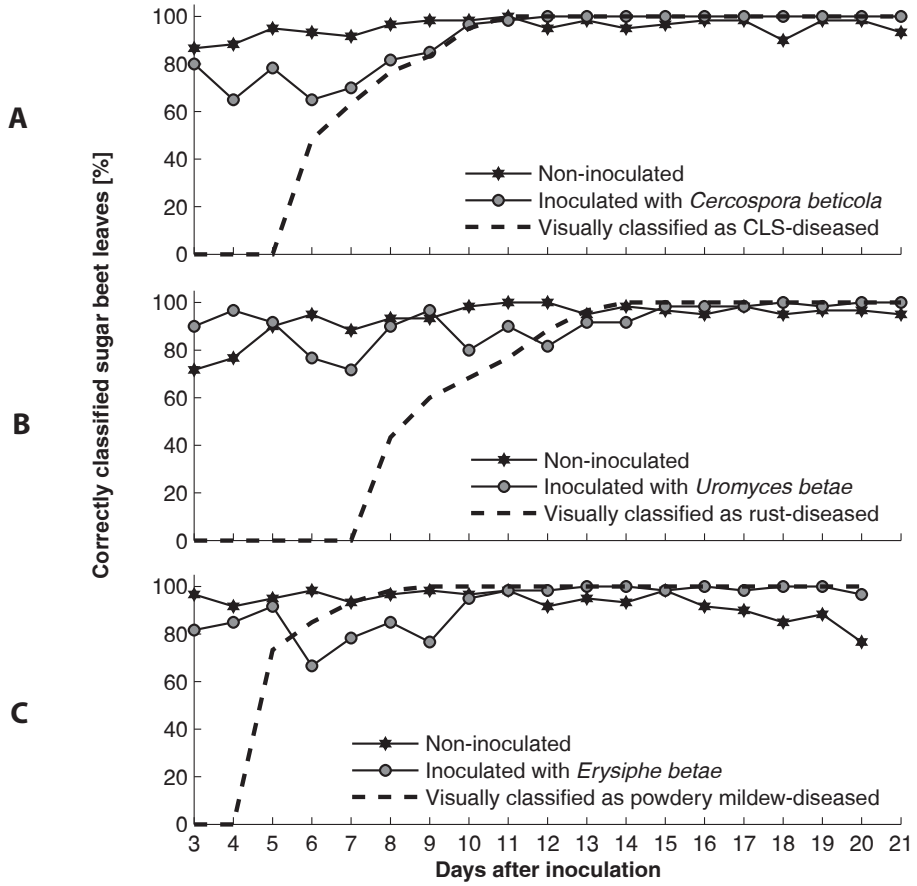


Figure 3.3: Effect of incubation time on the results of SVM classification between healthy sugar beet leaves inoculated with *Cercospora beticola* (A), *Uromyces betae* (B) and *Erysiphe betae* (C) (Rumpf et al., 2010).

3.2 Selection and combination of relevant features

SVMs are able to handle high dimensional feature spaces. Hence, the use of the original spectral signature as features seems promising. A high number of used features, however, results in a more complex model, including relations between every single feature and all classes. Usually every feature enhances classification accuracy until the maximum is asymptoticly achieved. Accordingly, an optimal ratio between classification accuracy and model complexity has to be determined. In addition, computation time during parameter

optimisation of SVMs can significantly be reduced by using fewer features. In some cases, features without information even decrease classification accuracy. In remote sensing Pal and Foody (2010) and Waske et al. (2010) showed that feature selection prior to classification using SVM improves accuracy, particularly if the size of the training set is small. The entire spectrum is not needed to detect biotic or abiotic stress in crop plants (Mewes et al., 2011). Just few regions of the spectral range are relevant for a specific detection of different causal agents of stress. Consequently, the question how to find the best subset of features for classification arises.
In feature selection filter and wrapper approaches are distinguished (Guyon and Elisseeff, 2003; Kohavi and John, 1997). The wrapper approach applies a classifier to many feature subsets and compares classification results. A statement about the relevance of a feature with regard to the used classifier is given. The filter approach evaluates features in a preprocessing step, independent of a specific classifier. This thesis focuses on the filter approach. The selected features are relevant for the classification task. The relevance of the features is determined by diverse information criteria. A most basic form of feature selection is feature ranking. Based on an information criterion each feature is weighted and the top k features are selected. The disadvantage of this approach is that every feature is evaluated and selected independent of the other features. However, a more sophisticated approach aims to find an optimal feature subset where the interaction between features is considered. In the following three different basic approaches, namely the minimum redundancy - maximal relevance criterion (mRMR) (Ding and Peng, 2005), the Hall criterion (Hall and Smith, 1997) and the RELIEF-F (Kononenko, 1994) were applied to identify relevant features for the early detection of plant diseases.

Minimum redundancy - maximal relevance criterion (mRMR):
Entropy and mutual information seem to be adequate concepts in order to identify optimal feature subsets for classification of different plant diseases even at an early stage of infestation (Rumpf et al., 2009a). The minimum redundancy – maximum relevance (mRMR) criterion uses entropy and mutual information in order to consider interactions between selected features during the determination of an appropriate feature subset. Entropy $H(X)$ (Shannon, 1948) is a measure of the uncertainty of a random variable X (in machine learning called feature) which quantifies the expected value of the contained information and is defined as follows

$$H(X) = -\sum_{x \in X} P(x)\, log_2\, P(x)\,. \tag{3.1}$$

The joint entropy $H(X_1, X_2)$ is the entropy of their pairing (Cover and Thomas, 1991). Mutual information $I(X_1; X_2)$ (Cover and Thomas, 1991) in contrast quantifies the dependencies between two random variables X_1 and X_2 which in turn can be calculated by entropy and joint entropy as follows

$$I(X_1; X_2) = \sum_{x_1, x_2} P(x_1, x_2)\, log_2 \frac{P(x_1, x_2)}{P(x_1) \cdot P(x_2)} = H(X_1) + H(X_2) - H(X_1, X_2)\,. \tag{3.2}$$

Features are independent if the mutual information is zero, meaning $P(x_1, x_2) = P(x_1) \cdot P(x_2)$. In this context mutual information measures the amount of information provided by a feature X about a label Y. Accordingly it is often called 'information gain' (Cover and

Thomas, 1991). The basic idea of the mRMR criterion is to find a subset of features $S \subseteq X$ with maximal information gain

$$D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; y) \tag{3.3}$$

and minimal mutual information

$$R = \frac{1}{|S|^2} \sum_{x_i \in S} I(x_i, x_j) \tag{3.4}$$

to each other selected feature (Ding and Peng, 2005). An incremental search method was used to find the optimal feature subset by maximising the difference $D - R$ (Ding and Peng, 2005).

Table 3.2 shows different feature subsets used for classifying specific plant diseases (Rumpf et al., 2009a). With regard to the number of features a classification of *Cercospora* leaf spot was already possible based on two VIs, whereas sugar beet rust needed three VIs and powdery mildew even five VIs. In addition to the different amount of features the feature selection was also different. The classification accuracy based on the feature subset selected by the mRMR criterion was always higher than selecting the feature subset based on correlation strength to label class. The differences were in the range of two and ten percent.

	Selected vegetation indices	**Accuracy** (specificity, sensitivity)
***Cercospora* leaf spot**		
Correlation:	NDVI, mCAI	**91.67%** (98.12%, 85,21%)
mRMR criterion:	ARI, SPAD	**93.39%** (98.75%, 88,02%)
Sugar beet rust		
Correlation:	mCAI, NDVI, ARI	**83.69%** (95.24%, 72.14%)
mRMR criterion:	SPAD, REP, ARI	**93.93%** (96.07%, 91.79%)
Powdery mildew		
Correlation:	SIPI, NDVI, PSSRb, PSSRa, SR	**84.65%** (97.75%, 70.73%)
mRMR criterion:	REP, SPAD, ARI, mCAI, NDVI	**90.25%** (93.33%, 86.98%)

Table 3.2: Selected feature subset of vegetation indices and classification results between different plant diseases and non-inoculated sugar beet leaves using SVMs. The first value presents the accuracy, the others specificity and sensitivity (Rumpf et al., 2009b, modified).

In contrast to VIs as features for classification, wavelengths are more informative for specific plant diseases and able to improve classification results. Dimensionality reduction with PCA according to Wu et al. (2008) surprisingly achieved no improvement in classification. Quite the contrary to using the combination of VIs for discrimination the results were even worse (Rumpf et al., 2010). Above principal components, which are linear combinations of

all available features, have no obvious biological interpretation. Thus the aim was to find relevant wavelengths.

Hall criterion:
In the last section mutual information was used. Mutual information affords the discretisation of features in a preprocessing step, which is only feasible for a rather limited number of dimensions. The information-loss by discretisation of every single wavelength is too high. Accordingly, a minimal subset of relevant wavelengths which is sufficient for separating healthy leaves and leaves inoculated with *Cercospora beticola* is identified using the correlation-based filter algorithm of Hall (Rumpf et al., 2009b). The algorithm of Hall analyses the relevance of a feature subset by considering the intercorrelation among the features (Hall and Smith, 1997). This approach is based on the assumption that good feature subsets contain features that are highly correlated with the class, yet uncorrelated with each other. The following equation formalises this assumption:

$$\text{Merit}_S = \frac{k\overline{r_{cf}}}{\sqrt{k + k\,(k-1)\,\overline{r_{ff}}}} \tag{3.5}$$

where Merit_S is the heuristic 'merit' of a feature subset S containing k features, $\overline{r_{cf}}$ the average feature-class correlation and $\overline{r_{ff}}$ the average feature-feature intercorrelation. In order to identify a nearly optimal feature subset a genetic algorithm (Goldberg, 1989) was used. In this case an identification of leaves inoculated with *Cercospora beticola* with low disease severity $\leq 5\%$ was enabled with classification accuracy over 84% using SVMs based on only seven selected wavelengths, viz. three in the visible spectrum and four in the infrared spectrum (Table 3.3). The classification result was even above 4% higher than based on the combination of VIs.

Used features	**Accuracy** (specificity, sensitivity)
Vegetation indices	**80.52%** (85.00%, 75.22%)
Criterion of Hall	**84.31%** (90.00%, 78.32%)

Table 3.3: Comparison of the classification results between healthy leaves and leaves inoculated with *Cercospora beticola* (disease severity $\leq 5\%$) based on nine vegetation indices and the seven relevant hyperspectral wavelengths identified by the correlation criterion of Hall.

Regarding different levels of disease severity the selected wavelengths by using the Hall criterion were always the same. This indicates that the main changes caused by *Cercospora beticola* influence the same ranges of the hyperspectral signature. With increasing disease severity classification accuracy rose up to almost 100%, without any misclassification (Table 3.4).

RELIEF-F:
The definition of optimal scanning positions in the whole reflection spectrum is also important to develop specific sensors for practical use. These sensors have to be robust, economically priced and user-friendly. Spectral vegetation indices have been shown to be useful

Disease severity	Classification accuracy		
	Accuracy	Healthy leaves	Diseased leaves
without symptoms	**68.09**%	76.67%	57.89%
1 - 5%	**93.16**%	98.00%	86.73%
> 5%	**99.81**%	100%	99.63%

Table 3.4: Classification results between healthy leaves and leaves inoculated with *Cercospora beticola* by different levels of disease severity. Without symptoms means that the inoculated leaf shows no visible symptoms (Rumpf et al., 2009a, modified).

for an indirect detection of plant diseases. Furthermore, it has been shown that normalised wavelength differences seem suitable for a transfer to other datasets. However, vegetation indices were not designed for the separation between different crop plant diseases. The design of specific spectral disease indices (SDIs) for the detection of diseases in major crops was the aim in the study of Mahlein et al. (2013). The best weighted combinations of a single wavelength and a normalised wavelength difference were exhaustively searched testing all possible combinations (Figure 3.4).

Most relevant wavelengths and two band normalised differences from 450 to 950 nm, describing the impact of a disease on sugar beet leaves were extracted from the data set. The amount of single wavelengths and normalised wavelength differences was reduced in order to evaluate all possible combinations of the two subsets without additional assumptions. This selection was accomplished by applying the RELIEF-F algorithm (Kira and Rendell, 1992) to handle non-linear relations between features and multiple classes. In contrast to the Hall criterion which explicitly models correlation between features under the assumption of linearity and Gaussian distribution, RELIEF-F ranks individual features according to their relevance in the context of others (Guyon, 2006). Remaining correlations between the selected features were implicitly considered in the final evaluation during the exhaustive search for developing SDIs.
The RELIEF-F algorithm is noise-tolerant and can handle highly correlated features, which is obviously the case for neighboured wavelengths. RELIEF-F measures the homogeneity of class labels in the local neighbourhood of randomly chosen samples. The key idea of the RELIEF algorithm is to evaluate features according to how well their values distinguish among samples that are near to each other. The extension RELIEF-F is more robust by regarding neighbourhoods of k elements subsuming RELIEF as a special case with $k = 1$ (Kononenko, 1994). RELIEF-F searches for two nearest neighbourhoods for a given sample l (see Listing 3.1). For a given k, 'hit' is the set of k nearest neighbours of the same class and 'miss' from the different class. Feature relevance is determined by the sum of the euclidean distances between nearest hits H_l and nearest misses M_l for all samples used to approximate probabilities.

The optimised disease indices were tested for their ability to detect and to classify healthy and diseased sugar beet leaves. With high accuracy and sensitivity healthy sugar beet leaves

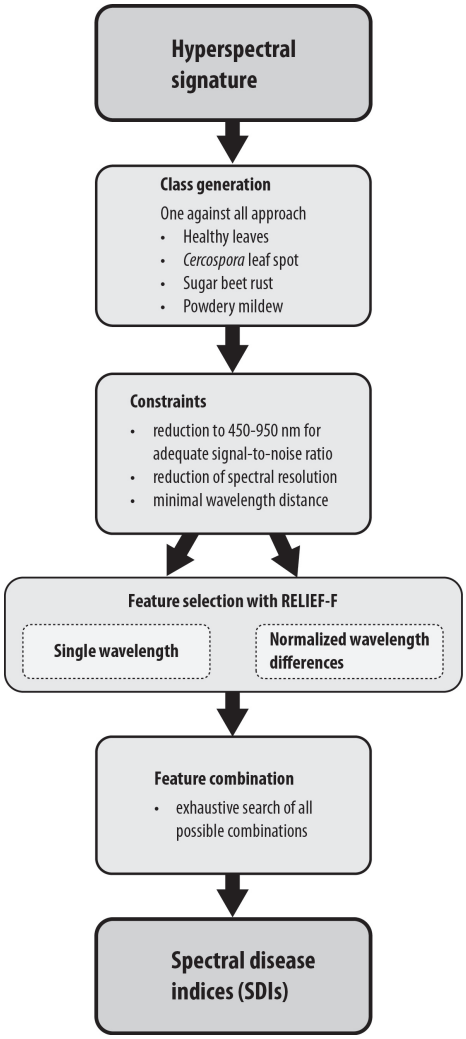


Figure 3.4: Systematical approach and development of spectral disease indices (SDIs) from hyperspectral reflectance data for the four classes healthy (HI), *Cercospora* leaf spot (CLSI), sugar beet rust (SBRI) and powdery mildew (PMI) (Mahlein et al., 2013).

and leaves infected with *Cercospora* leaf spot, sugar beet rust and powdery mildew were classified (balanced classification accuracy: 89%, 92%, 87%, 85%, respectively (Table 3.5)).

In comparison to VIs the classification accuracy using SDIs was much better. Healthy and diseased leaves could be separated by the HI with higher classification accuracy than using traditional VIs. In comparison to the best separating vegetation index photochemical reflection index (PRI) an increase in classification accuracy of 5.5% was achieved by the HI (Mahlein et al., 2013). A disease specific differentiation of the two other diseases and healthy leaves was only realised for *Cercospora* leaf spot by using the best suitable vegetation index mCAI with an accuracy of 89.5%, whereas the CLSI separated with 92.4% accuracy. Sugar beet rust and powdery mildew could not be identified by single VIs. Thus, disease

```
INPUT: A set of features F = F_1, ..., F_m, a set of samples
       R_1, ..., R_n and a class label is given for each R
OUTPUT: A set of feature weights W = W_1, ..., W_m

set all weights W(F) := 0
for i:= 1 to n do (number of samples approximating probabilities)
    begin
    randomly select an sample R_i:
    find k nearest hits H_l and nearest misses M_l;
    for j := 1 to m do (all features)
        begin
        W(F_j) := W(F_j) - sum_{l=1}^{k} difference(F_j, R_i, H_l) / (m · k)
                         + sum_{l=1}^{k} difference(F_j, R_i, M_l) / (m · k);
    end
end
```

Listing 3.1: Pseudo code of the RELIEF-F algorithm for two class classification

Spectral disease index (SDI)	**Index equation**	**Accuracy** (specificity, sensitivity)
Healthy-index (HI)	$\frac{R534-R698}{R534+R698} - \frac{1}{2} \cdot R704$	**89.02%** (83.77%, 94.27%)
Cercospora leaf spot-index (CLSI)	$\frac{R698-R570}{R698+R570} - R734$	**92.42%** (96.31%, 88.52%)
Sugar beet rust-index (SBRI)	$\frac{R570-R513}{R570+R513} + \frac{1}{2} \cdot R704$	**86.98%** (93.70%, 80.26%)
Powdery mildew-index (PMI)	$\frac{R520-R584}{R520+R584} + R724$	**84.52%** (95.28%, 73.77%)

Table 3.5: Classification result for each disease, based on the spectral disease indices (SDIs) *Cercospora* leaf spot-index (CLSI) for the classification of *Cercospora* leaf spot, Sugar beet rust-index (SBRI) for sugar beet rust detection and the Powdery mildew-index (PMI) for powdery mildew detection on sugar beet leaves.

specific indices improve disease detection, identification and monitoring for precision crop protection.

3.3 Extraction of suitable features out of spectral signatures[1]

Many methods for the analysis of hyperspectral or fluorescence data are based on the use of selected wavelengths. These approaches reveal several advantages like reduction of data

[1] The results of this section were mainly achieved by Chistoph Römer.

dimensionality or fast computational time. Fluorescence signature contains a particularly bad signal to noise ratio, so that additional smoothing is necessary (Figure 3.5). A piecewise approximation of the whole fluorescence curve by polynomials was used to cope with the bad signal-noise ratio. Thus the whole information contained in the curve has been considered. Evaluation techniques which take the full spectrum into account promise new insights into early stress reactions. If classifiers rely on single bands they are more prone to outliers and noise, which is especially harmful when the variations in the signature are subtle. Accordingly, filter feature selection algorithms may return single wavelengths, where separation between classes is randomly caused by noise and not by biologically plausible effects.

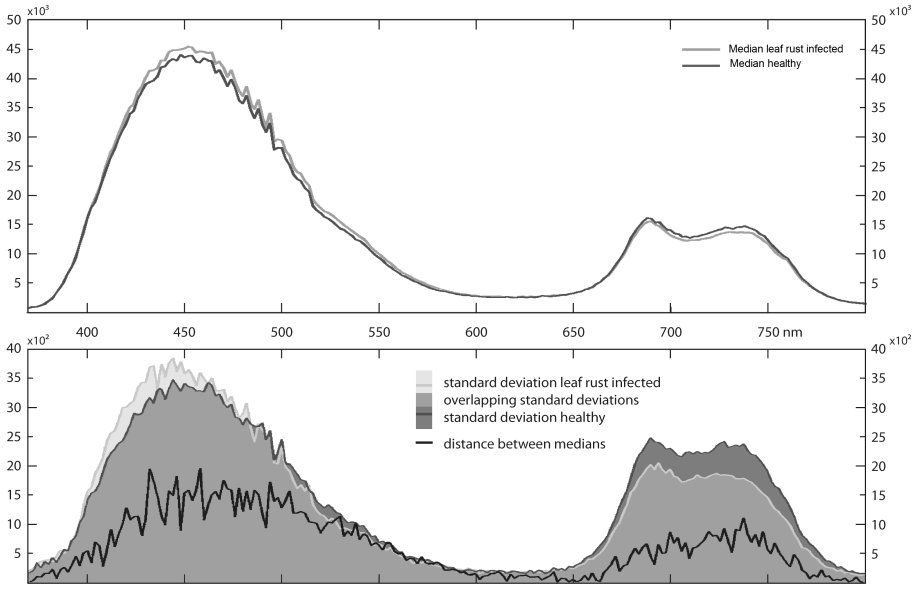


Figure 3.5: Medians of healthy and inoculated wheat leaves (top) and comparison of median difference with standard deviations two days after inoculation (bottom) (Römer et al. (2011)).

Polynomials of low order were used for a piecewise approximation of the whole fluorescence curve of wheat leaves inoculated with *Puccinia triticina* (Römer et al., 2011). This way, the polynomial coefficients contain the information about the form of the approximated piece of the curve. This has the advantage that the characteristic of several hundred wavelengths is compressed into a couple of holistic features. Robustness is also increased, as lower order coefficients are not susceptible to noise and single outliers.
A problem with polynomial interpolation is that the approximation error is comparably large near the break points. Otherwise, as much information as possible has to be extracted from the polynomial coefficients. It is important that wavelengths with high relevance are fitted as accurately as possible. Thus, polynomial break points have to be placed in areas of low interest. Less informative wavelengths were calculated using a filter feature selection algorithm. The break points were placed in an area of low relevance keeping the approximation error low (Figure 3.6).

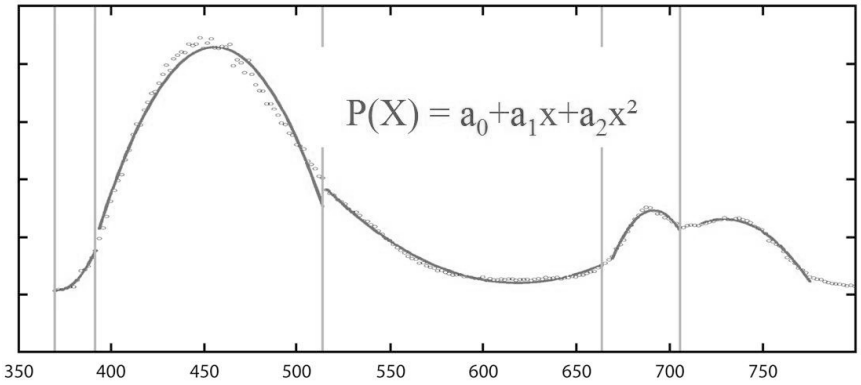


Figure 3.6: Piecewise polynomial approximation of the fluorescence curve of a wheat leaf inoculated with *Puccinia triticina*. The features a_0, a_1 and a_2 contain the information about the characteristics of the curve. Break points are placed in regions of low relevance for the classifier (Römer et al., 2011, modified).

It turned out that wavelengths between 450 and 500 nm and between 550 and 630 nm were informative for separating healthy leaves and leaves inoculated with *Puccinia triticina*. Classification achieved an accuracy of 93% already at the second day after inoculation (dai) (Table 3.6). At this time there were still no symptoms of wheat leaf rust visible.

Data set	SVMs (%)	Decision trees (%)	Artifical neural network (%)
All wavelengths	73.61	70.83	63.89
Polynomials	**93.05**	61.96	82.50

Table 3.6: SVMs compared with decision trees and artificial neural network (Römer et al., 2011, modified).

Polynomials achieved a superior classification accuracy of 93.05% (Table 3.7).

	RELIEF (%)	Principal components (%)	Polynomials (%)
Dai 2	72.22	72.22	**93.05**

Table 3.7: Results for SVMs trained on different, 20 dimensional feature spaces (Römer et al., 2011, modified).

3.4 Structured label space for sequential classification

In the following the detection and differentiation of various weeds based on bispectral images is discussed. To realise a weed specific herbicide application, discrimination, especially between weed species with high difference in the economic threshold is necessary. In comparison to the studies about the early detection and differentiation of plant diseases similarities

exist for the discrimination of different weeds using shape parameters. Again the problem of feature selection arises since some shape parameters are more important than others. Shape parameters were derived from difference images between near infrared (IR, 720 nm) and red light (R, 620–680 nm) which were subsequently filtered with a grey level threshold concluded with some preprocessing steps to reduce noise. Three types of shape parameters (features) were computed: (i) region-based features derived from the pixels of each segment, (ii) contour-based features derived from the border pixels and (iii) features derived from the skeleton of the segments (Jähne, 2001; Weis and Gerhards, 2007). However, a feature that is optimal for separation of two weeds may not be suitable to discriminate between two other weeds. The relevance of features depends on the specific separation task. A sequential classification with specific features may help to overcome this problem, instead of using the same feature set and SVM parameters like in common multi-class SVM.

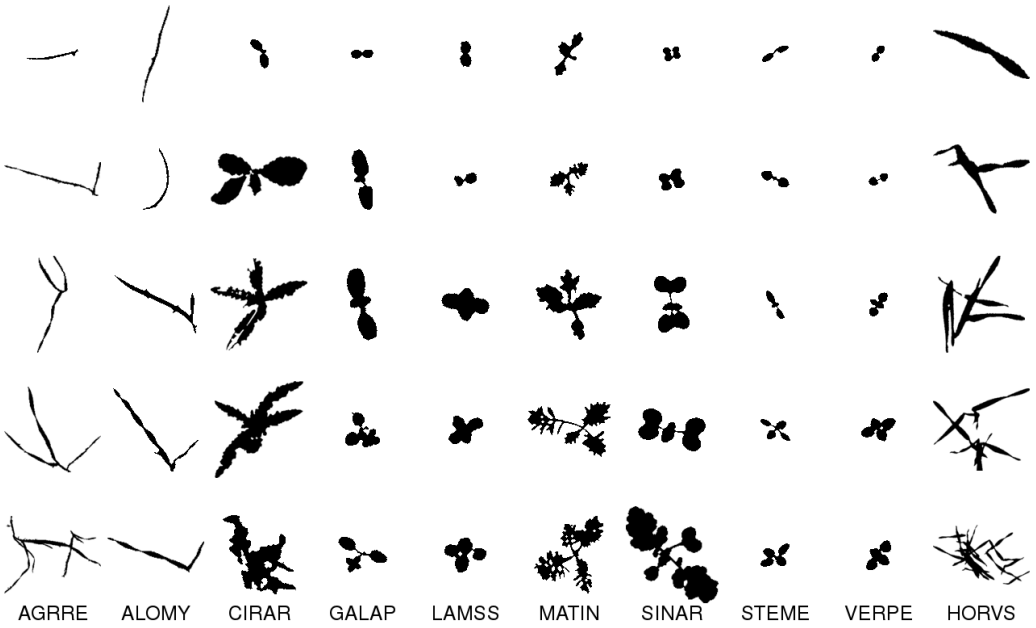


Figure 3.7: Samples of the training data, sorted by class assignment. The segments were scaled to a common maximum width for better visual comparison of the shape. The species are given by their EPPO-Codes: Monocotyledounous: AGRRE: *Agropyron repens* ALOMY: *Alopecurus myosuroides* Dicotyledonous: CIRAR: *Cirsium arvense* GALAP: *Galium aparine* LAMSS: *Lamium sp.* MATIN: *Matricaria inodora* SINAR: *Sinapis arvensis* STEME: *Stellaria media* VERPE: *Veronica persica* Crop: HORVS: *Hordeum vulgare* (Rumpf et al., 2012, modified).

As Figure 3.7 shows, some weeds look very different, whereas others are very similar. Since the fluctuation in economic loss due to the yield effect caused by weeds is high, it is necessary to distinguish correctly between the weed species before applying weed specific herbicides. At the first sight a standard differentiation between many weeds and the crop at an early stage in one step is not possible (Rumpf et al., 2012).

	Ground truth										
Prediction	AGRRE	ALOMY	CIRAR	GALAP	HORVS	LAMSS	MATIN	SINAR	STEME	VERPE	*Class precision*
AGRRE	57	21	0	0	1	0	0	0	0	0	72.15%
ALOMY	30	63	0	0	1	0	0	0	1	0	66.32%
CIRAR	0	0	103	7	3	1	3	8	1	1	81.10%
GALAP	0	0	10	87	5	19	6	1	7	10	60.00%
HORVS	1	2	2	1	78	0	4	0	0	0	88.64%
LAMSS	0	0	2	16	1	84	4	5	4	15	64.12%
MATIN	0	1	1	4	4	5	56	3	12	5	61.54%
SINAR	0	0	13	0	0	2	2	74	4	2	76.29%
STEME	0	1	2	2	1	4	11	1	58	6	67.44%
VERPE	0	0	0	8	0	28	3	3	5	74	61.16%
Class recall	64.77%	71.59%	77.44%	69.60%	82.98%	58.74%	62.92%	77.89%	63.04%	65.49%	**69.25%**

Table 3.8: Results of the one-against-all non-linear SVM classification in one step with weighted features by RELIEF-F. The equal coloured cells highlighted the groups of weeds where common misclassifications occurred (Rumpf et al., 2012).

Particularly, within the groups of weeds, highlighted with equal coloured cells (Table 3.8), misclassifications occurred. However, based on the classification result differences can be seen in severity grade of separation. The discrimination between the subgroups of monocotyledons and dicotyledons is obviously less difficult, whereas the differentiation within the group of dicotyledons seems difficult. Accordingly, the label space was structured into classification tasks which are manageable. In the first step the group of dicotyledons containing the weed species with a low economic threshold was discriminated with accuracy above 99%. In the next step groups of similar dicotyledons have to be identified in order to build additional separable subgroups. Subsequently a sequential classification followed (Figure 3.8). Especially the two weeds *Cirsium arvense* and *Galium aparine*, which cause high economic yield loss, needed high weighted relevant features. The second classification step adequately detected *Cirsium arvense* with almost 83% accuracy. In a final third step the separated subgroup of three dicotyledons containing *Galium aparine* of high interest was differentiated with an accuracy of nearly 89%. *Galium aparine* was classified with an accuracy of 80%.
In all three sequential classification steps different features turned out to be relevant (Table 3.9). The presented sequential classification divided the complex classification task into less complex parts. Thus, improving the classification accuracy and detection rates, especially of the weeds with low economic threshold, except the additional classification errors which occur by the identification of the subgroups. The mean accuracy increases from 69.3% separating all weed species in one step to 79.7% in the third step (Figure 3.8). This sequential classification allows the differentiation between crops and weeds and is further capable for discrimination within the weeds. Both parameters are crucial requirements for a site-specific application of herbicides.

classification step	**most relevant shape parameters**
1. step:	size of the skeleton, area size, mean distance to border, maximum distance to border
2. step:	mean distance from skeleton to border, rearmost distance to center of gravity along main axis, hu moment 2
3. step:	minimum distance to border, hu moment 2, compactness

Table 3.9: Results of the sequential classification using support vector machines based on different weighted shape features. In every step various shape parameters were of particular relevance.

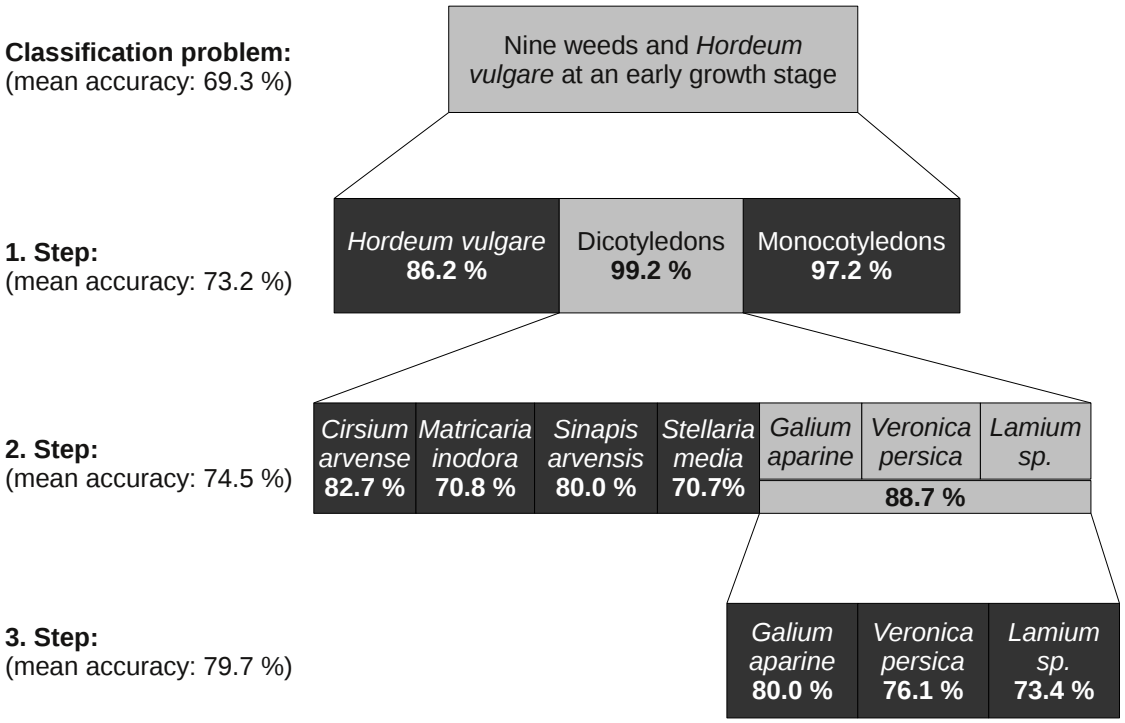


Figure 3.8: Results of the sequential classification of nine weeds and the crop *Hordeum vulgare* at an early growth stage. The well separated weed species are black, whereas grey color means that these weeds have to be discriminated in a separate step. The increasing mean accuracy from step to step is shown on the left side.

4 Conclusion and Perspectives

This current thesis used advanced methods of machine learning integrated with modern sensor techniques for early detection of plant diseases and weeds. The potential of machine learning for precision crop protection was outlined. Support vector machines (SVMs) equipped with non-linear kernels have proved as effective and robust classifiers. For a specific task in plant disease and weed detection, SVMs were adopted to the relevant biological phenomenon. A problem-specific extraction and selection of informative features in combination with task-oriented classification methods achieved a high level of robustness, specificity and earliness. The earliest detection of plant diseases was preponed to a point of time at which no characteristic symptoms became visible.

Plant diseases were detected and classified at an early stage based on spectral vegetation indices. Partly redundant information in the VIs were analysed by non-linear SVMs. In this way, both, discrimination of diseased and non-diseased sugar beet leaves and differentiation between the leaf diseases *Cercospora* leaf spot, sugar beet rust and powdery mildew, was performed. The classification accuracy differs depending on the type and stage of disease. Even, an identification of diseases before characteristic symptoms became visible, was realised with high accuracy.

Using high dimensional data of hyperspectral or fluorescence signatures affords a careful selection to identify disease specific features. In subsequent studies, it was shown that a carefully selected subset of features improves the classification result. For that reason, a large number of features with minor or no additional information, caused by high correlation within the original hyperspectral or fluorescence signature had to be removed. Suitable feature selection methods have to be chosen individually for optical sensor data based on the feature properties, especially number of features and their correlation level. Moreover, the chosen selection method depends on the further analysis of data.

Since fluorescence data has a higher level of noise, information and reliability of a single wavelength is lower. In this context feature extraction by polynomial coefficients proved to be more effective in order to get more robust features which analytically describe the spectral signature and additionally smooth the curve. These holistic features improved the accuracy of classification substantially.

Furthermore, a discrimination of various weeds at different growth stages based on shape parameters was realised by a sequential classification. The separation of the crop plant *Hordeum vulgare* and the weed groups of monocotyledons and dicotyledons was realised with high accuracy. Nevertheless the separation between different species of dicotyledons

needed particular features. To this end, a hierarchy of classes and enforced separation by a top-down traversal of the respective tree was designed. Thus, in each step a different optimal subset of features was used. Further, parameters of each SVM were adapted individually. The mean classification accuracy increased from step to step.

Whereas promising results of applying SVMs for weed and plant disease detection were achieved, exploitation of the potential of modern machine learning methods for the aims of precision agriculture is still in its infancy, at least with regard to early detection of biotic stress. In this thesis, basic principles were examined under controlled conditions. For precision agriculture applications the next step has to be the transfer of the achieved results to the field scale. Here, the influence of various stress factors with uncontrolled environmental and observation conditions, as site-specific characteristics and biologic variability between plants and cultivars, overlap the task-relevant information from sensor data. Instead of using sensors which measure a mean hyperspectral reflectance of a defined observation field, the use of hyperspectral imaging sensors, measuring a hyperspectral signature pixelwise, is one promising approach to cope with challenging field conditions. The proven methods can be applied to each pixel of the sensor image. Hence, a higher spatial resolution can be achieved by hyperspectral imaging. Thus, additional spatial patterns can be integrated in feature space as context knowledge about the neighbourhood. In order to transfer presented machine learning methods to the field, the decoding of the spectral signature is mandatory. This includes a modelling of the main influences, and the separation between the characteristic of stress factors and plant characteristics.

Many challenging boundary conditions have to be considered when transferring the proposed methods for tasks with multiple biotic and abiotic stressors in the field. Beneath the high amount of data which poses a challenge in itself, several additional disturbance variables complicate the process of interpretation and information extraction. Firstly, new stress factors have to be integrated into the classifying data model and the model has to be refined to different disease severities. Hereby, the problem of data labelling appears. Solutions can be found by automatic labelling using unsupervised learning methods. Further information on spatial and temporal development of plant diseases can be obtained by taking the actual disease severity into account. Instead of a binary separation between healthy and diseased plants, quantitative differentiation can be realised. Secondly, a single model for the spread of every stress factor of whole plants is required to distinguish between stress factors with similar spectral impact.

A deeper understanding of spectral signatures of plants is needed to model how several input variables affect the resulting spectra and how these variables are correlated. Machine learning, especially feature selection methods combined with SVMs, can cope with this complex challenge. Models based on data gathered under controlled conditions are a compulsory prerequisite to generate more complex models for field experiments. Sophisticated biological processes can be explained by focusing on single applied stressors and adopted environmental conditions.

Bibliography

Bishop, C., 2006. Pattern recognition and machine learning. Vol. 4. Springer New York.

Boser, E. B., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (COLT'92). ACM Press, Pittsburgh, Pennsylvania (United States), pp. 144–152.

Breiman, L., 2001. Statistical modeling: The two cultures. Statistical Science 16 (3), 199–231.

Brown, M. P. S., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M., Haussler, D., 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proceedings of the National Academy of Sciences 97 (1), 262–267.

Burks, T., Shearer, S., Heath, J., Donohue, K., Jul. 2005. Evaluation of neural-network classifiers for weed species discrimination. Biosystems Engineering 91 (3), 293–304.

Byvatov, E., Schneider, G., 2003. Support vector machine applications in bioinformatics. Applied Bioinformatics 2 (2), 67–77.

Carter, G. A., Knapp, A. K., 2001. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. American Journal of Botany 88, 677–684.

Cortes, C., Vapnik, N. V., 1995. Support-vector networks. Machine Learning 20 (3), 273–297.

Cover, T. M., Thomas, J. A., 1991. Elements of Information Theory, 2nd Edition. Wiley & Sons.

De Wolf, E., Francl, L., 2000. Neural network classification of tan spot and Stagonospora blotch infection periods in a wheat field environment. Phytopathology 90, 108–113.

Delalieux, S., Somers, B., Verstraeten, W. W., van Aardt, A. N. J., Keulemans, W., Coppin, P., 2009. Hyperspectral indices to diagnose leaf biotic stress of apple plants, considering leaf phenology. International Journal of Remote Sensing 30 (8), 1887–1912.

Ding, C., Peng, H., 2005. Minimum redundancy feature selection from microarray gene expression data. Journal of Bioinformatics and Computational Biology 3 (2).

FAO, 2009. Declaration of the world summit on food security. Rome, 16-18 November 2009: Food and Agriculture Organisation of the United Nations).

Finkel, E., 2009. With 'phenomics', plant scientists hope to shift breeding into overdrive. Science 325, 380–381.

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., Haussler, D., 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16 (10), 906–914.

Gaspar, T., Franck, T., Bisbis, B., Kevers, C., Jouve, L., Hasumann, J., Dommes, J., 2002. Concept in plant stress physiology. Application to plant tissue cultures. Plant Growth Regulation 37 (3), 263–285.

Gebbers, R., Adamchuk, V. I., 2010. Precision agriculture and food security. Science 327, 828–831.

Gerhards, R., Sökefeld, M., 2003. Precision farming in weed control – system components and economic benefits. Precision Agriculture 4, 229–234.

Gitelson, A. A., Kaufman, Y. J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote Sensing of Environment 80, 76–87.

Goldberg, D., 1989. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Professional.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., Lander, E. S., 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286, 531–537.

Guyon, I., 2006. Feature extraction: foundations and applications. Vol. 207. Springer Verlag.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157–1182.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Machine Learning 46, 389–422.

Hall, M., Smith, L., 1997. Feature subset selection: a correlation based filter approach. In: 4th International Conference on Neural Information Processing and Intelligent Information Systems. pp. 855–858.

Huang, Y., Lan, Y., Thomson, S., Fang, A., Hoffmann, W., Lacey, R., 2010. Development of soft computing and applications in agricultural and biological engineering. Computers and Electronics in Agriculture 71 (2), 107–127.

Jähne, B., 2001. Digital Image Processing, 5th Edition. Springer-Verlag, Berlin.

Jensen, J., 2007. Remote Sensing of the Environment: An Earth Resource Perspective, 2nd Edition. Prentice Hall.

Jones, D. G. J., Dangl, L. J., 2006. The plant immune system. Nature 444 (7117), 323–329.

Karimi, Y., Prasher, O. S., R.M. Patel, M. R., Kim, H. S., 2006. Application of support vector machine technology for weed and nitrogen stress detection in corn. Computers and Electronics in Agriculture 51 (1-2), 99–109.

Kira, K., Rendell, L. A., 1992. The feature selection problem: Traditional methods and a new algorithm. In: Proceedings of the 10th National Conference on Artificial Intelligence. AAAI'92. AAAI Press, pp. 129–134.

Kohavi, R., John, H. G., 1997. Wrappers for feature subset selection. Artificial Intelligence 97 (1-2), 273–324.

Kononenko, I., 1994. Estimating attributes: Analysis and extensions of relief. In: Proceedings of the European Conference on Machine Learning. Springer-Verlag New York, Inc., Secaucus, NJ, USA, pp. 171–182.

Kumar, S., Hebert, M., 2003. Discriminative random fields: A discriminative framework for contextual interaction in classification. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, pp. 1150–1157.

Lafferty, J., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001). Morgan Kaufmann, pp. 282–289.

Lamb, D., Brown, R., 2001. PA–Precision Agriculture: remote-sensing and mapping of weeds in crops. Journal of Agricultural Engineering Research 78 (2), 117–125.

Mahlein, A., Steiner, U., Hillnhütter, C., Dehne, H., Oerke, E., 2012a. Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. Plant Methods 8 (1), 3.

Mahlein, A.-K., Oerke, E.-C., Steiner, U., Dehne, H.-W., 2012b. Recent advances in sensing plant diseases for precision crop protection. European Journal of Plant Pathology 133 (1), 197–209.

Mahlein, A.-K., Rumpf, T., Welke, P., Dehne, H.-W., Plümer, L., Steiner, U., Oerke, E.-C., 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sensing of Environment 128, 21–30.

Mahlein, A.-K., Steiner, U., Dehne, H.-W., Oerke, E.-C., 2010. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. Precision Agriculture 11 (4), 413–431.

McCallum, A., Freitag, D., Pereira, F., 2000. Maximum entropy markov models for information extraction and segmentation. In: Proceedings of the Seventeenth International Conference on Machine Learning. pp. 591–598.

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. Geoscience and Remote Sensing, IEEE Transactions on 42 (8), 1778–1790.

Mewes, T., Franke, J., Menz, G., 2011. Spectral requirements on airborne hyperspectral remote sensing data for wheat disease detection. Precision Agriculture 12 (6), 795–812.

Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., Ramon, H., 2004. Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. Computers and Electronics in Agriculture 44, 173–188.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66 (3), 247–259.

Mucherino, A., Papajorgji, P., Paradalos, M. P., 2009. A survey of data mining techniques applied to agriculture. Operational Research 9 (2), 121–140.

Oerke, E.-C., Dehne, H.-W., 2004. Safeguarding production — losses in major crops and the role of crop protection. Crop Protection 23, 275–285.

Pal, M., Foody, G., 2010. Feature selection for classification of hyperspectral data by SVM. IEEE Transactions on Geoscience and Remote Sensing 48 (5), 2297–2307.

Patterson, D., 1995. Weeds in a changing climate. Weed Science, 685–701.

Römer, C., Bürling, K., Hunsche, M., Rumpf, T., Noga, G., Plümer, L., 2011. Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with support vector machines. Computers and Electronics in Agriculture 79 (2), 180–188.

Rubinstein, Y., Hastie, T., 1997. Discriminative vs informative learning. In: Proc. Third Int. Conf. on Knowledge Discovery and Data Mining. pp. 49–53.

Rumpf, T., Mahlein, A.-K., Dörschlag, D., Plümer, L., 2009a. Identification of combined vegetation indices for the early detection of plant diseases. In: Neale, M. C., Maltese, A. (Eds.), Proceedings of the SPIE Conference on Sensing for Agriculture, Ecosystems and Hydrology. Vol. 7472. Berlin (Germany).

Rumpf, T., Mahlein, A.-K., Römer, C., Plümer, L., 2009b. Optimal wavelengths for an early identification of *Cercospora beticola* with support vector machines based on hyperspectral reflection data. In: Institute of Electrical and Electronics Engineers (IEEE) (Ed.), 2010 IEEE International Geoscience and Remote Sensing Symposium. Hononulu (Hawai).

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., Plümer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and Electronics in Agriculture 74 (1), 91–99.

Rumpf, T., Römer, C., Weis, M., Sökefeld, M., Gerhards, R., Plümer, L., 2012. Sequential support vector machine classification for small-grain weed species discrimination with special regard to *Cirsium arvense* and *Galium aparine*. Computers and Electronics in Agriculture 80, 89–96.

Sajda, P., 2006. Machine learning for detection and diagnosis of disease. Annual Review Biomedical Engineering 8, 537–565.

Schölkopf, B., Smola, A., 2002. Learning with kernels: Support vector machines, regularization, optimization, and beyond. the MIT Press.

Shannon, C., 1948. A mathematical theory of communication. The Bell System Technical Journal 27, 379–423 and 623–656.

Stafford, J. V., 2000. Implementing precision agriculture in the 21st century. Journal of Agricultural Engineering Research 76, 267–275.

Steddom, K., Bredehoeft, W. M., Khan, M., Rush, M. C., 2005. Comparison of visual and multispectral radiometric disease evaluations of *Cercospora* leaf spot of sugar beet. Plant Disease 89 (2), 153–158.

Thorp, K., Tian, L., Oct. 2004. A review on remote sensing of weeds in agriculture. Precision Agriculture 5 (5), 477–508.

Vapnik, N. V., 2000. The nature of statistical learning theory, 2nd Edition. Statistics for engineering and information science. Springer-Verlag, New York.

Wang, X., Zhang, M., Zhu, J., Geng, S., 2008. Spectral prediction of *Phytophthora infestans* infection on tomatoes using artificial neural network (ANN). International Journal of Remote Sensing 29 (6), 1693–1706.

Waske, B., van der Linden, S., Benediktsson, J., Rabe, A., Hostert, P., 2010. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 48 (7), 2880–2889.

Weis, M., Gerhards, R., June 2007. Feature extraction for the identification of weed species in digital images for the purpose of site-specific weed control. In: Stafford, J. (Ed.), Precision agriculture '07. Vol. 6. 6th European Conference on Precision Agriculture (ECPA), Wageningen Academic Publishers, Netherlands, pp. 537–545.

West, S. J., Bravo, C., Oberti, R., Lemaire, D., Moshou, D., McCartney, H. A., 2003. The potential of optical canopy measurement for targeted control of field crop diseases. Annual Review of Phytopathology 41 (1), 593–614.

Wu, D., Feng, L., Zhang, C., He, Y., 2008. Early detection of *Botrytis cinerea* on eggplant leaves based on visible and near-infrared spectroscopy. Transactions of the ASABE 51 (3), 113–1139.

5 List of Own Publications

5.1 List of publications appended to this thesis

The following list of publications is most relevant for this thesis and appended below.

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., Plümer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and Electronics in Agriculture 74 (1), 91–99.

Rumpf, T., Römer, C., Weis, M., Sökefeld, M., Gerhards, R., Plümer, L., 2012. Sequential support vector machine classification for small-grain weed species discrimination with special regard to *Cirsium arvense* and *Gallium aparine*. Computers and Electronics in Agriculture 80, 89–96.

Mahlein, A.-K., Rumpf, T., Welke, P., Dehne, H.-W., Plümer, L., Steiner, U., Oerke, E.-C., 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sensing of Environment 128, 21–30.

Römer, C., Bürling, K., Hunsche, M., Rumpf, T., Noga, G., Plümer, L., 2011. Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with support vector machines. Computers and Electronics in Agriculture 79 (2), 180–188.

Rumpf, T., Mahlein, A.-K., Dörschlag, D., Plümer, L., 2009b. Identification of combined vegetation indices for the early detection of plant diseases. In: Neale, M. C., Maltese, A. (Eds.), Proceedings of the SPIE Conference on Sensing for Agriculture, Ecosystems and Hydrology. Vol. 7472. Berlin (Germany).

Rumpf, T., Mahlein, A.-K., Christoph, R., Plümer, L., 2009a. Optimal wavelengths for an early identification of cercospora beticola with support vector machines based on hyperspectral reflection data. In: Institute of Electrical and Electronics Engineers (IEEE) (Ed.), 2010 IEEE International Geoscience and Remote Sensing Symposium. Hononulu (Hawai).

5.2 List of publications relevant to this thesis

These publications are related work to which I have contributed. These publications are not included in this thesis, but document the relevance of this work in the scientific community.

Mahlein, A.-K., Rumpf, T., Welke, P., Steiner, U., Dehne, H.-W., Plümer, L., Oerke, E.-C., 2012. Spectral disease indices for the detection and differentiation of plant diseases. In:

Institute of Electrical and Electronics Engineers (IEEE) (Ed.), 2012 IEEE International Geoscience and Remote Sensing Symposium. Munich (Germany) (submitted).

Rumpf, T., Mahlein, A.-K., Dehne, H.-W., Steiner, U., Oerke, E.-C., 2011. Early detection and classification of foliar sugar beet diseases by means of leaf reflectance In: Proc. of 63th International Symposium on Crop Protection, Ghent

Plümer, L., Rumpf, T., Römer, C., 2010: Advanced Machine Learning Methods for Early Detection of Weed and Plant Diseases in Precision Crop Protection. In: Proc. of 3rd Conference on Precision Crop Protection, Bonn

Römer, C., Bürling, K., Rumpf, T., Hunsche, M., Noga, G., Plümer, L., 2010: Early identification of leaf rust on wheat leaves with robust fitting of hyperspectral signatures. In: Proc. of 10th International Conference Precision Agriculture, Denver, USA .

Weis, M., Rumpf, T., Gerhards, R., Plümer, L., 2009: Comparison of different classification algorithms for weed detection from images based on shape parameters. In: Zude, Manuela; Ruckelshausen, Arno (Hg.): Image analysis for agricultural products and processes. Potsdam-Bornim: Leibnitz Institute for Agricultural Enineering (ATB) (Bornimer Agrartechnische Berichte, 69), Bd. 69, S. 53?64.

A Appended Papers

A.1 Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., Plümer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and Electronics in Agriculture 74 (1), 91–99.

Abstract

Automatic methods for an early detection of plant diseases are vital for precision crop protection. The main contribution of this paper is a procedure for the early detection and differentiation of sugar beet diseases based on Support Vector Machines and spectral vegetation indices. The aim was I) to discriminate diseased from non-diseased sugar beet leaves, II) to differentiate between the diseases Cercospora leaf spot, leaf rust and powdery mildew, and III) to identify diseases even before specific symptoms became visible. Hyperspectral data were recorded from healthy leaves and leaves inoculated with the pathogens *Cercospora beticola*, *Uromyces betae* or *Erysiphe betae* causing Cercospora leaf spot, sugar beet rust and powdery mildew, respectively for a period of 21 days after inoculation. Nine spectral vegetation indices, related to physological parameters were used as features for an automatic classification. Early differentiation between healthy and inoculated plants as well as among specific diseases can be achieved by a Support Vector Machine with a radial basis function as kernel.

The discrimination between healthy sugar beet leaves and diseased leaves resulted in classification accuracies up to 97%. The multiple classification between healthy leaves and leaves with symptoms of the three diseases still achieved an accuracy higher than 88%. Furthermore the potential of presymptomatic detection of the plant diseases was demonstrated. Depending on the type and stage of disease the classification accuracy was between 65% and 90%.

Computers and Electronics in Agriculture 74 (2010) 91–99

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original paper

Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance

T. Rumpf[a,*,1], A.-K. Mahlein[b,1], U. Steiner[b], E.-C. Oerke[b], H.-W. Dehne[b], L. Plümer[a]

[a] Institute of Geodesy and Geoinformation, Department of Geoinformation, University of Bonn, Meckenheimer Allee 172, D-53115 Bonn, Germany
[b] Institute of Crop Science and Resource Conversation (INRES – Phytomedicine), University of Bonn, Nussallee 9, D-53115 Bonn, Germany

ARTICLE INFO

Article history:
Received 30 November 2009
Received in revised form 8 April 2010
Accepted 24 June 2010

Hyperspectral reflectance
Vegetation indices
Support Vector Machines
Automatic non-linear classification
Early detection
Cercospora beticola
Uromyces betae
Erysiphe betae
Sugar beet

ABSTRACT

Automatic methods for an early detection of plant diseases are vital for precision crop protection. The main contribution of this paper is a procedure for the early detection and differentiation of sugar beet diseases based on Support Vector Machines and spectral vegetation indices. The aim was (I) to discriminate diseased from non-diseased sugar beet leaves, (II) to differentiate between the diseases Cercospora leaf spot, leaf rust and powdery mildew, and (III) to identify diseases even before specific symptoms became visible. Hyperspectral data were recorded from healthy leaves and leaves inoculated with the pathogens *Cercospora beticola*, *Uromyces betae* or *Erysiphe betae* causing Cercospora leaf spot, sugar beet rust and powdery mildew, respectively for a period of 21 days after inoculation. Nine spectral vegetation indices, related to physiological parameters were used as features for an automatic classification. Early differentiation between healthy and inoculated plants as well as among specific diseases can be achieved by a Support Vector Machine with a radial basis function as kernel.

The discrimination between healthy sugar beet leaves and diseased leaves resulted in classification accuracies up to 97%. The multiple classification between healthy leaves and leaves with symptoms of the three diseases still achieved an accuracy higher than 86%. Furthermore the potential of presymptomatic detection of the plant diseases was demonstrated. Depending on the type and stage of disease the classification accuracy was between 65% and 90%.

1. Introduction

In the past different automatic classification methods have been used to classify remote sensing data and plant observations. Machine learning methods, such as artificial neural networks (ANNs), Decision Trees, K-means, *k* nearest neighbors, and Support Vector Machines (SVMs) have been applied in agricultural research (Mucherino et al., 2009). There is an extensive potential of automatic classification methods in site-specific weed detection (Gutiérrez et al., 2008; Karimi et al., 2006). Wang et al. (2008) predicted *Phytophthora infestans* infection on tomatoes by using ANNs, Camargo and Smith (2009) identified visual symptoms of cotton diseases using SVMs, Ferreiro-Armán et al. (2006) have discriminated between grape varieties from hyperspectral airborne data using SVMs. Furthermore, SVMs turned out as a powerful machine learning technique for general-purpose supervised prediction in biological research like, for example, in the classification of proteins (Park et al., 2005) or gene expression levels (Friedel et al., 2005), for yield prediction in agricultural sciences (Ruß et al., 2008; Ruß, 2009), environmental modeling or stress detection (Karimi et al., 2008) and in remote sensing for land cover classification (Waske and Benediktsson, 2007; Gonçalves et al., 2005) or change detection (He and Laptev, 2009; Nemmour and Chibani, 2006).

Support Vector Machine is a powerful classification method based on the statistical learning theory of Vapnik (1998). In general, classification algorithms aim at finding patterns in empirical data (training data or input data) with regard to label classes. The resulting classification model is used to make a prediction for new unlabeled data. In a sense, supervised learning concludes in finding a function *f* which fits the training data in the best way possible.

However, the identification of such a function is an 'ill-posed problem' since there is an infinite number of functions which describe the discrete data equally well. On the other hand, machine learning is mainly interested in predicting the class of unseen data, i.e. generalisation ability of the classifier, rather than fitting the training data. It is a basic insight of machine learning that the class of feasible functions *f* has to be restricted. In this respect the selection of the function class is a trade-off between a sufficiently modest complexity to achieve good generalisation and to avoid overfitting. Both issues are balanced by SVMs in an optimal way (Vapnik, 2000). SVMs separate two different classes through a hyperplane which

* Corresponding author.
E-mail address: rumpf@igg.uni-bonn.de (T. Rumpf).
[1] Both authors equally contributed to this work.

0168-1699/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2010.06.009

is specified by its normal vector $\vec{w}$ and the bias b. By using a kernel function SVMs are also able to discriminate non-linear, which is required in the classification of plant diseases and especially the discrimination between healthy and diseased plants at early stages.

This study presents a procedure for an automatic classification of foliar sugar beet diseases with special focus on early detection. Foliar diseases are serious threats in sugar beet cultivation. The fungal pathogens *Cercospora beticola* (Sacc), *Erysiphe betae* (Vanha) Weltzien, and *Uromyces betae* (Persoon) Lev. causing Cercospora leaf spot, powdery mildew, and leaf rust, respectively, may induce losses in yield quantity and sugar yield; economic losses may reach up to US$ 1500 ha^{-1} (Wolf and Verreet, 2002). Usually fungal leaf diseases are managed by cultural practices that reduce primary inoculums, planting resistant cultivars, and by applying fungicides (Steddom et al., 2005). Because of the high costs of chemical control and its ecologic impact, one aim of precision farming is to reduce and optimize pesticide applications. The detection and differentiation of several diseases at early stages of epidemics allow a more efficient application of agrochemicals (Hillnhuetter and Mahlein, 2008). But visual monitoring of diseases at early stages in the field is time-consuming and expensive (Steddom et al., 2005; Steiner et al., 2008). Hence, alternative evaluation methods are required. Optical methods like hyperspectral imaging and non-imaging sensors have proved to be useful tools in order to detect changes in plant vitality (Hatfield et al., 2008; West et al., 2003). A high potential of reflectance data in discriminating between healthy and diseased plants has been shown (Bravo et al., 2003; Delalieux et al., 2007; Carroll et al., 2008; Larsolle and Muhammed, 2007; Naidu et al., 2009; Steddom et al., 2003; Wang et al., 2008; Zhang et al., 2002). Spectral vegetation indices (VIs) related to specific physiological parameters have been exploited for the differentiation of healthy and diseased plants (Delalieux et al., 2009; Naidu et al., 2009; Graeff et al., 2006; Steddom et al., 2003, 2005). These VIs from remote sensing are not disease-specific, hence disease discrimination using a single VI is not feasible.

For precision plant protection new disease detection methods must facilitate an automatic classification of the diseases. Data mining techniques, the process of extracting important and useful information from a large set of data (Mucherino et al., 2009; Wu et al., 2008b), seems to solve this complex agricultural problem. Different techniques have been proposed for mining data in terms of disease detection. Bravo et al. (2003) investigated the difference in spectral reflectance between healthy and rust diseased wheat plants. Using a quadratic discriminating model based on the reflectance of the most discriminating four wavebands determined by analysis of covariance (ANCOVA) F-test, they correctly classified diseased and healthy spectra with a classification accuracy of 96%. In a next step they used the neural network Self-Organizing Maps (SOM) successfully to discriminate between healthy plants, nitrogen deficiency, and rust diseased wheat plants in field (Moshou et al., 2006). Wang et al. (2008) spectrally predict late blight infections on tomatoes based on artificial neural networks (ANNs).

Because plant diseases are often associated with specific physiological and visual modifications of their host plants, we applied a method based on the combination of various VIs derived from hyperspectral data and used SVMs to fully exploit their combined information. As a consequence of learning with SVMs several advantages arise. First the function class for the classification model may be arbitrarily complex thus providing the flexibility for difficult classification tasks. A radial basis function (RBF) is parametrized by a simple parameter σ which controls the smoothness of the decision boundary, and therefore the handling is manageable. Reflecting the expectation that some of the training and test samples have been given a wrong label a second parameter C controls the penalisation of this kind of error. In summary, learning with SVMs amounts to find adequate parameters for C and σ.

The objectives of this study were (I) to discriminate diseased from non-diseased sugar beet leaves, (II) to differentiate between the diseases Cercospora leaf spot, leaf rust and powdery mildew, and (III) to identify diseases even before specific symptoms became visible. This study has shown that combined VIs, together with SVMs using an appropriate radial basic function are able to discriminate between the foliar diseases Cercospora leaf spot, sugar beet rust, powdery mildew and healthy plants and as well as between the plant diseases themselves. Furthermore latent infection could be predicted.

2. Materials and methods

2.1. Plant cultivation

Greenhouse experiments were conducted to assess spectral characteristics of sugar beet leaves under controlled conditions. Sugar beet plants (cv. Pauletta, KWS GmbH, Einbeck, Germany) were grown in a commercial substrate (Klasmann-Deilmann GmbH, Germany) in plastic pots (⌀ 13 cm) at 23/20 °C (day/night), 60% relative humidity (RH) and a photoperiod of 16 h. Plants were watered as necessary and fertilized weekly with 100 ml of a 0.2% solution of Poly Crescal (Aglukon GmbH, Düsseldorf, Germany).

2.2. Pathogens

For each treatment, 15 plants were inoculated with the pathogens at growth stage (GS) 14 (= four leaves fully developed). As healthy control 15 plants were kept non-inoculated at 23/20 °C and 60 ± 10% RH. *Cercospora beticola* was inoculated by spraying a spore suspension (4×10^4 conidia ml^{-1}) onto the leaves using a hand sprayer. Subsequently, the plants were covered with plastic bags to realize 100% RH at 25/20 °C for 48 h. Suspensions of *Uromyces betae* (4×10^4 urediniospores ml^{-1}), were sprayed onto the leaves before covering the plants with plastic bags and incubating them for 48 h at 19/16 °C. For further incubation the plants inoculated with *C. beticola* and *U. betae* were transferred to 23/20 °C and 60 ± 10% RH. Plants heavily infested with powdery mildew were used as inoculum source of *Erysiphe betae*. Healthy plants were inoculated in a chamber where a ventilator ran for 25 s to distribute *E. betae* conidia evenly on their leaves. Plants were left over night and afterwards transferred to 23/20 °C, however, separated from the other plants.

2.3. Data recording

Spectral reflectance was measured using a handheld non-imaging spectroradiometer (ASD FieldSpec Pro FR spectrometer, Analytic Spectral Devices, Boulder, USA) with a plant probe foreoptic and a leaf clip holder. The spectral range was from 400 nm to 1050 nm with a spectral resolution of 1.4 nm. The contact probe foreoptic has a 10 mm field of view and an integrated 100 W halogen lamp. Instrument optimization and reflectance calibration were performed prior to sample acquisition; the average of 25 dark current measurements was calibrated to the average of 25 barium sulphate white reference (Spectralon, Labsphere, North Sutton, NH, USA) measurements. Because of the internal light source the integration time was adjusted to 17 ms per scan constantly. Final reflectance spectra were obtained by determining the ratios of data acquired for a sample to data acquired for the white reflectance standard. Each sample scan represented an average of 25 reflectance spectra. In addition, a SPAD-502 chlorophyll meter (Minolta Camera Ltd., Osaka, Japan) was used to measure the leaf greenness as an indicator of the chlorophyll content.

Data from inoculated and non-inoculated leaves were recorded daily till 21 days after inoculation. For each treatment, spectra from

Table 1
Vegetation indices and equations used in this study (R = hyperspectral reflectance). These eight VIs and the SPAD-value are used as features for classification.

Index	Equation	Related to	Reference
Normalized difference vegetation index	$NDVI = \frac{R_{800}-R_{670}}{R_{800}+R_{670}}$	Biomass, leaf area	Rouse et al. (1974)
Simple ratio	$SR = \frac{R_{800}}{R_{670}}$	Biomass, leaf area	Birth and McVey (1968)
Structure insensitive vegetation index	$SIPI = \frac{R_{800}-R_{445}}{R_{800}+R_{680}}$	Ratio carotinoids/chlorrphyll a	Penuelas et al. (1995)
Pigments specific simple ratio	$PSSRa = \frac{R_{800}}{R_{680}}$	Chlorophyll (a/b) content	Blackburn (1998)
	$PSSRb = \frac{R_{800}}{R_{635}}$		
Anthocyanin reflectance index	$ARI = \left(\frac{1}{R_{550}}\right) - \left(\frac{1}{R_{700}}\right)$	Anthocyanin	Gitelson et al. (2001)
Red edge position	$REP = 700 + \frac{40\cdot(R_{RE}-R_{700})}{(R_{740}-R_{700})}$	Inflection point red edge	Guyot and Baret (1988)
	$R_{RE} = \frac{R_{670}+R_{780}}{2}$		
Modified chlorophyll absorption integral	$mCAI = \frac{(R_{545}+R_{752})}{2}\cdot(752-545) - \left(\sum_{R_{545}}^{R_{752}} R\cdot 1.423\right)$	Chlorophyll content	Laudien et al. (2003)

the adaxial surface of the two youngest, fully developed leaves of 15 plants were taken ($n = 30$). RGB images of the leaves were additionally taken. Disease severity of each pathogen was evaluated daily and classified according to Wolf and Verreet (2002). Experiments were repeated twice.

2.4. Spectral vegetation indices

In order to evaluate the suitability of VIs to identify and discriminate among foliar diseases, VIs related to different physiological parameters were calculated (Table 1). Correlation and regression analyses between VIs and disease severity were conducted for each disease (Mahlein et al., 2010).

Relationships between disease severity and VI values were determined by Pearson's correlation coefficient using the Superior Performing System SPSS 17.0 (SPSS Inc., Chicago, IL, USA).

2.5. Classification

In classification generally supervised and unsupervised learning methods are distinguished. Clustering is the most prominent method of unsupervised learning. There are several cluster analysis methods available, viz. DB-SCAN (Ester et al., 1997), K-Means, X-Means (Pelleg and Moore, 2000) or SVM-Clustering (Ben-Hur et al., 2001). In our study we intentionally focused on supervised learning methods, viz. Decision Trees, ANNs and SVMs. Supervised learning needs labeled training samples to learn a model which enables the classifier to predict the class of unseen pattern. For these three classification methods, eight VIs (Table 1) and the SPAD-value were used as features.

2.5.1. Decision Trees

Learning with Decision Trees is one of the most popular and widely used methods for inductive inference. This techniques have their roots in an algorithm which was proposed and refined on by Quinlan (1993). The basic data structure is a tree where the nodes represent features and the edges represent decisions in favor of a (range of) value(s) of the latter. The root and every interior node contains a decision criterion for that feature which has the best chance to reduce the entropy of the respective sample set. Entropy reduction is achieved by splitting the sample set in two or more subsets. This procedure is applied recursively until a given threshold for the minimal entropy is reached. In the case of numeric data, a split value which maximizes entropy reduction is calculated (Mitchell, 1998). After the split into two parts based on the feature with the highest relevance, the next feature which splits the data optimally is determined. Since always one feature is considered at a time, a stepwise axis-parallel boundary is formed in the domain of real numbers. Following a path from the root to a leaf node the Decision Tree corresponds to a rule based classifier.

2.5.2. Artificial neural networks

Artificial neural networks have been inspired by the research on human brain. In the network each node represents a neuron and each link the way two neurons interact. The most popular kind of ANNs is the multilayer perceptron, in which neurons are organized in layers. Each neuron can receive input signals from neurons of the previous layer and send it output to neurons of the successive layer. Further, the activation value is computed as weighted sum of all received signals. The weights between the linked neuron can either increase or decrease the signal. In the simplest case a linear combination of basis functions, especially sigmoid functions, is used to process the activation value of each neuron. ANNs are discussed extensively in Bishop (2005).

2.5.3. Support Vector Machines

In this paragraph Support Vector Machines are discussed in more detail. The simple case of linear SVMs assumes that two classes are linear separable, i.e. their discriminant can be described by a linear function. Starting point is a training data set.

$$\left(\vec{x}_1, y_1\right), \quad \left(\vec{x}_2, y_2\right), \ldots, \left(\vec{x}_n, y_n\right). \tag{1}$$

where $\vec{x}$ denotes a vector with m features $x_1, \ldots, x_m$ and $y_k = 1$ if $\vec{x}_k$ belongs to label class one and $y_k = -1$ if $\vec{x}_k$ was in label class two. Different labels define different classes. In our case the two classes correspond to were healthy sugar beet leaves and leaves inoculated with the different pathogens, respectively. The dot product, which is defined by the formula $\langle \vec{\omega}, \vec{x} \rangle = \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_m x_m$, intuitively specifies the distance (in an m-dimensional Euclidean space). This is used as similarity measure. The basic idea behind SVMs is to separate the two different classes through a hyperplane which is specified by its normal vector $\vec{\omega}$ and the bias b. The hyperplane can be given as

$$\langle \vec{\omega}, \vec{x} \rangle + b = 0 \quad \text{where} \quad \vec{\omega} \in \mathbb{R}^m,\ b \in \mathbb{R}. \tag{2}$$

This yields the corresponding decision function

$$f\left(\vec{x}\right) = \operatorname{sgn}\left(\langle \vec{\omega}, \vec{x} \rangle + b\right). \tag{3}$$

The sign of $f\left(\vec{x}\right)$ depends on the side of the hyperplane where the sample lies. As Vapnik (1998) shows the optimal separating hyperplane is the one which maximizes the distance between the hyperplane and the nearest points of both classes (called margin) and results in the best prediction for unseen data. The samples with minimal distance to the hyperplane are called Support Vectors (SV) and only they define the hyperplane. Thereby the separation is

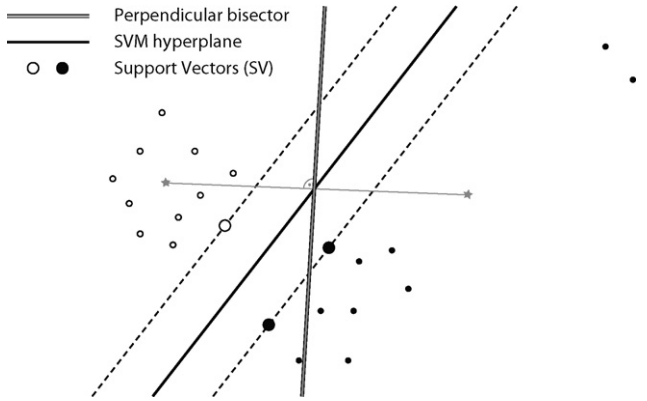


Fig. 1. Comparison of separating two classes. The perpendicular bisector of the line segment connecting the respective centroids (asterisk) of the two classes, obviously differs from the hyperplane with maximal margin constructed by SVMs. The maximal margin hyperplane is more robust against outliers.

robust against outliers — in contrast to the well-known Fisher discriminant which calculates a separating hyperplane based on the centroids of both classes (Fig. 1).

It is assumed that both classes are linearly separable. The specification of a hyperplane by ω and b is unique up to a factor λ. By requiring the scaling of ω and b to be such that the closest point(s) to the hyperplane satisfy $\left|\left\langle \vec{\omega}, \vec{x}_i \right\rangle + b\right| = 1$ the canonical form of a hyperplane is obtained. Accordingly the norm of the normal vector ω is equal to the inverse of the distance, of the closest sample(s) of both classes to the hyperplane. Hence, the optimal separating hyperplane with maximal margin can be formulated as the following quadratic optimization problem:

$$\min_{\vec{\omega}\in\mathbb{R}^m, b\in\mathbb{R}} \quad \tau(\vec{\omega}) = \frac{1}{2}\left|\vec{\omega}\right|^2 \tag{4}$$

$$\text{subject to} \quad y_i \cdot \left(\left\langle \vec{\omega}, \vec{x} \right\rangle + b\right) \geq 1 \quad \forall\, i = 1, \ldots, n. \tag{5}$$

The constraint (5) ensures that $f\left(\vec{x}_i\right)$ yields +1 for $y_i \in \left\{+1\right\}$ and -1 for $y_i \in \left\{-1\right\}$ so that the two classes are separated correctly.

Chapelle (2007) showed that already the primal optimization problem (4) can efficiently be solved, but it is often advisable to solve the dual maximization problem (Boyd and Vandenberghe, 2004). Using the Karush Kuhn Tucker conditions and introducing Lagrange multipliers α_i, ω and b are eliminated (Schölkopf and Smola, 2002).

$$\max_{\alpha\in\mathbb{R}^n} \quad W(\vec{\alpha}) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j \left\langle \vec{x}_i, \vec{x}_j \right\rangle \tag{6}$$

$$\text{subject to} \quad \alpha_i \geq 0 \quad \forall\, i = 1, \ldots, n \tag{7}$$

$$\text{and} \quad \sum_{i=1}^{n}\alpha_i y_i = 0. \tag{8}$$

The optimal solution of (6) only depends on the $\alpha_i \neq 0$ (SV). All other $\vec{x}_i$ which are no SV become zero and have no influence on the construction of the optimal hyperplane which is given by

$$\vec{\omega}^* = \sum_{i=1}^{n}\alpha_i y_i \vec{x}_i \tag{9}$$

$$b^* = -\frac{1}{2}\left\langle \vec{\omega}^*, \vec{x}_a + \vec{x}_b \right\rangle, \tag{10}$$

where $\vec{x}_a \in \left\{+1\right\}$ and $\vec{x}_b \in \left\{-1\right\}$ are any support vectors of both classes. The decision function is then

$$f(\vec{x}) = \operatorname{sgn}\left(\left\langle \vec{\omega}^*, \vec{x} \right\rangle + b^*\right) \tag{11}$$

So far the discussion has been restricted to linear separation. However, in general data will not be linear separable. For this reason a kernel function is introduced to enable an efficient computation (Boser, 1992; Hofmann et al., 2008). It contains an implicit mapping in the feature space without explicitly computing the mapping Φ. The 'kernel trick' can be applied since all feature vectors only occurred in dot products (4), (10), (11) (Schölkopf et al., 1998). Accordingly, the dot product $\left\langle \vec{x}, \vec{x}_i \right\rangle$ as measure of similarity can be substituted by a kernel

$$k\left(\vec{x}, \vec{x}_i\right) = \left(\Phi\left(\vec{x}\right) \cdot \Phi\left(\vec{x}_i\right)\right). \tag{12}$$

One kernel commonly used in practice is the radial basis function (RBF) kernel, i.e.

$$k\left(\vec{x}, \vec{x}_i\right) = \exp\left(\frac{-\left|\vec{x} - \vec{x}_i\right|^2}{2\sigma^2}\right), \tag{13}$$

where the parameter σ controls the smoothness of the decision boundary in the feature space. In our case, this kernel was used to differentiate between healthy and inoculated sugar beet leaves.

Beyond specifying non-linear discriminants by RBF kernels, another generalisation has been proposed which replaces hard margins by soft margins. This way allows to handle noise and prelabeling errors, which often occur in practice. Slack-variables ξ_i are used to relax the hard-margin constraint (Cortes and Vapnik, 1995). The condition in (5) changes to

$$y_i\left(\left\langle \vec{\omega}, \vec{x} \right\rangle + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall\, i = 1, \ldots, n \tag{14}$$

so that some classification errors which depend on ξ_i are allowed. In conclusion the quadratic optimization problem (4) converts to

$$\min_{\vec{\omega}\in\mathbb{R}^m, b\in\mathbb{R}} \quad \tau\left(\vec{\omega}, \xi_i\right) = \frac{1}{2}\left|\vec{\omega}\right|^2 + C\sum_{i=1}^{n}\xi_i, \tag{15}$$

with the regularisation parameter $C \geq 0$. A larger C penalizes a wrong classification more strongly.

Learning an optimal SVM classifier with RBF kernel for the identification of sugar beet diseases is structured as follows. After deducing the vegetation indices out of the original spectral signature the relevance of the indices may be checked. There are several methods to determine the relevance of features, but their detailed discussion is beyond the scope of this paper. In our case it turned out to be appropriate to use all features. In order to specify the best radial basis function and to find an appropriate factor for penalizing classification errors, the parameter C and σ have to be optimized. In this respect, we applied an grid-based approach as recommended by Hsu et al. (2008). Afterwards the classification result of this model is evaluated by cross-validation (see Section 2.7).

2.6. Multi-class classification

Several methods are available to extend dichotomous classifiers such as SVMs for multi-class classification effectively. In this study Chang and Lin's library for SVMs (LIBSVM) has been used for classification (Chang and Lin, 2001). Here the 'one against one' approach (Knerr et al., 1990) is applied. The classifier is constructed for $k(k-1)/2$ times, where k is the number of classes. Each one is trained with data of two classes. The classification decision is based on a majority vote of the class assignments. If classes have identical votes, the one with the smallest index is selected.

Table 2
Important measures in a statistical classification task.

	Condition (e.g. disease detection)		
	True (inoculated leaf)	**False** (healthy leaf)	
Test result	No. of true positive (TP)	No. of false positive (FP)	Pos. predictive value (Precision) ($TP\,(TP+FP)$)
	No. of false negative (FN)	No. of true negative (TN)	Neg. predictive value ($TN\,(TN+FN)$)
	Sensitivity (Recall) $\frac{TP}{TP+FN}$	**Specificity** $\frac{TN}{TN+FP}$	**Accuracy** $\frac{TP+TN}{TP+FP+TN+FN}$

2.7. Evaluation

In general training samples and test samples have to be separated to evaluate the learned model. However, using only half of the data for learning may not be enough to detect all patterns. To realise an optimal utilisation of all information in the training data the model learned was evaluated by cross-validation. The cross-validation splits all examples into a defined number of groups S, out of which $S-1$ are applied to learn a model and the remaining one is used for evaluation. This approach was repeated for all possible choices of evaluation groups. The performance resulted from the average of all possibilities. Frequently 10 groups were chosen. Important performance measures are described in Table 2. Specificity gives the proportion of the correctly classified healthy leaves of all classified healthy leaves. Sensitivity (Recall) gives the proportion of correctly classified inoculated sugar beet leaves in relation to all classified inoculated leaves. The accuracy is given by the average of sensitivity and specificity.

3. Results

3.1. Disease development

Non-inoculated plants stayed healthy over the experiment period. Inoculated plants were first colonized without symptoms, after a latency period typical symptoms appeared. The development of diseases and the symptoms varied significantly for the three pathogens (Fig. 2). Small chloroses were the first symptoms of Cercospora leaf spot. After 6–8 days of incubation, these spots became necrotic and the characteristic red margin of the spots became visible. Fourteen days after inoculation the spots coalesced and formed large necrotic areas. First chloroses due to *Uromyces betae* became visible 9 days after inoculation (dai). At later stages, rust spores ruptured the epidermis and amber uredinia became visible on the upper and lower side of leaves. Symptoms of powdery mildew firstly appeared 5 dai. Small colonies visible on the upper side of leaves in the beginning rapidly expanded and after 14 dai the white, fluffy mycelium covered the total leaf surface.

3.2. Dichotomous classification between healthy leaves and leaves with disease symptoms

In a first dichotomous approach SVMs were used for the differentiation between the two classes, non-inoculated, healthy leaves and leaves inoculated with one of the three leaf pathogens. VIs calculated from characteristic reflectance spectra (Fig. 3) and the SPAD-value were used for classification.

The results showed that the specificity of the classification was always lower than the sensitivity. The classification error range was 7% to almost 3% (Table 3). In comparison with the classification results of Decision Trees and ANNs the classification error of SVMs was always lower.

The classification accuracy increased with increasing disease severity (Fig. 4). Differences in the number of leaves in the disease class give additional information on the reliability of classification results. With only 1–2% diseased leaf area, the classification accuracy was about 65% for all diseases. The accuracy of differentiating between healthy leaves and leaves with Cercospora leaf spot symptoms rapidly increased with a disease severity of 3–5%. When more than 10% of the leaf area was covered by leaf spots, the classification accuracy reached 100% (Fig. 4).

The accuracy of detecting sugar beet rust and powdery mildew symptoms increased less rapidly. At sugar beet rust disease severity stages of 6–9% the classification accuracy was about 95%. Leaves with powdery mildew could be differentiated from healthy leaves

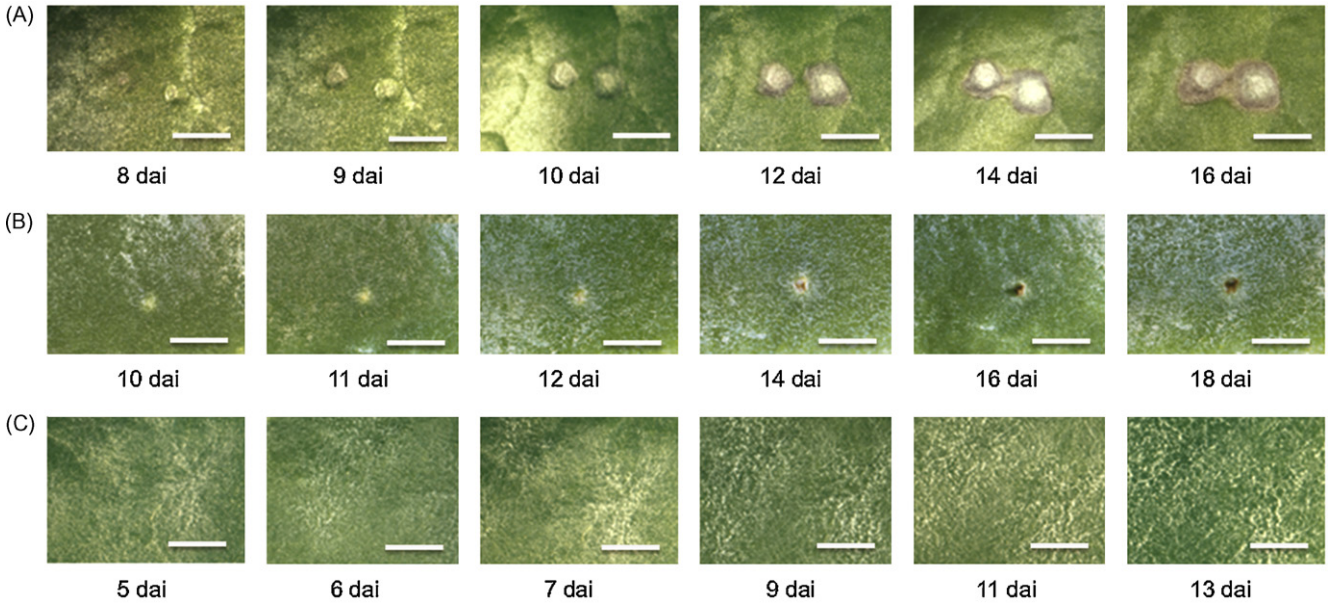


Fig. 2. Development of disease specific symptoms of three diseases on sugar beet leaves; Cercospora leaf spot (A), sugar beet rust (B), and powdery mildew (C) (dai = days after inoculation; bar = 4 mm).

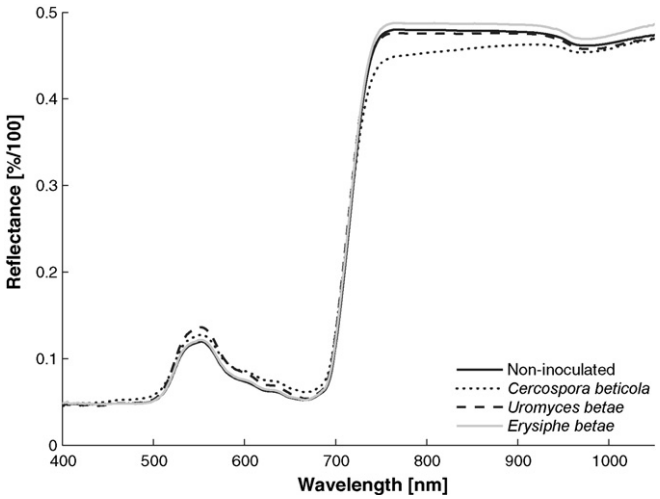


Fig. 3. Spectral signatures of the non-inoculated sugar beet leaves and sugar beet leaves inoculated with the different pathogens. The showed reflectance spectra are the mean of all sugar beet leaves with diseases severity under 10%, respectively.

Table 3
Results of the dichotomous classification based on spectral vegetation indices. Comparison between the three different classification methods Decison Tree, ANNs and SVMs.

Leaf disease	Classification error [%]		
	Accuracy	Specificity (healthy leaves)	Sensitivity (diseased leaves)
Decision Trees			
Cercospora leaf spot	4.67	1.15	8.52
Sugar beet rust	7.15	4.76	9.86
Powdery mildew	13.18	5.59	21.51
ANNs			
Cercospora leaf spot	3.64	3.02	4.32
Sugar beet rust	4.30	2.38	6.49
Powdery mildew	8.41	11.67	4.84
SVMs			
Cercospora leaf spot	2.88	1.35	4.55
Sugar beet rust	3.73	3.21	4.32
Powdery mildew	6.92	5.29	8.71

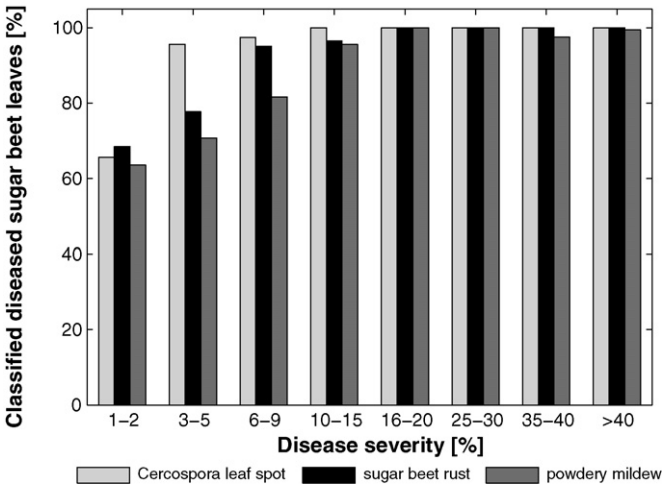


Fig. 4. Classification results of non-inoculated sugar beet leaves and sugar beet leaves inoculated with the different pathogens depending to disease severity.

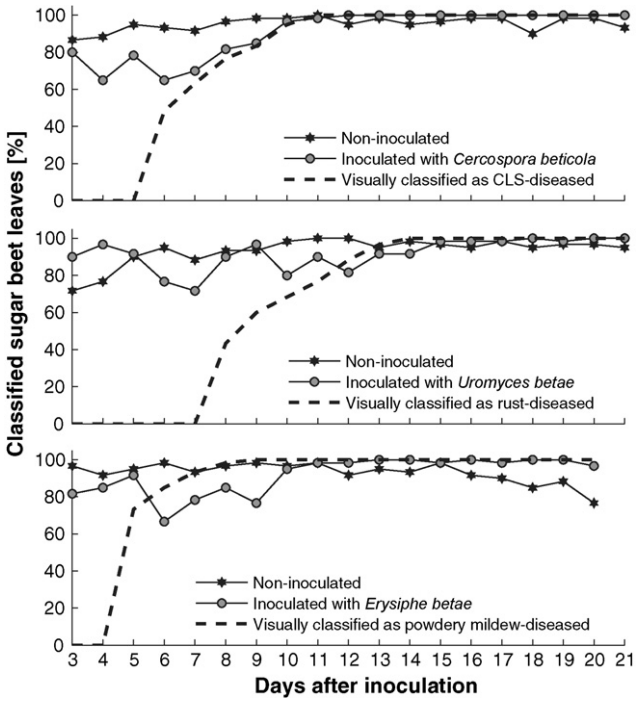


Fig. 5. Effect of incubation time on the results of SVM classification between healthy sugar beet leaves inoculated with *Cercospora beticola* (A), *Uromyces betae* (B) and *Erysiphe betae* (C), respectively.

with an accuracy of about 95% when 10–15% of the leaf area was covered with fluffy mycelium (Fig. 4).

3.3. Multi-class classification among healthy leaves and leaves with symptoms of three diseases

Table 4 summarises the results of the model learned which classified healthy sugar beet leaves and leaves diseased with Cercospora leaf spot, sugar beet rust and powdery mildew, respectively (multi-class classification). The overall classification accuracy was better than 86%, but the difference between the classes were low. The class recall of each class ranged between 84% and more than 92%. The class of healthy leaves was classified best. Classification difficulties occurred in separating between sugar beet rust and Cercospora leaf spot and also in the classification between powdery mildew and healthy sugar beet leaves.

3.4. Classification of healthy leaves and leaves inoculated with fungal pathogens at early stages of pathogenesis

For the differentiation between healthy sugar beet leaves and leaves inoculated with one of the different pathogens before specific disease symptoms became visible — VI data were used starting 3 dai. First symptoms of Cercospora leaf spot appeared 6 dai, rust 8 dai and powdery mildew 5 dai. Leaves inoculated with *C. beticola* were correctly classified by SVMs with an accuracy range from 65% to 80%, even before symptoms became visible (Fig. 5A). When specific symptoms occurred 6 dai, the classification accuracy steadily increased until 12 dai where it converged at 100%. Throughout the 21 days of the experiment, the classification accuracy of the automatic procedure was consistent to visually classified Cercospora leaf spot-infected leaves. The classification accuracy of healthy leaves was 3 dai almost 87% and reached >95% starting from 8 dai.

Although first symptoms of sugar beet rust appeared only 8 dai, a classification accuracy of 90% for *U. betae*-infected leaves was reached 3–5 dai (Fig. 5B). One day before first rust symp-

Table 4
Results of the Support Vector Machines multi-class classification based on spectral vegetation indices.

	Ground truth				
Prediction	Healthy	Cercospora leaf spot	Sugar beet rust	Powdery mildew	*Class precision*
Healthy	942	32	47	69	86.42%
Cercospora leaf spot	12	748	61	13	89.69%
Sugar beet rust	20	88	622	14	83.60%
Powdery mildew	46	12	10	834	92.46%
Class recall	92.35%	85.00%	84.05%	89.68%	*86.42%*

toms became visible, the sensitivity decreased to about 71% and increased again 15 dai to over 98%. The results of the automatic classification were inferior to the visual ratings only between days 12 and 14 after inoculation. Healthy sugar beet leaves were classified with an accuracy of 72% at early stages of leaf colonization by *U. betae*, but from 10 dai the classification accuracy was always greater than 95%.

Already 3 dai the classification accuracy of powdery mildew was above 80% and increased to almost 92% 5 dai (Fig. 5C). After the appearance of first visible colonies the classification rate decreased to under 70% 6 dai, subsequently increased again from 10 dai to 20 dai to over 95%. In contrast with the results for the leaves colonized by the other pathogens from 6 dai to 9 dai the visual classification of *E. betae*-infected leaves was superior (13–23%) to the automatic classification procedure. In the beginning of the experiment the classification rate of healthy leaves was above 91%; in the third week, however, it decreased to 77% 20 dai.

4. Discussion and conclusion

This article demonstrated the feasibility of presymptomatic identification of foliar sugar beet diseases. SVMs proved to be a powerful tool for automatic classification. The main advantages of SVMs go back to their generalisation ability which is achieved by using the maximum margin hyperplane for separation and the application of non-linear discriminant functions such as RBF kernels (Vapnik, 2000). In addition high performance in learning the best model, the comparison of the different classifiers shows that SVMs use the inherent information of the vegetation indices in an optimal way. Above, not only the identification of diseased leaves (dichotomous approach), but also the differentiation between distinct diseases (multi-class approach) can be realised. Comparing SVMs and ANNs both are discriminant methods without any assumption of a density distribution. There are important differences, however, which are important both from a theoretical and practical perspective. In order to find the best discriminant SVMs have to solve a quadratic optimization problem. Convexity of this problem ensures a unique, global solution. The applied techniques tend to converge rapidly thus providing good performance. In contrast, the optimization problem for ANNs is not convex, there are in general several local optima rather than a unique global solution, and good performance of the optimizer is not guaranteed (Suykens, 2002). Above, ANNs prone to overfitting by using empirical risk minimization, while SVMs use structural risk minimization for controlling the generalisation ability (Vapnik, 2000).

Recently, there has been growing interest in exploring the potential of SVMs for early detection of plant diseases. In a recent publication Camargo and Smith (2009) used SVMs for the identification of visual symptoms of plant diseases based on RGB images. They extracted a set of features, like shape, texture or grey level, from segmented diseased regions of cotton leaves. Using all features they reached a classification accuracy of 90%. Whereas the differentiation between different plant diseases, especially at a very earlier stage before symptoms are visible, has not been considered until now. Wu et al. (2008a) recently showed that an early detection of *Botryis cinerea* on eggplant leaves is possible, even before symptoms appeared. Owing to the complexity of the original spectral data, principal component analysis was applied to reduce the numerous wavelengths to several principal components (PCs) in order to decrease the amount of calculation and improve the accuracy. These PCs were set as input variables of back-propagation neural networks. In contrast to Wu et al. (2008a), we achieved classification errors between barely 7% and under 3%, dependent on the plant disease, by using VIs, which contains combinations of individual wavelength. In Table 5 a comparison of the results, by using the first six PCs similar to Wu et al. (2008a), is presented. For each plant disease the classification error by using VIs as features is noticeably inferior (Table 5). In order to additionally improve the detection of the three above mentioned plant diseases one could try to find wavelengths or combination of wavelengths more specific for the given task.

Most VIs are based on only two or three specific wavelengths of the visible- and near infrared region of the reflection spectrum. These wavelengths are highly correlated to specific physiological plant parameters, like pigment content, biomass or vitality (Thenkabail et al., 2000). During pathogenesis and symptom development, plant pathogens are affecting these physiological parameters. Earlier studies successfully used VIs to discriminate between healthy and diseased plants. Derived from airborne hyperspectral imagery Carroll et al. (2008) have shown that a detection of European corn borer is possible. In literature the potential of leaf spectral reflectance changes and VIs for detecting leaf biotic stress of apple plants (Delalieux et al., 2009) or virus infected grapevines (Naidu et al., 2009) was investigated. Changes in reflectance of sugar beet plants caused by diseases are strongly correlated to the stage of pathogenesis and diseases severity (Steddom et al., 2005; Mahlein et al., 2009). All these researchers could deduce changes in plant health using VIs, but the detection of a specific disease using VIs was not feasible so far. As we know, this work has been the first study in literature using combinations of VIs to predict plant diseases in a very early stage. A combination of spec-

Table 5
Comparison of the results of the dichotomous classification based on spectral vegetation indices or principal components.

Leaf disease	Classification error [%]		
	Accuracy	Specificity (healthy leaves)	Sensitivity (diseased leaves)
SVMs (VIs)			
Cercospora leaf spot	2.88	1.35	4.55
Sugar beet rust	3.73	3.21	4.32
Powdery mildew	6.92	5.29	8.71
SVMs (PCs)			
Cercospora leaf spot	3.47	0.62	6.58
Sugar beet rust	9.34	1.90	17.96
Powdery mildew	14.26	4.51	24.95

tral vegetation indices, based on different wavelengths, describing different physiological parameters enhances the information content for an automatic classification and improves the classification accuracy.

By using combined VIs both, the feature combination and the number of features proves to be dependent on the specific plant diseases (Rumpf et al., 2009). For detection of Cercospora leaf spot only two VIs are necessary whereas three and more are needed to detect leaf rust or powdery mildew. Concerning an early detection of plant diseases before visible symptoms appeared the best classification accuracy is achieved by using the information of all nine VIs (Rumpf et al., 2009).

This study proved the feasibility of early detection of sugar beet leaves respectively inoculated with different pathogens by using SVMs. Inoculated leaves were already identified 3 days after inoculation. Even before characteristic symptoms are visible, several putative modifications in cellular leaf structure, for example changes in water content at infection sites, initiating cell death caused by fungal toxins or resistance reactions of plant tissue occur (Daub and Ehrenshaft, 2000; Jones and Dangl, 2006; Knogge, 1996). These modifications are associated with changes in spectral reflectance characteristics (Jacquemoud and Ustin, 2001). Comparative observations of the pathogenesis of the three foliar sugar beet pathogens suggested that the influence on spectral signatures depends on the intensity of physiological changes and on the extent of the symptoms (Mahlein et al., 2010), because physiological interactions between a fungal pathogen and a host plant vary depending on the pathogen (Glazebrook, 2005). Further investigations are necessary to describe these interactions for each pathogen in detail. The classification accuracy differs also between the pathogens. Difficulties emerge especially at early development stages of the characteristic symptoms. Slight variations can be accounted for the original source of the reflectance data. The reflectance curve, measured with a non-imaging sensor, is always the mean of the reflectance of healthy and diseased plant tissue. This results in a number of problems that are typical for such single point measurements (Scholten et al., 2005). Due to the very small size of sugar beet rust colonies (0.5–1.5 mm), a precise classification at early stages or in the case that only few pustules occur is very complex. Also a distinctive detection of powdery mildew at early stages is challenging. First powdery mildew symptoms are fluffy white mycelia covering the leaf surface. This fungal tissue on the leaf surface shifts the spectral signature like a dusty coat. It could be anticipated that imaging hyperspectral sensors can improve a hyperspectral disease detection through a better understanding of the pathogen host interactions (Chaerle and van der Straeten, 2001). With imaging sensor systems a pixel wise attribution of disease specific symptoms and tissue could be facilitated (Steiner et al., 2008).

In this study, our results showed that the SVMs method based on VIs has successfully been applied to identify leaves inoculated with *Cercospora beticola*, *Uromyces betae* or *Erysiphe betae*. One aim of our work has been to prove a simple disease detection technique which can easily be modified for online application in the field. Based on this results specific sensors for practical use may be developed in future. These sensors have to be robust, economically priced and user-friendly. By using the SVMs method based on VIs related to physiological parameters of plants, this procedure is even applicable to other plant–pathogen systems. In order to improve the early identification and discrimination of plant diseases on sugar beet leaves, machine learning techniques will be applied on the original hyperspectral data. In this context the focus will be the definition of optimal scanning positions in the whole reflection spectrum, and the development of specific indices for the detection of leaves that are inoculated with one of the three foliar sugar beet diseases.

Acknowledgement

This study has been conducted within the Research Training Group 722 'Information Techniques for Precision Crop Protection', funded by the German Research Foundation (DFG).

References

Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V., 2001. Support vector clustering. Journal of Machine Learning Research 2, 125–137.
Birth, S.G., McVey, R.G., 1968. Measuring the color of growing turf with a reflectance spectrophotometer. Agronomy Journal 60 (6), 640–643.
Bishop, M.C., 2005. Neural Networks for Pattern Recognition, reprint Edition. Oxford University Press, Oxford, Great Britain.
Blackburn, A.G., 1998. Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. International Journal of Remote Sensing 19 (4), 657–675.
Boser, E.B., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (COLT'92), ACM Press, Pittsburgh, PA, United States, pp. 144–152.
Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, New York, NY, USA.
Bravo, C., Moshou, D., West, J., McCartney, A., Ramon, H., 2003. Early disease detection in wheat fields using spectral reflectance. Biosystems Engineering 84 (2), 137–145.
Camargo, A., Smith, J., 2009. Image pattern classification for the identification of disease causing agents in plants. Computers and Electronics in Agriculture 66 (2), 121–125.
Carroll, W.M., Glaser, A.J., Hellmich, L.R., Hunt, E.T., Sappington, W.T., Calvin, D., Copenhaver, K., Fridgen, J., 2008. Use of spectral vegetation indices derived from airborne hyperspectral imagery for detection of European corn borer infestation in Iowa corn plots. Journal of Economic Entomology 101 (5), 1614–1623.
Chaerle, L., van der Straeten, D., 2001. Seeing is believing: imaging techniques to monitor plant health. Biochimica et Biophysica Acta (BBA) Gene Structure and Expression 1519 (3), 153–166.
Chang, C.C., Lin, J.C., 2001. Libsvm: A Library for Support Vector Machines. Accessed 10/2009.
Chapelle, O., 2007. Training a support vector machine in the primal. Neural Computation 19 (5), 1155–1178.
Cortes, C., Vapnik, N.V., 1995. Support-vector networks. Machine Learning 20 (3), 273–297.
Daub, E.M., Ehrenshaft, M., 2000. The photoactivated cercospora toxin cercospor: contributions to plant disease, fundamental biology. Annual Review of Phytopathology 38, 461–490.
Delalieux, S., Somers, B., Verstraeten, W.W., van Aardt, A.N.J., Keulemans, W., Coppin, P., 2009. Hyperspectral indices to diagnose leaf biotic stress of apple plants, considering leaf phenology. International Journal of Remote Sensing 30 (8), 1887–1912.
Delalieux, S., van Aardt, J., Keulemans, W., Schrevens, E., Coppin, P., 2007. Detection of biotic stress (*venturia inaequalis*) in apple trees using hyperspectral data: non-parametric statistical approaches and physiological implications. European Journal of Agronomy 27 (1), 130–143.
Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1997. Density-connected sets and their application for trend detection in spatial databases. In: Proceeding of 3rd International Conference Knowledge Discovery and Data Mining (KDD'97), AAAI Press, Newport Beach, CA, USA, pp. 10–15.
Ferreiro-Armán, M., Da Costa, J., Homayouni, S., Martn-Herrero, J., 2006. Hyperspectral image analysis for precision viticulture. In: Image Analysis and Recognition. Springer-Verlag, pp. 730–741.
Friedel, C.C., Jahn, H.V.K., Sommer, S., Rudd, S., Mewes, W.H., Tetko, V.I., 2005. Support vector machines for separation of mixed plant–pathogen EST collections based on codon usage. Bioinformatics 21 (8), 1383–1388.
Gitelson, A.A., Merzlyak, N.M., Chivkunova, B.O., 2001. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. Photochemistry and Photobiology 74 (1), 38–45.
Glazebrook, J., 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annual Review of Phytopathology 43, 205–227.
Gonçalves, P., Carrao, H., Pinheiro, A., Caetano, M., 2005. Land cover classification with support vector machine applied to modis imagery. In: Proceedings of the 25th EARSeL Symposium, Porto, Portugal, pp. 517–525.
Graeff, S., Link, J., Claupein, W., 2006. Identification of powdery mildew (*Erysiphe graminis* sp. tritici) and take-all disease (*Gaeumannomyces graminis* sp. tritici) in wheat (*Triticum aestivum* l.) by means of leaf reflectance measurements. Central European Journal of Biology 1 (2), 275–288.
Gutiérrez, P.A., López-Granados, F., Pena-Barragán, J.M., Jurado-Expósito, M., Gómez-Casero, M.T., Hervás-Martinez, C., 2008. Mapping sunflower yield as affected by *Ridolfia segetum* patches and elevation by appling evolutionary product unit neural networks to remote sensed data. Computers and Electronics in Agriculture 60 (2), 122–132.
Guyot, G., Baret, F., 1988. Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux. In: Guyenne, D.T., Hunt, J.J. (Eds.), Spectral Signatures of Objects in Remote Sensing. Vol. 287 of ESA Special Publication, pp. 279–286.

Hatfield, L.J., Gitelson, A.A., Schepers, S.J., Walthall, L.C., 2008. Application of spectral remote sensing for agronomic decisions. Agronomy Journal 100 (3), 117–131.
He, L., Laptev, I., 2009. Robust change detection in dense urban areas via svm classifier. In: 5th GRSS/ISPRS Workshop on Data Fusion and Remote Sensing over Urban Areas (URBAN 2009), Shanghai, China.
Hillnhuetter, C., Mahlein, A.-K., 2008. Early detection and localisation of sugar beet diseases: new approaches. Gesunde Pflanzen 60 (4), 143–149.
Hofmann, T., Schölkopf, B., Smola, J.A., 2008. Kernel methods in machine learning. Annals of Statistics 36 (3), 1171–1220.
Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2008. A practical guide to support vector classification. In: Technical Report. Department of Computer Science, National Taiwan University, Taipei 106, Taiwan.
Jacquemoud, S., Ustin, L.S., 2001. Leaf optical properties: a state of the art. In: 8th International Symposium of Physical Measurements & Signatures in Remote Sensing, CNES, Aussois, France, pp. 223–332.
Jones, D.G.J., Dangl, L.J., 2006. The plant immune system. Nature 444 (7117), 323–329.
Karimi, Y., Prasher, O.S., Patel, R.M., Kim, M.R.H.S., 2006. Application of support vector machine technology for weed and nitrogen stress detection in corn. Computers and Electronics in Agriculture 51 (1–2), 99–109.
Karimi, Y., Prasher, S., Madani, A., Kim, S., 2008. Application of support vector machine technology for the estimation of crop bio, physical parameters using aerial hyperspectral observations. Canadian Biosystems Engineering 50 (7), 13–20.
Knerr, S., Personnaz, L., Dreyfus, G., 1990. Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: Fogel man, J. (Ed.), Neurocomputing: Algorithms, Architectures and Applications. Springer-Verlag.
Knogge, W., 1996. Fungal infection of plants. Plant Cell 8 (10), 1711–1722.
Larsolle, A., Muhammed, H.H., 2007. Measuring crop status using multivariate analysis of hyperspectral field reflectance with application to disease severity and plant density. Precision Agriculture 8 (1), 37–47.
Laudien, R., Bareth, G., Doluschitz, R., 2003. Analysis of hyperspectral field data for detection of sugar beet diseases. In: Proceedings of the EFITA Conference, Debrecen, Hungary, pp. 375–381.
Mahlein, A.-K., Steiner, U., Dehne, H.-W., Oerke, E.-C., 2009. Spectral signatures of diseased sugar beet leaves. In: 7th JIAC Conference, Wageningen, Netherlands, pp. 239–246.
Mahlein, A.-K., Steiner, U., Dehne, H.-W., Oerke, E.-C., 2010. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. Precision Agriculture 11 (4), 413–431.
Mitchell, M.T., 1998. Machine Learning. McGraw Hill.
Moshou, D., Bravo, C., Wahlen, S., West, J., McCartney, A., Baerdemaeker, J., Ramon, H., 2006. Simultaneous identification of plant stresses and diseases in arable crops using proximal optical sensing and self-organising maps. Precision Agriculture 7 (3), 149–164.
Mucherino, A., Papajorgji, P., Paradalos, M.P., 2009. A survey of data mining techniques applied to agriculture. Operational Research 9 (2), 121–140.
Naidu, A.R., Perry, M.E., Pierce, J.F., Mekuria, T., 2009. The potential of spectral reflectance technique for the detection of grapevine leaf-roll associated virus fi 3 in two red-berried wine grape cultivars. Computers and Electronics in Agriculture 66 (1), 38–45.
Nemmour, H., Chibani, Y., 2006. Multiple support vector machines for land cover change detection: an application for mapping urban extensions. ISPRs Journal of Photogrammetry & Remote Sensing 61 (2), 125–133.
Park, J.K., Gromiha, M.M., Horton, P., Suwa, M., 2005. Discrimination of outer membrane proteins using support vector machines. Bioinformatics 21 (23), 4223–4229.
Pelleg, D., Moore, A., 2000. X-means: extending k-means with efficient estimation of the number of clusters. In: Proceedings of the 17th International Conference on Machine Learning, Morgan Kaufmann, pp. 727–734.
Penuelas, J., Baret, F., Filella, I., 1995. Semiempirical indexes to assess carotenoids chlorophyll-a ratio from leaf spectral reflectance. Photosynthetica 31 (2), 221–230.
Quinlan, R.J., 1993. C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA (USA).
Rouse, J.W.J., Haas, H.R., Schell, A.J., Deering, W.D., pp. 309–317 1974. Monitoring vegetation systems in the great plains with ERTS. In: NASA Special Publication 1.
Rumpf, T., Mahlein, A.-K., Dörschlag, D., Plümer, L., 2009. Identification of combined vegetation indices for the early detection of plant diseases. In: Neale, M.C., Maltese, A. (Eds.), Proceedings of the SPIE Conference on Sensing for Agriculture, Ecosystems and Hydrology. vol. 7472. Berlin, Germany.
Ruß, G., 2009. Data mining of agricultural yield data: A comparison of regression models. In: Advances in Data Mining. Applications and Theoretical Aspects. Springer-Verlag, pp. 24–37.
Ruß, G., Kruse, R., Schneider, M., Wagner, P., 2008. Data mining with neural networks for wheat yield prediction. In: Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects. Springer-Verlag, pp. 47–56.
Schölkopf, B., Smola, J.A., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond: Support Vector Machines, Regularization, Optimization, And Beyond, [reprint.] Edition. The MIT Press and MIT Press, Cambridge, MA, USA.
Schölkopf, B., Smola, J.A., Müller, K.-R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10 (5), 1299–1319.
Scholten, J., Klein, M., Steemers, A., de Bruin, G., 2005. Hyperspectral imaging – a novel non-destructive analytical tool in paper and writing durability research. In: 8th International Conference on Non-Destructive Investigations and Microanalysis for the Diagnostics and Conservation of the Cultural and Environmental Heritage, Lecce, Italy.
Steddom, K., Bredehoeft, W.M., Khan, M., Rush, M.C., 2005. Comparison of visual and multispectral radiometric disease evaluations of cercospora leaf spot of sugar beet. Plant Disease 89 (2), 153–158.
Steddom, K., Heidel, G., Jones, D., Rush, M.C., 2003. Remote detection of rhizomania in sugar beets. Phytopathology 93 (6), 720–726.
Steiner, U., Buerling, K., Oerke, E.-C., 2008. Sensor use in plant protection. Gesunde Pflanzen 60 (4), 131–141.
Suykens, A.K.J., 2002. Least Squares Support Vector Machines. World Scientific, Singapore.
Thenkabail, S.P., Smith, B.R., Pauw, E., 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. Remote Sensing of Environment 71 (2), 158–182.
Vapnik, N.V., 1998. Statistical Learning Theory. Wiley, New York.
Vapnik, N.V., 2000. The Nature of Statistical Learning Theory, Statistics for Engineering and Information Science, 2nd Edition. Springer-Verlag, New York.
Wang, X., Zhang, M., Zhu, J., Geng, S., 2008. Spectral prediction of *Phytophthora infestans* infection on tomatoes using artificial neural network (ANN). International Journal of Remote Sensing 29 (6), 1693–1706.
Waske, B., Benediktsson, A.J., 2007. Fusion of support vector machines for classification of multisensor data. IEEE Transactions on Geoscience and Remote Sensing 45, 3858–3866.
West, S.J., Bravo, C., Oberti, R., Lemaire, D.D., Moshou, M.H., 2003. The potential of optical canopy measurement for targeted control of field crop diseases. Annual Review of Phytopathology 41 (1), 593–614.
Wolf, F.J.P., Verreet, A.J., 2002. An integrated pest management system in Germany for the control of fungal leaf diseases in sugar beet: The IPM sugar beet model. Plant Disease 86 (4), 336–344.
Wu, D., Feng, L., Zhang, C., He, Y., 2008a. Early detection of *Botrytis cinerea* on eggplant leaves based on visible and near-infrared spectroscopy. Transactions of the ASABE 51 (3), 1133–1139.
Wu, X., Kumar, V., Quinlan, R.J., Ghosh, Joydeep, Y.Q., Motoda, H., McLachlan, J., Geofirey, N.A., Liu, B., Yu, S.P., Zhou, Z.-H., Steinbach, M., Hand, J.D., Steinberg, D., 2008b. Top 10 algorithms in data mining. Knowledge and Information Systems 14 (1), 1–37.
Zhang, M., Liu, X., O'Neill, M., 2002. Spectral discrimination of *Phytophthora infestans* infection on tomatoes based on principal component and cluster analyses. International Journal of Remote Sensing 23 (6), 1095–1107.

A.2 Sequential support vector machine classification for small-grain weed species discrimination with special regard to *Cirsium arvense* and *Gallium aparine*

Rumpf, T., Römer, C., Weis, M., Sökefeld, M., Gerhards, R., Plümer, L., 2012. Sequential support vector machine classification for small-grain weed species discrimination with special regard to *Cirsium arvense* and *Gallium aparine*. Computers and Electronics in Agriculture 80, 89–96.

Abstract

Site-specific weed management can reduce the amount of herbicides compared to classical broadcast applications. The ability to apply herbicides on weed patches within the field requires automation. In this work the focus is on the automatic detection of different species with imaging sensors. Image processing algorithms determine shape features for the plants in the images. With these shape descriptions classification algorithms can be trained to identify the weed and crop species. Since the weeds differ in the economic loss due to their yield effect and are controlled by different herbicides, it is necessary to correctly distinguish between the species. Image series of different measurements with plant samples of different growth stages were analysed. For classification a sequential classification approach was chosen, involving three different support vector machine (SVM) models. In a first step groups of similar plant species were successfully identified (monocotyledons, dicotyledons and barley). To differentiate further between the dicotyledons, the species in this group were subject to a second and third classification step. For each classification step different, best suited features were selected by SVM-Weighting and RELIEF-F algorithm. This way the overall classification approach was focused on the problem and adapted to identify the most important species, *Cirsium arvense* and *Galium aparine*, with higher accuracy than using a non-sequential classification approach. Despite an overall classification accuracy of 97.7%, the group of dicotyledons was more difficult to separate. With the two subsequent classifiers correct classification rates of 80% and more were achieved for *Cirsium arvense* and *Galium aparine*.

Computers and Electronics in Agriculture 80 (2012) 89–96

Contents lists available at SciVerse ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Sequential support vector machine classification for small-grain weed species discrimination with special regard to *Cirsium arvense* and *Galium aparine*

Till Rumpf [a,*], Christoph Römer [a], Martin Weis [b], Markus Sökefeld [b], Roland Gerhards [b], Lutz Plümer [a]

[a] Institute of Geodesy and Geoinformation, Department of Geoinformation, University of Bonn, Germany
[b] Institute of Phytomedicine (360), Department of Weed Science, University of Hohenheim, Germany

ARTICLE INFO

Article history:
Received 7 June 2011
Received in revised form 24 October 2011
Accepted 28 October 2011

Keywords:
Early weed detection
Cirsium arvense
Galium aparine
Feature selection
Sequential classification
Support vector machines

ABSTRACT

Site-specific weed management can reduce the amount of herbicides used in comparison to classical broadcast applications. The ability to apply herbicides on weed patches within the field requires automation. This study focuses on the automatic detection of different species with imaging sensors. Image processing algorithms determine shape features for the plants in the images. With these shape descriptions classification algorithms can be trained to identify the weed and crop species. Since weeds differ in their economic loss due to their yield effect and are controlled by different herbicides, it is necessary to correctly distinguish between the species. Image series of different measurements with plant samples at different growth stages were analysed. For the classification a sequential classification approach was chosen, involving three different support vector machine (SVM) models. In a first step groups of similar plant species were successfully identified (monocotyledons, dicotyledons and barley). Distinctions within the class of dicotyledons proved to be particularly difficult. For that purpose species in this group were subject to a second and third classification step. For each of these steps different features were found to be most important. Feature weighting was done with the RELIEF-F algorithm and SVM-Weighting. The focus was on the early identification of the two most harmful species *Cirsium arvense* and *Galium aparine*, with optimal accuracy than using a non-sequential classification approach. An overall classification accuracy of 97.7% was achieved in the first step. For the two subsequent classifiers accuracy rates of 80% and more were obtained for *C. arvense* and *G. aparine*.

1. Introduction

Weed populations have been found to be distributed heterogeneously within agricultural fields (Marshall, 1998; Johnson et al., 1996; Christensen and Heisel, 1998; Gerhards and Christensen, 2003; Christensen et al., 2009). Due to the lack of automatic weed detection techniques and site-specific herbicide application, the majority of farmers spray herbicide uniformly across the field. An exact herbicide application to the weeds within a weed patch requires not only detailed information on the weed density but also on weed species distribution. Based on this information and the application of economic weed thresholds selective herbicides can be sprayed site-specifically. Gerhards and Oebel (2006) realised herbicide savings in cereals, maize and sugar beet field from 6% to 81% with site-specific herbicide application based on weed species distribution maps. A site-specific application of a mixture of the individual herbicides on the same field achieved only savings of 19% compared to a uniform treatment of the whole field (Gerhards and Sökefeld, 2003).

According to the demand of Christensen et al. (2009) the identification of single weed species a species discrimination of dicotyledons is necessary. A discrimination of species yields information on weed distribution and weed species composition, laying the foundations for the site-specific application of selective herbicides. A suitable application technology, which allows a simultaneous application of several herbicidal agents on-the-go, based on sensor signals or weed distribution maps, is a further requirement for the adoption of site-specific and selective herbicide application. Possible technical solutions for patch spraying with several herbicides were outlined by Schulze-Lammers and Vondricka (2010) and kefeld (2010).

A major step towards a practical solution for site-specific weed management is the development of precise and powerful data acquisition techniques to automatically and continuously determine in-field variation of weed populations. The most promising techniques to identify weed species in arable crops are based on image processing (Weis and Sökefeld, 2010). Infrared, multispectral and RGB (red, green, blue channel) cameras were used to take pictures of crops and weed species from a low distance above the ground. Plant properties were then extracted from the images by image processing algorithms. Those properties were computed as features, which were used to separate species from each other.

* Corresponding author. Tel.: +49 228 736335.
E-mail address: rumpf@igg.uni-bonn.de (T. Rumpf).

0168-1699/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2011.10.018

Åstrand and Baerveldt (2004) combined colour and shape features, which were derived after an initial plant segmentation step. The colour features were computed as standard deviation and mean of the RGB values. Shape features were computed as area of segment, form factors (distance variance to centre of gravity, compactness and moments). Segments were merged if the distance between them was small. In addition to the colour and shape features a row distance measure was introduced to locate the position 'in-row' or 'between-row'. Classification was finally done with a Bayesian approach. Burks et al. (2000) computed 33 unique colour texture features (co-occurence) from a hue, saturation and intensity (HSI) representation to distinguish between five weed species and the soil. Different neural network classifiers were tested for their performance based on the data given in Burks et al. (2005), resulting in a backpropagation training algorithm with high classification success. Blasco et al. (2002) implemented a vision system for the detection of weeds in a lettuce crop. Segmentation into plant/soil was based on a Bayes-classifier for RGB colours and size features were used to separate weeds from lettuce plants. Cho et al. (2002) used a discriminant function for shape feature selection and neural networks to identify weeds in a radish crop. Tellaeche et al. (2011) successfully applied Support Vector Machines to identify weeds between crop rows based on weed and crop cover measures for image parts. Zhu and Zhu (2009) present a Support Vector Machines approach for weed detection based on shape and texture features for single leaves.

Still, there was no approach for reliable and robust classification between weed species yet. This, however, is a necessary prerequisite for site-specific weed management.

The objective of this paper was the automatic classification of crop and the four weed classes *Galium aparine*, *Cirsium arvense*, other dicotyledonous weeds and monocotyledonous weeds using image processing and classification. Diverse economic thresholds for these weed classes and the availability of selective herbicides against these weeds are the reason for this separation. Gerowitt and Heitefuss (1990) determined weed thresholds for *G. aparine* between 0.1 and 0.5 plants m^{-2}, 40–50 plants m^{-2} for dicotyledonous weeds in total and 20–30 plants m^{-2} for grass weeds. For the very competitive weed species *C. arvense* Börner (1995) indicated a threshold of 2 plants m^{-2}.

The approach in this study is based on an imaging system capturing two images of different wavelengths to differentiate plants from background. The plants are then extracted with image processing algorithms and classified according to their shape. The shape description is expressed as shape features. While the differences in shape features between crop and weed were rather large, features of weeds were highly similar and therefore specific weeds are hard to identify. In addition different growth stages amplify the problem, especially during two-leaf stage. For instance, the dicotyledons *G. aparine* and *Veronica persica* appear to be nearly identical, but due to the high economic loss caused by *G. aparine* accurate classification is critical. This means that very specific features and classifiers are needed to solve this problem. This, however, is not possible in a single multi-classification approach with all weed species and crops. Hence, a sequential classification approach was developed. The main idea is to separate between similar subgroups of the dataset, like crop, monocotyledons and dicotyledons, which are well separable. In next steps dicotyledons were identified by features specialised for the current subgroup. Relevant features were determined by SVM-Weighting (Guyon et al., 2002) or RELIEF-F (Kononenko et al., 1994). Depending on the current task for classification linear and non-linear Support Vector Machines (SVMs) were used. This way it is possible to differentiate weed species of similar appearance.

The image analysis system in combination with automatic algorithms for plant species discrimination can be included into real-time and map-based approaches for site-specific weed control.

2. Materials and methods

2.1. Data acquisition

Images were taken from greenhouse series and in the field: weed and crop species were grown in pots in the years 2006 and 2008, field data were acquired in maize (2008), winter wheat (2007) and sugar beet crops (2007). Images of the red (R, ca. 580 nm) and infrared (IR, >720 nm) spectrum of the light were taken simultaneously and subtracted from each other (IR-R) to generate difference images (Sökefeld et al., 2007). Most of the red light is absorbed by plants for the photosynthesis, whereas the infrared light is reflected. All other materials in field (soil, mulch, stones) have a similar reflection at both wavelengths. Plant material therefore appears bright in the difference images due to the typical 'red edge' in the reflectance spectrum.

The data set consisted of samples for 10 species, which were to be differentiated: two monocotyledonous, seven dicotyledonous weed species and summer barley were chosen because of their relevance for weed management. All species were in early growth stages, ranging from germination to two-leaf stage, only *C. arvense* appeared with up to five leaves. The different growth stages of each weed and crop (*Hordeum vulgare*, summer barley) were merged, because they have no economic relevance in the management practice, since the management thresholds are set according to the number of plants per m^2.

2.2. Segmentation and feature extraction

The acquired difference images are converted using image processing techniques: a grey value threshold is used to separate plants and background, resulting in binary images with two values, one for the foreground (plant) and the other for the background. In the binary images objects are identified by segmentation of connected foreground components. These segments correspond to plants or parts thereof, if single leaves are not connected after the thresholding step. Fig. 1 shows the resulting objects after binarisation and segmentation for some of the training samples of this study.

Overlapping plants in this step lead to complex objects, containing parts of different plants. These objects would be difficult to separate into their components and even then a proper shape description is likely to fail for the following analysis. Therefore they are put into classes for overlapped plants and as such handled parallel to the other species classes in the system. Overlaps can also be identified according to their shape description, as they lead to large objects and their complexity expresses itself in some of the shape features. Since the monocotyledonous species with their long leaves naturally tend to overlap, these objects are usually assigned to a monocotyledonous class, which especially for *H. vulgare* crops leads to a valid decision. In Fig. 1 some overlapping of *H. vulgare* plants (HORVS) can be seen. The general assumption for the application of this approach is, that the plants are measured in early development stages, where overlapping does not affect the overall sampling accuracy. The shape of single plants cannot be extracted from cluttered scenes, limiting the analysis to early growth stages. The most important herbicide applications, which can benefit from this technology, take place shortly after germination, the problems due to overlapping are limited during this period.

To identify different weeds and crop species, shape parameters were computed for the objects in the image. Some of the shape parameters were derived from the set of pixels belonging to an object, like areasize, inertia values according to main axes of the different object, central moments and moment invariants. Other features were derived from the border representation, like border

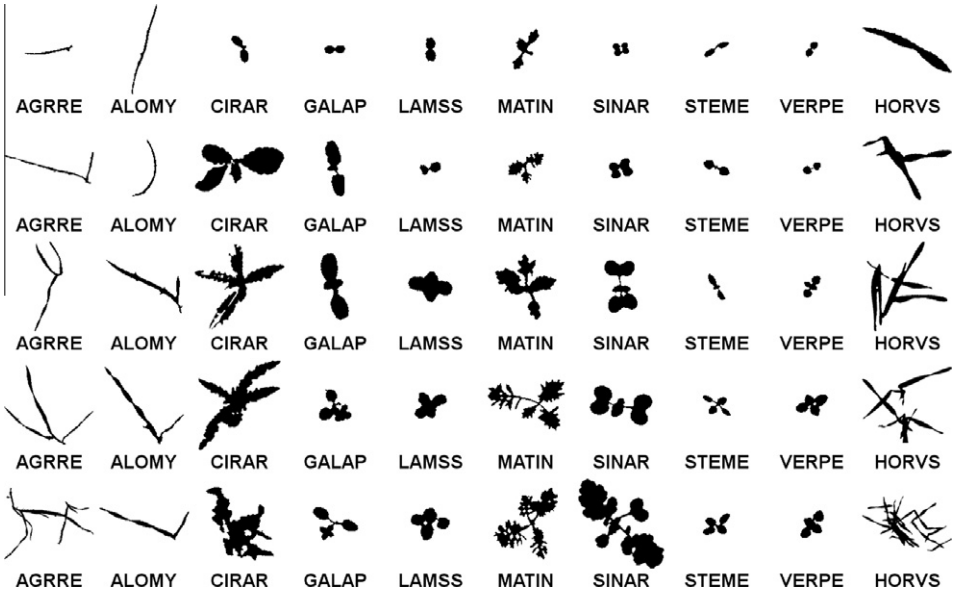


Fig. 1. Samples of the training data, sorted by class assignment. The segments were scaled to a common maximum width for better visual comparison of the shape. The species are given by their EPPO-Codes: Monocotyledounous: AGRRE: *Agropyron repens* ALOMY: *Alopecurus myosuroides* Dicotyledonous: CIRAR: *C. arvense* GALAP: *Galium aparine* LAMSS: *Lamium* sp. MATIN: *Matricaria inodora* SINAR: *Sinapis arvensis* STEME: *Stellaria media* VERPE: *Veronica persica* Crop: HORVS: *Hordeum vulgare*.

length and Fourier features. A combination of these feature representations led to the features *compactness* and distance features of the border to either the centre of gravity of the region (*rmean, rmax*) or to the main inertia axes (*drear*). To derive further features, a distance transform was applied to the region, resulting in a distance value (to the border) for each pixel. A skeletonisation step additionally created skeletons of the objects, which are the central lines of a region, defined by the local maxima of the distance values (Weis et al., 2007). Combining the skeleton lines with the distance transform, the distance values for all skeleton pixels were assembled and statistical measures were computed (number of values: *skelsize*, maximum, mean: *skelmean*, variance). These measures describe the overall thickness of an object: typically monocotyledons leaves have a larger skeleton and the skeleton does not have a large distance to the border lines. In contrast to these dicotyledons with more compact leaves have a shorter skeleton representation with bigger distances to the border. Finally, the values of each feature had to be normalized to a mean of zero and a standard deviation of one. The normalisation avoids numerical problems introduced by different scales.

After the feature weighting steps, which were conducted during the analysis the following features proved to be the most relevant and are therefore named here:

- *areasize*, the number of pixels of an object;
- *rmean* and *rmax*, the mean and maximum distance of the border to the centre of gravity;
- *drear*, vertical distance of the border to the main axis of the object;
- *eccentricity* and *compactness*, describing the overall elongatedness and computed as the ratio of the (squared) border length to the area (Burger and Burge, 2009);
- *hu1* and *hu2*, the first two moment invariants as given by Hu (1962);
- *skelsize* and *skelmean*, two features derived from the skeletonisation step, denoting the length of the skeleton (measured in pixel) and the mean distance to the border.

Segments of the data sets were selected for the training and classes were assigned to them, resulting in a large training data set of plants grown under various conditions. Since the appearance of the plants changes during the growth process, several growth stages of each species were distinguished during a training step. The training was done by manual assignment of classes to visually selected segments and their feature vectors, which contain the shape parameters. Samples of the training data set are shown in Fig. 1. Obviously the shape variation within each class can be high (top–down) and weeds of each group in early growth stages have similar appearance (left–right).

2.3. Classification with support vector machines

The simple case of linear SVMs assumes that two classes (dichotomous problem) are linearly separable, i.e. their discriminant can be described by a linear function. Consider a set of samples $(\vec{x}_1, y_1) \dots (\vec{x}_l, y_l), l \in \mathbb{N}$. Each sample has a vector $\vec{x}_i \in \mathbb{R}^m$ consisting of m known features describing it. The label $y_i \in \pm 1$ indicates to which class it belongs and usually cannot be directly observed. Each sample is generated by an unknown, underlying probability distribution $P(\vec{x}, y) = P(\vec{x})P(y|\vec{x})$. The task at hand for a binary classifier is now to decide to which class an unknown sample belongs, based on the observed feature vector $\vec{x}$.

To achieve this a training data set of $m < l$ samples is needed where both the feature vectors and the labels are observed for each sample. The classifier aims at finding an indicator function $f : \vec{x} \to \{\pm 1\}$ which approximates the unknown $P(\vec{x}, y)$ and gives the best prediction of the class y of a new sample $\vec{x}$ which was not part of the training data set.

Without a restriction of the set of functions, however, a function f which does well on the training data might not generalize well on unseen data. This risk of "overfitting" occurs especially if the indicator function f may be arbitrarily complex.

Hence, Support Vector Machines (SVMs) (Vapnik, 2000; Schölkopf et al., 1998) restrict their set of indicator functions to separating hyperplanes of the form

$$f(\vec{x}) = \mathrm{sgn}(\langle \vec{\omega}, \vec{x} \rangle + b), \tag{1}$$

where the c denotes the orientation of the hyperplane and $b \in \mathbb{R}$ the offset from the origin. For a set of linearly separable data, however,

there are many different hyperplanes and the concept of finding a separating hyperplane is not exclusively used by Support Vector Machines. SVMs are special about the specific selection on the separating hyperplane. SVMs adopt the hyperplane which maximises the distance to the closest samples of both classes in the training data, i.e. the hyperplane with the maximal margin. As Vapnik (1998) proves, this choice of the hyperplane minimised the risk of overfitting.

To get the optimal separating hyperplane with maximal margin, we have to minimise

$$h(\vec{\omega}) = \frac{1}{2}|\vec{\omega}|^2, \tag{2}$$

$$\text{subject to } y_i(\langle\vec{\omega},\vec{x}\rangle + b) \geqslant 1, \forall i. \tag{3}$$

Note that only those few samples closest to the hyperplane determine the position of the hyperplane. Those samples are called support vectors.

To solve the constrained minimisation problem, it is often easier to solve the dual maximisation problem (Boyd and Vandenberghe, 2004). Using the Karush Kuhn Tucker conditions (Rockafellar, 1993) the dual Lagrangian is formed

$$\underset{\vec{\alpha}}{\text{maximise}}\, W(\vec{\alpha}) = \sum_i^m \alpha_i - \frac{1}{2}\sum_i^m\sum_j^m \alpha_i\alpha_j y_i y_j \langle\vec{x}_i,\vec{x}_j\rangle, \tag{4}$$

$$\text{subject to } \alpha_i \geqslant 0 \tag{5}$$

$$\text{and } \sum_i^m \alpha_i y_i = 0, \tag{6}$$

where α denotes the Lagrangian variable. One key feature of the Lagrangian is that every vector $\vec{x}_i$ which is no support vector becomes zero in the optimisation process and has no weight in the construction of the optimal hyperplane

$$\hat{\vec{\omega}} = \sum_i^m \alpha_i y_i \vec{x}_i, \tag{7}$$

$$\hat{b} = -\frac{1}{2}\left\langle\hat{\vec{\omega}},\vec{x}_a + \vec{x}_b\right\rangle, \tag{8}$$

where $\vec{x}_a \in \{+1\}$ and $\vec{x}_b \in \{-1\}$ are any support vectors of both classes.

2.3.1. Nonlinear support vector machines

Until now linear separability was assumed, though this is indeed a rather special case. As every dot product can be replaced by a kernel of the form $k(\vec{x}_1,\vec{x}_2)$ (Schölkopf and Smola, 2002), using the dual expression in Eq. (4) enables to use the so called kernel trick and modify Eq. (4) to

$$\underset{\vec{\alpha}}{\text{maximise}}\, W(\vec{\alpha}) = \sum_i^m \alpha_i - \frac{1}{2}\sum_i^m\sum_j^m \alpha_i\alpha_j y_i y_j \mathrm{k}(\vec{x}_i,\vec{x}_j). \tag{9}$$

In our approach, we use a very popular kernel, namely Gaussian radial basis functions (RBF). The idea is similar to radial basis networks, though instead of placing the radial basis functions into the centres of both classes, they are placed at each support vector, i.e. those samples which are critical for the classification task (Schölkopf et al., 1997). The RBF kernel used is

$$\mathrm{k}(\vec{x}_1,\vec{x}_2) = \exp\left(-\frac{|\vec{x}_1 - \vec{x}_2|^2}{2\sigma}\right). \tag{10}$$

2.3.2. Soft margin support vector machine

In order to improve the generalisation ability of the SVM classifier, the risk of outliers, noise and wrongly labelled training samples have to be taken into account. To achieve this, slack variables are introduced which penalise classification errors. This modifies the optimisation problem to

$$\underset{\vec{\omega}}{\text{maximise}}\, h(\vec{\omega})\frac{1}{2}|\vec{\omega}|^2 + C\sum_{i=1}^m \xi_i, \tag{11}$$

$$\text{subject to } y_i(\langle\vec{\omega},\vec{x}\rangle + b) \geqslant 1 - \xi_i, \forall i \tag{12}$$

$$\text{and } \xi_i \geqslant 0, \forall i. \tag{13}$$

The parameter C handles the tradeoff between margin maximisation and training error minimisation (Schölkopf and Smola, 2002). Interestingly, apart from an additional boundary on constraint 5 to $C \geqslant \alpha_i \geqslant 0$, the general form of the dual problem does not change (Cortes and Vapnik, 1995).

2.4. Feature weighting

It is unlikely that all 50 shape parameters are equally important for class separation. This is important as many bad features may overweight a small number of good ones. Out of many possible feature weighting methods two approaches were chosen. For linear classification SVM-Weighting (Guyon et al., 2002) was used and for non-linear classification tasks the RELIEF-F algorithm (Kira et al., 1992) proved to be suitable.

2.4.1. SVM-Weighting

As stated above SVMs are 'multivariate' classifiers, which means they are able to handle multiple features in training simultaneously. The weight vector $\vec{\omega}$ (Eq. (7)), which is a linear combination of training patterns, specifies the weight of every feature in the construction of the optimal hyperplane. Features where separation is good have a large corresponding coefficient in $\vec{\omega}$ due to Eq. (7). These weights can also be used to rank features (Guyon et al., 2002).

2.4.2. RELIEF

First, a dichotomous case with non-linear classification was considered. The key idea of the RELIEF algorithm is to estimate features according to how well their values distinguish among samples that are near to each other (Kononenko et al., 1994). RELIEF searches for two nearest neighbours for a given sample. Each neighbourhood consists of k elements. For a given k, 'hit' is the set of k nearest neighbours of the same class and 'miss' from the different class. The relevance of a feature F_j is determined by the sum of the euclidean distances between nearest misses M_l and nearest hits H_l for all samples used to approximate probabilities.

Listing 1: Pseudo code of the RELIEF-F algorithm for two class classification.

```
1   INPUT: A set of features F = F_1, ..., F_m, a set of
2     samples R_1, ..., R_n and a class label is given for each R
3   OUTPUT: A set of feature weights W = W_1, ..., W_m
4   set all weights W(F) := 0
5   for i := 1 to n do (number of samples for approximating
    probabilities)
6     begin
7       randomly select an sample R_i:
8       find k nearest hits H_l and nearest misses M_l;
9       for j := 1 to m do (all features)
10          begin
11          W(F_j) := W(F_j) - sum_{l=1}^{k} difference(F_j, R_i, H_l)/(m · k)
                              + sum_{l=1}^{k} difference(F_j, R_i, M_l)/(m · k);
12        end
13    end;
```

For multi-class problems the extension RELIEF-F Kononenko et al. (1994) can be used. Instead of finding the nearest miss M from the different class, the algorithm finds one nearest miss of each different class and averages their contribution for updating the weight $W(F_j)$. The average is weighted with the prior probability of each class.

3. Results

3.1. One-against-all classification

In a first approach a classifier to differentiate all classes in one step was applied. Due to the number of 10 classes and difficulties of class separation, especially dicotyledons, a non-linear SVM was used. Features were weighted according to the RELIEF-F algorithm. This way, features with good class separability properties got more weight in classification.

After the application of a non-linear SVM with RBF kernel in order to classify between all classes only an overall accuracy of 69.25% was achieved (Table 1). The classification result was not sufficient for our task, however, we learned which classes were well separable and which classes were often mixed up. *H. vulgare* separation with a class recall of 82.98% and the two dicotyledons *C. arvense* and *Sinapis arvensis* with a class recall over 77% were adequately identified. While the other weeds were systematically misclassified.

First, misclassification occurred in the group of monocotyledons *Agropyron repens* and *Alopecurus myosuroides* (highlighted dark grey), whereas the separation of *H. vulgare*, monocotyledons and dicotyledons was excellent. Second, within the dicotyledons also symmetric misclassifications could be observed. Especially two groups of weeds were interchanged, viz. the first group *G. aparine*, *Lamium* sp. and *V. persica* (highlighted black) and the second group *Matricaria inodora* and *Stellaria media* (highlighed light grey). The group of misclassifications within dicotyledons is of peculiar interest, because *G. aparine* is a weed with low economic threshold which accordingly has to be identified with high accuracy.

In order to improve the classification result specific models for special groups of weeds had to be learned. The classification problem have sequentially been divided into several separation models. Thus, in each classification particularly suitable features could be weighted higher and parameters of each SVM could individually adapted.

The advantage of this approach is that by reduction of class numbers more specific features could be weighted higher for this particular classification. As the used non-linear kernel is based on radial basis functions, it naturally depends on the density of the data points, especially near the margin. The more classes are used, the higher is the risk that their densities differ significantly. This can be expected here as some classes (e.g. *G. aparine*, *A. repens* and *A. myosuroides*) are very dissimilar in feature space and others (e.g. *G. aparine*, *V. persica* and *Lamium* sp.) are very close. Due to the variance in density within the different classes, the classification problem had to be sequentially divided into robustly separable groups of weeds. The features skelsize and areasize were the most important to separate the crop *H. vulgare* from the classes monocotyledons and dicotyledons, but were not suitable to separate the different classes of dicotyledons. Other features have more relevance for the separation of the dicotyledonous species.

First, the crop *H. vulgare* and the groups of monocotyledons and dicotyledons were clearly classified. Second, sequentially specific classifiers for the more difficult classification of the dicotyledons were applied. This approach is introduced below.

3.2. Sequential classification

The aim of sequential classification was the identification of weeds with high economic effect. The classification task was divided into three steps. In a first step the three well separable groups *H. vulgare*, monocotyledons and dicotyledons, as shown in Table 1, were differentiated. Using SVM-Weighting the features were weighted by relevance for this classification step. The most relevant features separating the three groups were skelsize, areasize, rmean and rmax. The overall classification accuracy using a linear SVM was 97.74% (Table 2). The class recall of the dicotyledons which have the highest economic effect even amount 99.24%. Further, only a classification between several dicotyledons was necessary.

Compared to SVM weighting in the first classification we applied RELIEF-F to weight the features in the second non-linear classification by SVM with RBF-kernel. The two features most weighted by RELIEF-F were eccentricity and hu1. A second group of features, namely skelmean, drear and hu2 were almost as relevant. In comparison to the first classification step the most relevant features were different. The second classification step separated the dicotyledons with an accuracy of 69.37% (Table 3). Regarding the classification result in more detail four classes were detected with a high accuracy of above 70%, whereas the classification result of *C. arvense*, the weed with highest economic effect, was even above 82.71%. Symmetric misclassifications, however, still occurred. The identification of the second weed with high economic effect, namely *G. aparine*, was still unsatisfactory with an accuracy of 64.00%. The two classes by which *G. aparine* was misclassified are *Lamium amplexicaule* and *Lamium purpureum* (merged as LAMSS) and *V. persica* (highlighted grey).

Thus, in a third step we learned a non-linear SVM classifier with RBF-kernel only for these three dicotyledons. The most relevant

Table 1
Results of the one-against-all non-linear SVM classification in one step with weighted features by RELIEF-F. The equal coloured cells highlighted the groups of weeds where common misclassifications occurred.

Prediction	Ground truth										Class precision (%)
	AGRRE	ALOMY	CIRAR	GALAP	HORVS	LAMSS	MATIN	SINAR	STEME	VERPE	
AGRRE	57	21	0	0	1	0	0	0	0	0	72.15
ALOMY	30	63	0	0	1	0	0	0	1	0	66.32
CIRAR	0	0	103	7	3	1	3	8	1	1	81.10
GALAP	0	0	10	87	5	19	6	1	7	10	60.00
HORVS	1	2	2	1	78	0	4	0	0	0	88.64
LAMSS	0	0	2	16	1	84	4	5	4	15	64.12
MATIN	0	1	1	4	4	5	56	3	12	5	61.54
SINAR	0	0	13	0	0	2	2	74	4	2	76.29
STEME	0	1	2	2	1	4	11	1	58	6	67.44
VERPE	0	0	0	8	0	28	3	3	5	74	61.16
Class recall (%)	64.77	71.59	77.44	69.60	82.98	58.74	62.92	77.89	63.04	65.49	69.25

Table 2
Results of the first step of sequential classification with linear SVM.

Prediction	Ground truth			Class precision (%)
	Monocotyledons	Dicotyledons	HORVS	
Monocotyledons	171	3	3	96.61
Dicotyledons	2	784	10	98.49
HORVS	3	3	81	93.10
Class recall (%)	97.16	99.24	86.17	97.74

Table 3
Results of the second step of sequential classification for the dicotyledonous weed species with non-linear SVM. The equal coloured cells highlighted the group of weeds where common misclassifications occurred.

Prediction	Ground truth							Class precision (%)
	CIRAR	GALAP	LAMSS	MATIN	SINAR	STEME	VERPE	
CIRAR	110	7	2	4	9	0	2	82.09
GALAP	8	80	22	6	0	5	9	61.54
LAMSS	2	19	94	5	4	3	26	61.44
MATIN	2	6	2	63	1	15	5	67.02
SINAR	10	0	1	1	76	3	4	80.00
STEME	1	4	3	9	2	65	7	71.43
VERPE	0	9	19	1	3	1	60	64.52
Class recall (%)	82.71	64.00	65.73	70.79	80.00	70.65	53.10	69.37

Table 4
Results of the third step of sequential classification with non-linear SVM.

Prediction	Ground truth			Class precision (%)
	GALAP	VERPE	LAMSS	
GALAP	100	11	21	75.76
VERPE	11	86	17	75.44
LAMSS	14	16	105	77.78
Class recall (%)	80.00	76.11	73.43	76.36

features for this classifier were rmin, hu2 and compactness. The average classification performance of these three dicotyledons increased considerably to 76.36% (Table 4). The weed *G. aparine* with high economic effect was reliable perceived with class recall of 80%.

The sequential classification steps and achieved classification accuracies are summarised in Table 5.

4. Discussion

Automatic weed detection is a prerequisite for accurate and profitable site-specific weed control. Furthermore, not only differentiation between crop and weed, but especially between different weeds, is of high interest. This is important, as the economic threshold varies highly between different weed species. Therefore an unidentified weed which cause high economic loss, and hence wrongly applied herbicides, may be fatal for yield. For instance, Pallutt and Flatter (1998) calculated yield losses in winter barley of 10–30 kg/ha for one *G. aparine* plants m^{-2}, whereas weeds with lower competitiveness like *Lamium* sp. caused losses between 1 and 2 kg/ha per plants m^{-2}.

The focus of this study is to discriminate between different weed groups, single weed species and crop plants of various image series taken in the greenhouse and under field conditions. This means that each class consists of individuals from images of diverse growth stages and origin, resulting in a very demanding classification task due to highly similar morphology of different weeds. Despite the high class variance an optimal classification scheme that can distinguish between several weed species and the crop was developed.

Due to the fact that the classification problem involves separation of a high number of classes and the separability of some species with similar morphology is difficult in one classification step, the problem was divided into several parts of classification. Well separable groups of weed species were identified by a standard SVM with RBF kernel, which was used to separate the different weed classes simultaneously in one step (Table 1). The classification results turned out to be unsuitable for herbicide specific weed management. Especially weed species with a high negative impact on crop yield were often misclassified.

In order to achieve an adequate classification accuracy, a sequential classification approach was necessary. One key to adequate classification results is the weighting of relevant features. For instance, we have great morphologic differences between certain weed species such as *C. arvense* and *V. persica* on the one hand, and very similar shapes like *G. aparine* and *Lamium* sp. on the other hand. Naturally, the optimal features to describe the differences between named classes are not the same. Hence, features highly

Table 5
The weed species for the three sequential classification steps from top to bottom with the achieved classification accuracies.

	HORVS	AGRRE	ALOMY	CIRAR	MATIN	SINAR	STEME	GALAP	VERPE	LAMSS
Step 1	Crop (%) 86.2	Monocotyledons (%) 97.2		Dicotyledons (%) 99.2						
Step 2				82.7	70.8	80.0	70.7	88.7		
Step 3								80.0	76.1	73.4

weighted for an SVM classification of all plants in a one-step approach are not the same features which are best suited for a separation between two or three species. Therefore a sequential classification approach first separates between dissimilar subgroups of weeds, and then step by step specializes the classifier into separating the resulting subgroups of weeds with specialised features and particular kernel parameters.

Consequently, in the first classification step groups of weed species (monocotyledons, dicotyledons and crop) was separated with high accuracy of more than 97% by using a linear SVM combined with SVM-Weighting (Table 2). Párez et al. (2000) reduced the weed-crop classification to a binary problem and achieved classification rates using pattern recognition from almost 90% for the crop and between 75% and 80% for the weed depending on the classification process.

The subdivision of the dicotyledons in the classes of *C. arvense* and *G. aparine* during the second classification step follows their extremely diverse economic thresholds in comparison to the other dicotyledons (Gerowitt and Heitefuss, 1990; Börner, 1995). Furthermore selective herbicides are available for these weed species as well as for monocotyledons. The high weighted shape features for the first classification step are not suitable to distinguish the dicotyledons among each other. Moreover, discrimination between the dicotyledons is a non-linear task and therefore the RBF kernel was chosen for the SVM combined with the feature weighting algorithm RELIEF-F. Misclassifications between *V. persica*, *Lamium* sp. and *G. aparine*, however, still occurred.

This is the reason for the third classification step to detect *G. aparine* with high accuracy, such that the classification accuracy reached 80% for this class. Gerhards and Oebel (2006) used an image analysis approach for weed detection with a similar subdivision of the weeds. In comparison to this approach the sequential classification increased the results for *G. aparine* to 80% versus 72% and for monocotyledons to 97% versus 74%. The realised classification results by means of sequential classification with SVM (Table 5), which are presented in this paper, are a solid basis concerning differentiation of weed species or rather groups of weed species for site-specific weed control. This kind of classification not only allows the differentiation between crops and weeds which is a crucial requirement for the site-specific application of broad spectrum herbicides or herbicide mixtures but it is also capable for the discrimination within the weeds dividing them in monocotyledons and dicotyledons.

5. Conclusion

The conducted work shows that an early detection and classification of different weed species using image processing and classification based on shape features is feasible. To focus the approach on the identification of weeds with high economic effects, in particular *C. arvense* and *G. aparine*, a sequential classification was implemented. The presented approach divides the overall complexity of the classification problem into less complex parts, thus improving the classification accuracy and detection rates. The results of the classification can be used by decision support systems for site-specific weed management. This will lead to an improvement of herbicide saving in terms of cost reduction and minimisation of environmental impact.

References

Åstrand, B., Baerveldt, A.-J., 2004. Plant recognition and localization using context information. In: Proceedings of Mechatronics and Robotics 2004 (MechRob2004). Sascha Eysoldt Verlag, Aachen, Germany, Special Session Autonomous Machines in Agriculture, pp. 1191–1196.

Blasco, J., Aleixos, N., Roger, J.M., Rabatel, G., Moltó, E., 2002. Ae-automation and emerging technologies, Robotic weed control using machine vision. Biosystems Engineering 83, 149–157.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, New York, NY, USA.

Burger, W., Burge, M.J., 2009. Principles of Digital Image Processing Core Algorithms Undergraduate Topics in Computer Science, 1st edition. Springer.

Burks, T., Shearer, S., Heath, J., Donohue, K., 2005. Evaluation of neural-network classifiers for weed species discrimination. Biosystems Engineering 91 (3), 293–304.

Burks, T.F., Shearer, S.A., Payne, F.A., 2000. Classification of weed species using color texture features and discriminant analysis. In: Transactions of the ASAE, vol. 43. American Society of Agricultural Engineers, pp. 441–448.

Börner, H., 1995. Unkrautbekämpfung. Gustav Fischer, Jena.

Cho, S.I., Lee, D.S., Jeong, J.Y., 2002. Weed–plant discrimination by machine vision and artificial neural network. Biosystems Engineering 83, 275–280.

Christensen, S., Heisel, T., 1998. Patch spraying using historical, manual and real-time monitoring of weeds in cereals. Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz Sonderheft XVI, 257–263.

Christensen, S., Søgaard, H., Kudsk, P., Nørremark, M., Lund, I., Nadimi, E., Jørgensen, R., 2009. Site-specific weed control technologies. Weed Research 49 (3), 233–241.

Cortes, C., Vapnik, N.V., 1995. Support-vector networks. Machine Learning 20 (3), 273–297.

Gerhards, R., Christensen, S., 2003. Real-time weed detection, decision making and patch spraying in maize, sugar beet, winter wheat and winter barley. Weed Research 43 (6), 385–392.

Gerhards, R., Oebel, H., 2006. Practical experiences with a system for site-specific weed control in arable crops using real-time image analysis and GPS-controlled patch spraying. Weed Research 46 (3), 185–193.

Gerhards, R., Sökefeld, M., 2003. Precision farming in weed control – system components and economic benefits. Precision Agriculture 1, 229–234.

Gerowitt, B., Heitefuss, R., 1990. Weed economic thresholds in the F R. Germany. Crop Protection 9, 323–331.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Machine Learning 46, 389–422, 10.1023/A:1012487302797.

Hu, M.K., 1962. Visual pattern recognition by moment invariants. IRE Transactions Information Theory 8 (2), 179–187.

Johnson, G., Mortensen, D., Gotway, C., 1996. Spatial and temporal analysis of weed seedling populations using geostatistics. Weed Science 44, 704–710.

Kira, K., Rendell, L.A., 1992. The feature selection problem: traditional methods and a new algorithm. In: Proceedings of the 10th National Conference on Artificial Intelligence, AAAI'92. AAAI Press, pp. 129–134.

Kononenko, I., 1994. Estimating attributes: analysis and extensions of relief. In: Proceedings of the European Conference on Machine Learning. Springer-Verlag, New York, Inc., Secaucus, NJ, USA, pp. 171–182.

Marshall, E., 1998. Field-scale estimates of grass populations in arable land. Weed Research 28, 191–198.

Pallutt, B., Flatter, A., 1998. Variabilität der Konkurrenz von Unkräutern in Getreide und daraus resultierende Auswirkungen auf die Sicherheit von Schwellenwerten. Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz Sonderheft XVI, 333–344.

Párez, A., López, F., Benlloch, J., Christensen, S., 2000. Colour and shape analysis techniques for weed detection in cereal fields. Computers and Electronics in Agriculture 25, 197–212.

Rockafellar, R.T., 1993. Lagrange multipliers and optimality. SIAM Review 35 (2), 183–238.

Schölkopf, B., Smola, J.A., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond: Support Vector Machines, Regularization, Optimization, and Beyond (reprint) Edition. The MIT Press and MIT Press, Cambridge, MA, USA.

Schölkopf, B., Smola, J.A., Müller, K.R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10 (5), 1299–1319.

Schölkopf, B., Sung, K., Burges, J.C., Girosi, F., Niyogi, P., Poggio, T., Vapnik, N.V., 1997. Comparing support vector machines with gaussian kernels to radial basis function classifiers. IEEE Transactions on Signal Processing 45, 2758–2765.

Schulze-Lammers, P., Vondricka, J., 2010. Precision crop protection-the challenge and use of heterogeneity. In: Direct Injection Sprayer, first ed. Springer Verlag, Dordrecht, Heidelberg, London, New York, Ch., pp. 295–310.

Sökefeld, M., 2010. Precision crop protection-the challenge and use of heterogeneity. In: Variable Rate Technology for Herbicide Application, first ed. Springer Verlag, Dordrecht, Heidelberg, London, New York, Ch., pp. 335–347.

Sökefeld, M., Gerhards, R., Oebel, H., Therburg, R.-D., 2007. Image acquisition for weed detection and identification by digital image analysis. In: Stafford, J. (Ed.), Precision Agriculture '07, Sixth European Conference on Precision Agriculture, vol. 6. Wageningen Academic Publishers, The Netherlands, pp. 523–529.

Tellaeche, A., Pajares, G., Burgos-Artizzu, X.P., Ribeiro, A., 2011. A computer vision approach for weeds identification through support vector machines. Applied Soft Computing 11 (1), 908–915.

Vapnik, N.V., 1998. Statistical learning theory. Wiley, New York.

Vapnik, N.V., 2000. The nature of statistical learning theory, Statistics for engineering and information science, 2nd Edition. Springer-Verlag, New York.

Weis, M., Gerhards, R., 2007. Feature extraction for the identification of weed species in digital images for the purpose of site-specific weed control. In: Stafford, J. (Ed.), Precision Agriculture '07, Sixth European Conference on Precision Agriculture, ECPA, vol. 6. Wageningen Academic Publishers, The Netherlands, pp. 537–545.

Weis, M., Sökefeld, M., 2010. Precision crop protection-the challenge and use of heterogeneity. In: Detection and Identification of Weeds, first ed. Springer Verlag, Dordrecht, Heidelberg, London, New York, Ch., pp. 119–134.

Zhu, W., Zhu, X., 2009. The application of support vector machine in weed classification. In: Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on, vol. 4., pp. 532–536.

A.3 Development of spectral indices for detecting and identifying plant diseases

Mahlein, A.-K., Rumpf, T., Welke, P., Dehne, H.-W., Plümer, L., Steiner, U., Oerke, E.-C., 2013. Development of spectral plant disease indices. Remote Sensing of Environment 128, 21–30.

Abstract

Spectral vegetation indices (SVIs) have been shown to be useful for an indirect detection of plant diseases. However, these indices have not been evaluated to detect or to differentiate between plant diseases on crop plants. The aim of this study was to develop specific spectral disease indices (SDIs) for the detection of diseases in crops. Sugar beet plants and the three leaf diseases *Cercospora* leaf spot, sugar beet rust and powdery mildew were used as model system. Hyperspectral signatures of healthy and diseased sugar beet leaves were assessed with a non-imaging spectroradiometer at different developing stages and disease severities of pathogens. Significant and most relevant wavelengths and two band normalized differences from 450 to 950 nm, describing the impact of a disease on sugar beet leaves were extracted from the data-set using the RELIEF-F algorithm. To develop hyperspectral indices for the detection of sugar beet diseases the best weighted combination of a single wavelength and a normalized wavelength difference was exhaustively searched testing all possible combinations. The optimized disease indices were tested for their ability to detect and to classify healthy and diseased sugar beet leaves. With a high accuracy and sensitivity healthy sugar beet leaves and leaves, infected with *Cercospora* leaf spot, sugar beet rust and powdery mildew were classified (balanced classification accuracy: 89%, 92%, 87%, 85%, respectively). Spectral disease indices were also successfully applied on hyperspectral imaging data and on non-imaging data froma sugar beet field. Specific disease indices will improve disease detection, identification and monitoring in precision agriculture applications.

Remote Sensing of Environment 128 (2013) 21–30

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Development of spectral indices for detecting and identifying plant diseases

A.-K. Mahlein [a,*,1], T. Rumpf [b,1], P. Welke [b], H.-W. Dehne [a], L. Plümer [b], U. Steiner [a], E.-C. Oerke [a]

[a] Institute of Crop Science and Resource Conversation (INRES–Phytomedicine), University of Bonn, Nussallee 9, D-53115 Bonn, Germany
[b] Institute of Geodesy and Geoinformation, Department of Geoinformation, University of Bonn, Meckenheimer Allee 172, D-53115 Bonn, Germany

ARTICLE INFO

Article history:
Received 26 October 2011
Received in revised form 28 August 2012
Accepted 19 September 2012
Available online xxxx

Keywords:
Spectral disease indices
Hyperspectral reflectance
Band selection
Sugar beet
Plant diseases
Precision crop protection
Cercospora leaf spot
Powdery mildew
Sugar beet rust

ABSTRACT

Spectral vegetation indices (SVIs) have been shown to be useful for an indirect detection of plant diseases. However, these indices have not been evaluated to detect or to differentiate between plant diseases on crop plants. The aim of this study was to develop specific spectral disease indices (SDIs) for the detection of diseases in crops. Sugar beet plants and the three leaf diseases *Cercospora* leaf spot, sugar beet rust and powdery mildew were used as model system. Hyperspectral signatures of healthy and diseased sugar beet leaves were assessed with a non-imaging spectroradiometer at different developing stages and disease severities of pathogens. Significant and most relevant wavelengths and two band normalized differences from 450 to 950 nm, describing the impact of a disease on sugar beet leaves were extracted from the data-set using the RELIEF-F algorithm. To develop hyperspectral indices for the detection of sugar beet diseases the best weighted combination of a single wavelength and a normalized wavelength difference was exhaustively searched testing all possible combinations. The optimized disease indices were tested for their ability to detect and to classify healthy and diseased sugar beet leaves. With a high accuracy and sensitivity healthy sugar beet leaves and leaves, infected with *Cercospora* leaf spot, sugar beet rust and powdery mildew were classified (balanced classification accuracy: 89%, 92%, 87%, 85%, respectively). Spectral disease indices were also successfully applied on hyperspectral imaging data and on non-imaging data from a sugar beet field. Specific disease indices will improve disease detection, identification and monitoring in precision agriculture applications.

1. Introduction

Near-range and remote sensing methods, like hyper- and multispectral sensors or thermography possess multiple opportunities to increase the productivity of agricultural production systems (Oerke et al., 2006; Steiner et al., 2008). Remote sensing technologies can provide an automatic and objective alternative to visual disease assessment of plant diseases (Hillnhütter & Mahlein, 2008; Mahlein et al., 2012a; Nutter et al., 1990). Many researchers have shown the capabilities of remote sensing techniques in the area of agriculture and crop production (i.e. Doraiswamy et al. (2003); Galvao et al. (2009); Thenkabail et al. (2000)) and also in the field of plant disease detection. Hillnhütter et al. (2011), Mahlein et al. (2010, 2012 b, a), Moshou et al. (2004) and Steddom et al. (2005) have proven the potential of spectral sensor systems for the detection of fungal diseases.

Efficient use of spectral reflectance measurements for disease detection relies on the identification of most significant spectral wavelength, highly correlated to a specific disease. Depending on the application area and aim, just a few regions of the spectrum are of interest. In the visible region from 400 to 700 nm, the composition of pigments has a predominant impact on the spectral signature (Blackburn & Steele, 1999; Gitelson et al., 2002). The near infrared from 700 to 1100 nm is mainly influenced by structural leaf traits and water content (Jacquemoud & Baret, 1990).

The use of spectral vegetation indices (SVIs) is a common method to analyze and to detect changes in plant physiology and chemistry. These indices, based on the information of few wavelengths have been developed to specify different plant parameters, such as pigment content (Blackburn, 1998a; Gitelson et al., 2002; Peñuelas et al., 1995), leaf area (Rouse et al., 1974) or water content (Peñuelas et al., 1993). A quantitative relation between the trait of interest, i.e. high correlation to pigment or water content, and SVIs is typical. Several approaches however have shown that SVIs have additionally the potential to detect plant diseases (Hatfield et al., 2008; Thenkabail et al., 2000). But a quantitative statement or the identification of a specific disease based on common SVIs is not possible so far, since these indices lack disease specificity. Therefore, disease specific data analysis methods and algorithms are required. A combination of different wavelengths to so called spectral disease indices (SDI) can be useful to simplify disease detection by spectral sensors, since each disease influences the spectral signature in a characteristic way.

In contrast to common SVIs the developed SDIs aimed to detect diseased sugar beet leaves as well as to identify and differentiate one disease from others. Experiments under controlled conditions were performed to identify the most important spectral wavelength for plant disease detection and identification (Mahlein et al., 2010). A methodology for the

* Corresponding author. Tel.: +49 228 734 997.
E-mail address: amahlein@uni-bonn.de (A.-K. Mahlein).
[1] Both authors equally contributed to this work.

0034-4257/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.rse.2012.09.019

extraction and combination of most relevant wavelengths to SDIs was developed. For a reduction of data dimensionality each SDI consists of the combination of a single and a normalized wavelength difference according to Blackburn and Steele (1999), Carter and Miller (1994), Delalieux et al. (2009), and Thenkabail et al. (2000). To validate the transferability of the developed indices, independent data sets from field spectroscopy and from a hyperspectral camera were used.

Based on the above background, the main objective of this paper was (I) to identify disease specific single wavelengths and wavelength differences, (II) to combine these specific wavelengths to spectral disease indices, and (III) to approve the developed indices on independent datasets. Therefore a statistical approach was adopted to compute and to evaluate spectral disease indices.

2. Material and methods

2.1. Experimental setup

A data set of spectral signatures from healthy sugar beet leaves and from sugar beet leaves inoculated with foliar diseases was used as groundwork for DI development. A specific index was evaluated for each of the four classes (I) healthy, (II) *Cercospora* leaf spot, (III) powdery mildew and (IV) sugar beet rust. The spectral reflectance was assessed under constant conditions in a controlled environment, and experiments were repeated twice.

2.2. Plant cultivation and pathogen inoculation

Sugar beet plants, cultivar Pauletta (KWS GmbH, Einbeck, Germany), were grown in a commercial substrate (Klasmann-Deilmann GmbH, Germany) in plastic pots (ø 13 cm) in a greenhouse at 23/20 °C (day/night), 60% relative humidity (RH) and a photo-period of 16 h (>300 μmol/m^{-2} s^{-1}) per day. Plants were watered as necessary and fertilized weekly with 100 ml of a 0.2% solution of Poly Crescal (Aglukon GmbH, Düsseldorf, Germany).

For each class, 15 plants were inoculated with the pathogens when four leaves were fully developed. Production of inoculum of the pathogens *Cercospora beticola*, *Uromyces betae* and *Erysiphe betae* and inoculation of plants were carried out according to Mahlein et al. (2010). Each treatment consisted of 15 plants.

2.3. Measurement of leaf reflectance

Spectral reflectance was measured with a handheld non-imaging spectro-radiometer using a plant probe foreoptic with a leaf clip holder (ASD FieldSpec Pro FR spectrometer, Analytic Spectral Devices, Boulder, USA). The spectral range was from 350 nm to 1100 nm with a spectral resolution of 3 nm. Because the reflectance spectra data were noisy at the extremes, values between 400 nm and 1050 nm were analyzed. The contact probe foreoptic has a 10 mm field of view and an integrated 100 W halogen reflector lamp. The instrument was warmed up for 90 min previous to measurement to increase the quality and homogeneity of spectral data. Instrument optimization and reflectance calibration were performed prior to sample acquisition. The average of 25 dark-current measurements was calibrated to the average of 25 barium sulfate white reference (Spectralon, Labsphere, North Sutton, NH, USA) measurements. Because of the internal light source, the integration time was adjusted to 17 ms per scan constantly. Finally, reflectance spectra were obtained by determining the ratio of recorded sample data to data acquired for the white reflectance standard. Each sample scan represented an average of 25 reflectance spectra. Because reflectance spectra were assessed under constant light and temperature conditions with the plant probe foreoptic, pre-processing to smooth the spectrum and reduce signal noise was not necessary.

Data of infected and non-infected leaves were collected and recorded daily until 21 days after inoculation (dai). In each class, spectra from 15 plants and 2 leaves per plant from the adaxial leaf surface were taken. Disease severity caused by each pathogen was evaluated daily and classified according to Wolf and Verreet (2002). The experiment was repeated twice.

2.4. Development of specific disease indices

A statistical approach was adapted to compute and to evaluate SDIs. Fig. 1 summarizes the steps from spectral signatures to disease-specific

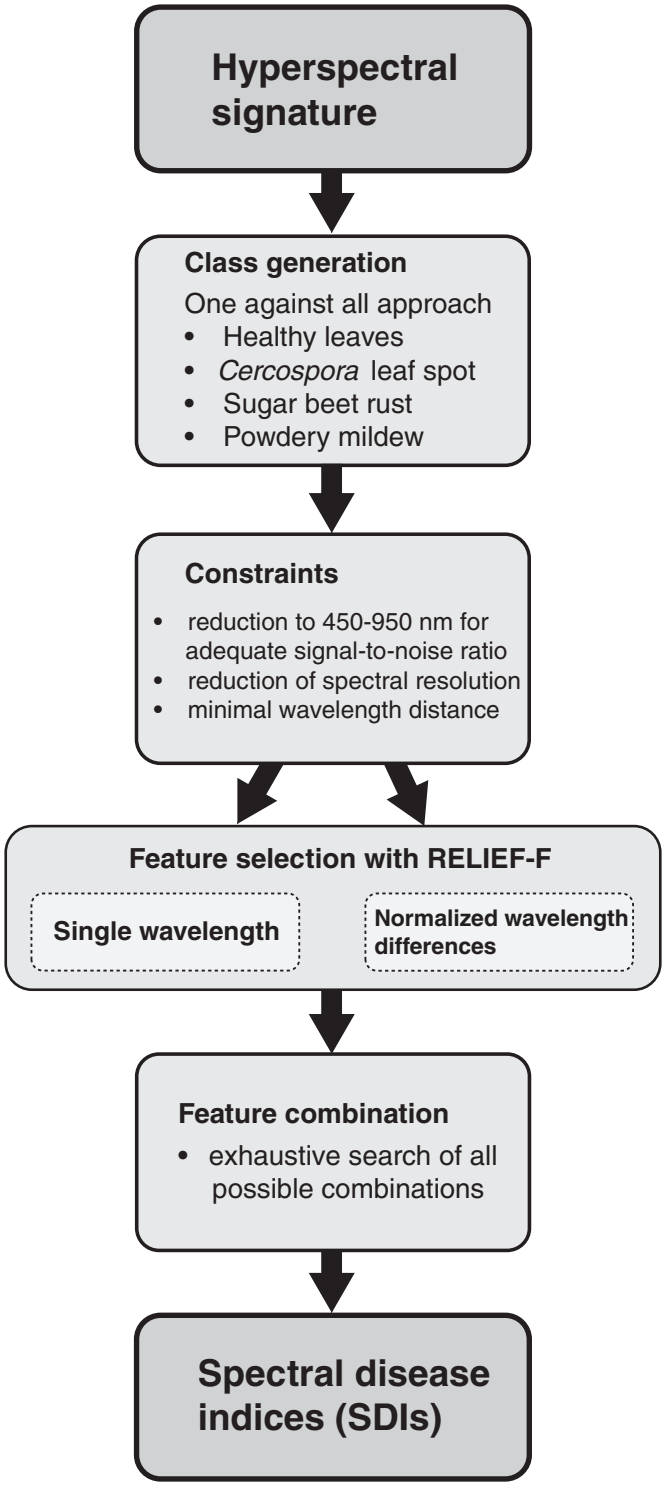


Fig. 1. Systematical approach and development of specific disease indices (DIs) from hyperspectral reflectance data.

indices. Hyperspectral signature from the classes (I) healthy, (II) *Cercospora* leaf spot, (III) powdery mildew and (IV) sugar beet rust were the basis for the generation of SDIs.

Reflectance data from disease severities of 1 to 30% was used for index development to be sensitive to early disease stages and to use characteristic hyperspectral signatures of mature disease symptoms as a basis. Hyperspectral signatures of every class were compared to all other classes in order to ensure disease specificity in a one against all approach. Several constraints were set to reduce dimensionality of the data and to increase the robustness and transferability to other data sets in advance. Since wavelengths next to each other are highly correlated, the minimal distance between wavelengths combined in normalized differences was set to 50 nm (a counter plot indicating wavelength correlation can be found in the supplementary data, Appendix A). Further the spectral resolution was reduced to 6 nm due to high correlation between neighboring wavelengths. The developed indices aim at qualitatively recognizing a specific plant disease independent of disease severity. Thus a combination of a single wavelength and a normalized wavelength difference seemed suitable. A combination of two normalized two band wavelength differences was tested simultaneously and resulted in lower classification results. Single wavelengths are especially important to differentiate samples with higher disease severity. Differences were suitable to assess changes in the hyperspectral signature caused by *Cercospora* leaf spot and sugar beet rust, as the course of the disease cause opposed changes at different ranges on the hyperspectral signature. As powdery mildew, however, caused a parallel shift in the hyperspectral signature, normalized differences performed separation of powdery mildew and the other classes' best.

Before finding the best combination of single wavelength and wavelength difference, feature selection was necessary. The most relevant single wavelengths and wavelength differences were extracted by the RELIEF-F algorithm (Robnik-Šikonja & Kononenko, 2003) (for detailed information on feature selection and the RELIEF algorithm please see Appendix B in the supplementary data). The RELIEF algorithm (Kononenko, 1994) estimates the relevance of wavelength according to their goodness to separate samples of both classes which are near to each other. The advantage of the RELIEF-F algorithm is to correctly estimate the quality of features with strong dependencies. Moreover, the algorithm is robust against outliers. The RELIEF-F searches for the two nearest neighbors of a given sample. Each neighborhood consists of k samples. For a given k, 'hit' is the set of k nearest neighbors of the same class and 'miss' from the different class. In fact, the relevance of the wavelengths is determined by the sum of the Euclidean distance between nearest misses and nearest hits for all samples (List. 1).

Listing 1: Pseudo code of the Relief-F algorithm for two class classification

INPUT: A set of features F, a set of samples X,
a label function l, a number of iterations $p \in \mathbb{N}$
and a number of neighbors $k \in \mathbb{N}$
OUTPUT: A set of feature weights $w = \{w(F_1), \cdots, w(F_m)\}$
Set $w(F_i) := 0 \; \forall \; i \in \{0, \ldots, m\}$
for $i := 1$ to p
  Pick $x \in X$ at random
  Find k nearest hits $A \subseteq X$ and nearest misses $B \subseteq X$
  for $j := 1$ to m
    $w(F_j) := w(F_j) - \frac{1}{k} \sum_{a \in A} |x_j - a_j| + \frac{1}{k} \sum_{b \in B} |x_j - b_j|$
  end
end
Set $w(F_i) := \frac{w(F_i)}{m} \; \forall \; i \in \{0, \ldots, m\}$
return $\{w(F_1), \ldots, w(F_m)\}$

Finally, the best combination of an individual wavelength and a normalized wavelength difference has to be found. The best weighted 25% of single wavelengths and normalized wavelength differences were included for SDI development. All possible combinations were exhaustively searched to find the best index with a related threshold for separating the index-specific class. Because of the high variance in hyperspectral signature between individual plants and high variance especially at low disease severities, the normalized wavelength difference is mandatory for building a specific index. Furthermore a weighting factor for the single wavelength has to be determined. We set the possible weights -1, -0.5, 0, 0.5 and 1. The final indices were evaluated for their classification result using an 8-fold cross validation by stratified sampling. The balanced accuracy was used for classification accuracy, whereby both classes were equally weighted. The two independent data sets from two experiments afford optimal conditions to develop indices as robust as possible. In general all calculations were programmed in MATLAB 7.11 using parallel processing by an eight-core processor.

2.5. Comparison with common vegetation indices

The ability to identify disease with SDIs was tested and compared to vegetation indices from literature. The classification accuracies of the NDVI (Rouse et al., 1974), the PRI (Gamon et al., 1992), the ARI (Gitelson et al., 2001), the SIPI (Peñuelas et al., 1995), the mCAI (Laudien et al., 2003), the PSSRa, PSSRb, and PSSRc (Blackburn, 1998b), the GM1 and GM2 (Gitelson & Merzlyak, 1997), the ZM (Zarco-Tejada et al., 2001) and the TCARI/OSAVI (Haboudane et al., 2002) were compared to the developed SDIs using an 8-fold cross validation by stratified sampling.

2.6. Independent data sets

To test the transferability of the developed disease indices on independent data sets, non-imaging data from a field experiment and hyperspectral imaging data were used. The datasets were assessed with different sensors and under different measuring conditions.

2.6.1. Field data

A field experiment was conducted at the research station Klein-Altendorf (50°36′55.36″N, 7°0′0.10″E) of the University of Bonn in the growing season of 2008 as described by Mahlein et al. (2009). The field was divided in two treatments; treatment 1 without fungicide application in order to monitor the incidence of fungal diseases over the growing season; the plants of treatment 2 were treated once with the fungicide Spyrale® (Syngenta Agro GmbH; 1 l/ha, difenoconazol 100 g/l; fenpropidin 375 g/l) to avoid fungal infections. On September 9, 2008, spectral reflectance from sugar beet canopy was assessed using an ASD FieldSpecPro FR with a pistol grip foreoptic, which was placed 1 m above the sugar beet canopy constantly. The spectral range of the spectroradiometer was from 400 nm to 1050 nm. The spectral sampling interval was automatically interpolated from 1.4 nm to 1 nm steps using a linear equation by the ASD software. Instrument optimization and reflectance calibration were performed prior to sample acquisition. The average of 50 dark current measurements was calibrated to the average of 50 barium sulfate white reference (Spectralon, Labsphere, North Sutton, NH, USA) measurements. The integration time was adjusted to 34 ms per scan constantly. Finally reflectance spectra were obtained by determining the ratios of data acquired for a sample to data acquired for the white reflectance standard. Each sample scan represented an average of 50 reflectance spectra. Spectral reflectance and ground truth data, in particular incidence and severity of diseases were collected and geo-referenced at 50 sampling points.

2.6.2. Hyperspectral imaging data

Hyperspectral images from diseased sugar beet plants were taken using an imaging system. The hyperspectral imaging system combines an imaging spectrograph and a mirror scanner. The line scanning

spectrograph ImSpector V10E (Spectral Imaging Ltd., Oulu, Finland) has a spectral range from 400 to 1000 nm and a spectral resolution of up to 2.8 nm. The maximal image size of the 30 μm sensor slot results in 1600 pixels per line with a sensor pixel size of 0.0074 mm. Limited by the distance between target and sensor system (0.9 m) a spatial resolution of 0.29 mm per pixel was obtained. A mirror scanner (Spectral Imaging Ltd.) – maximal field of view 80° – mounted in front of the objective lens provided the second spatial dimension of images. The hyperspectral sensor system was mounted on a manual positioning XY-frame, surrounded by six ASD-Pro-Lamps (Analytical Spectral Devices Inc., Boulder, USA) radiating a near-solar light spectrum. The distance between lamps and leaves was 0.8 m with a vertical orientation of 45°. Imaging data were recorded in a dark chamber in order to realize optimal and reproducible illumination and constant measurement conditions. For subsequent calculation of reflectance, three images were grabbed. A dark current image was recorded by closing an internal shutter of the camera, followed by an image of a white reference bar (Spectral Imaging Ltd., Oulu, Finland), with the same horizontal size and on the same level as the object area, both with the same exposure time. Subsequently an image of the leaf area was recorded with improved exposure time. Calculation of reflectance, relative to a white reference bar and the dark current measurement were performed using the software ENVI 4.6 + IDL 7.0 (ITT Visual Information Solutions, Boulder, USA). Prior to the calculation of disease indices on imaging data the background was masked out. For detailed information on data assessment, data pre-processing and normalization see Mahlein et al. (2012b).

3. Results

3.1. Spectral signatures

First symptoms of *Cercospora* leaf spot appeared 6 dai, of powdery mildew 5 dai and sugar beet rust 8 dai, respectively. For the differentiation of leaf diseases based on reflectance measurement, spectral signatures at different disease severities have been measured and compared. Fig. 2A summarizes the averaged spectral signatures of healthy sugar beet leaves and sugar beet leaves with *Cercospora* leaf spot, powdery mildew, and sugar beet rust at disease severity of 10 to 30%, respectively. Compared to the spectra of healthy leaves, each disease gave a divergent, characteristic reflectance curve, strongly correlated to the occurrence of disease-specific symptoms. Reflectance of *C. beticola*-infected leaves increased in the VIS mostly in the green and red ranges of the spectrum between 500 and 700 nm and decreased from 700 to 900 nm. The slope at the red edge position between VIS and NIR decreased. A blue shift of the red edge position depending on *Cercospora* leaf spot disease severity was obvious. Reflectance of leaves colonized by the ectoparasite *E. betae* rose consecutively within the measuring period and increased with disease severity (Fig. 2A). This increment was most distinctive in the VIS and less pronounced in the NIR. Powdery mildew rather affected the overall level of reflectance than the profile of spectra. Due to the small symptoms of the biotroph pathogen *U. betae* scattered on the leaf area, changes in reflectance spectra were comparatively low for sugar beet rust (Fig. 2A). Reflectance of leaves with 10 to 30% disease severity was high between 550 and 700 nm and low from 700 to 900 nm.

The variance of the dataset was analyzed by standard deviation of reflectance and by the absolute differences between the reflectance median (Fig. 2B). Standard deviation of diseased sugar beet leaves increased in the VIS from 400 to 700 nm and in the NIR from 720 to 780 nm with increasing disease severity. The reflectance difference likewise increased with disease severity. At a disease severity below 10% differences were detected in the VIS from 500 to 700 nm and around the red edge position. At disease severities in the range 10 to 30% differences increased obviously in the VIS with a peak around 700 nm. Reflectance differences at

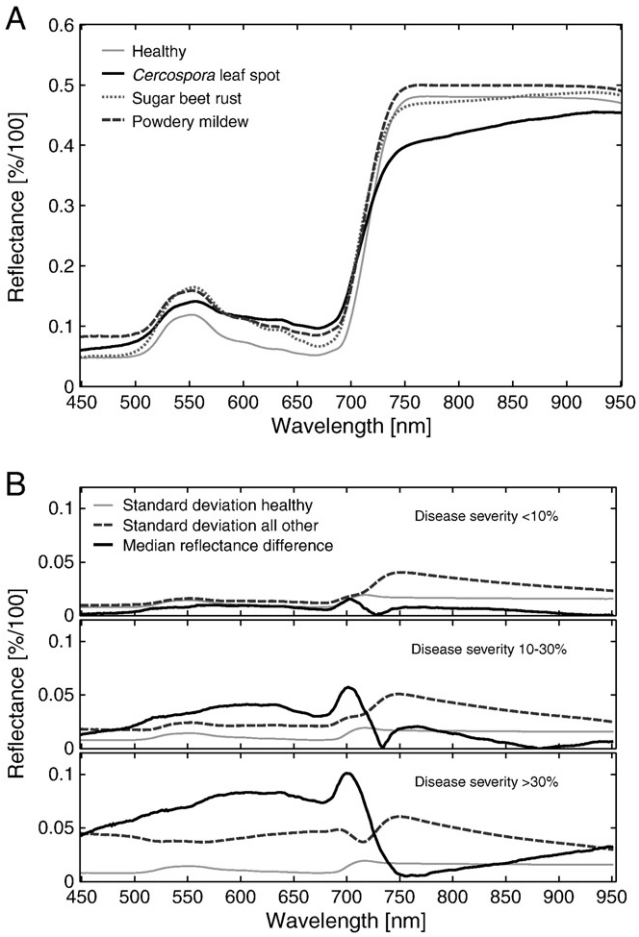


Fig. 2. Spectral signatures of healthy sugar beet leaves and sugar beet leaves with 10 to 30% disease severity of *Cercospora* leaf spot, sugar beet rust and powdery mildew, respectively; and differences between reflectance medians and standard deviation of healthy sugar beet leaves and diseased sugar beet leaves at different disease severities.

disease severities above > 30% were high in the VIS and at the red edge position.

3.2. Spectral plant disease indices

For the evaluation of specific DIs the different classes were parted into two groups. In the first step a binary classification of healthy sugar beet leaves from diseased leaves was proven in a one against all approach. In addition to a binary classification of healthy and diseased sugar beet leaves, the detection and identification of plant diseases by specific SDIs were proven. Therefore disease specific indices, based on significant wavelengths were evaluated. To reduce the information of the entire spectrum, the most relevant single wavelengths and wavelength differences were evaluated. The result of this data reduction was the basis for index development.

Relevant single wavelengths for the classes were assessed calculating the RELIEF algorithm (Fig. 3). Single wavelengths of high relevance were from 670 to 690 nm (Fig. 3A). Single wavelengths around 500 nm, 690 nm and 730 nm were highly relevant for sugar beet leaves diseased with *Cercospora* leaf spot (Fig. 3C). Single wavelengths relevant to sugar beet rust infection according to the RELIEF-F algorithm were around 500 nm, 570 nm and 700 nm (Fig. 3E), and for powdery mildew infection at 450 and 720 nm (Fig. 3G).

The contour plot (Fig. 3B) visualizes normalized wavelength differences highly correlated to the class 'healthy'. High relevance was detected for normalized differences including reflectance difference between 600 and 670 nm and at 700 nm. Normalized reflectance

differences, highly correlated to *Cercospora* leaf spot included values around 550 nm and 700 nm (Fig. 3D), for sugar beet rust infected leaves from 500 to 600 nm and normalized wavelength differences including 700 nm (Fig. 3F), and for powdery mildew infection normalized wavelength differences between 500 and 600 nm (Fig. 3H).

Based on single wavelengths of high relevance according to the RELIEF algorithm and the two band normalized differences highly correlated to the class 'healthy', possible wavelength combinations and algorithms for a class specific vegetation index were calculated. The final health-index (HI) is based on the absolute reflectance at 704 nm and the normalized reflectance difference 534 nm to 698 nm (Eq. (1)). Testing all possible wavelength combinations, the *Cercospora* leaf spot-index (CLSI) is based on the reflectance at 734 nm and the normalized reflectance difference of 698 nm and 570 nm (Eq. (2)). After this data dimensionality reduction the sugar beet rust-index (SBRI) was evaluated (Eq. (3)) based on reflectance at 704 nm and

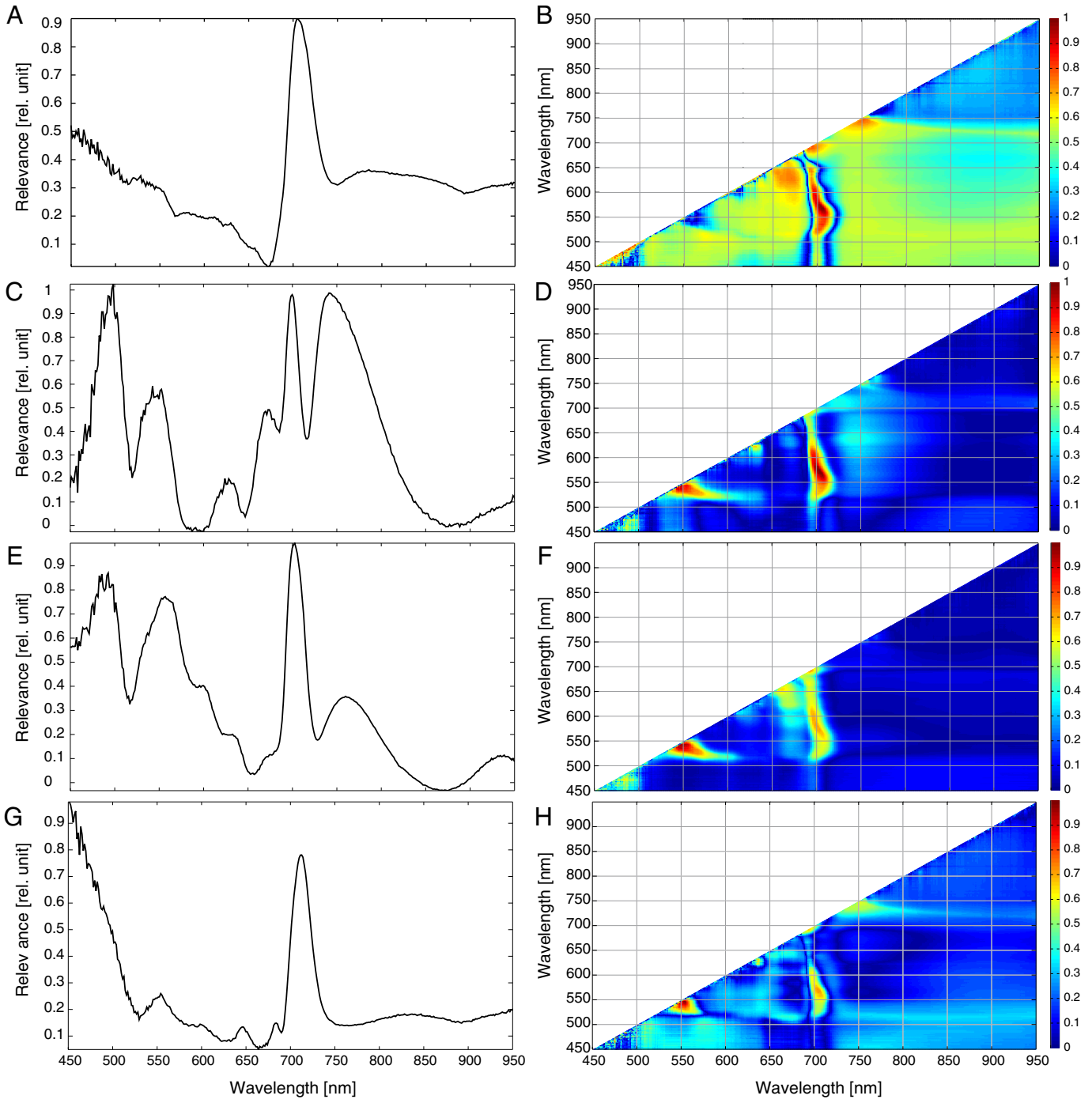


Fig. 3. Relevant single wavelengths for the class healthy sugar beet leaves (A), *Cercospora* leaf spot (C), sugar beet rust (E) and powdery mildew (G) diseased sugar beet leaves, according to the RELIEF-F algorithm, and contour plot for the relevance of all possible wavelength differences of the class label healthy (B),and the classes *Cercospora* leaf spot (D), sugar beet rust (F) and powdery mildew (H) at a disease severity of 10 to 30%, respectively. The color scale represents the level of relevance according to the RELIEF-F algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the normalized difference of 570 and 513 nm. Previous information resulted in the powdery mildew-index (PMI; Eq. (4)), related to absolute reflectance at 724 nm and the normalized reflectance difference of 520 nm and 584 nm.

$$Healthy-index\,(HI):\frac{R534-R698}{R534+R698}-\frac{1}{2}\cdot R704 \quad (1)$$

$$Cercospora\ leaf\ spot-index\,(CLS):\frac{R698-R570}{R698+R570}-R734 \quad (2)$$

$$Sugar\ beet\ rust-index\,(SBRI):\frac{R570-R513}{R570+R513}+\frac{1}{2}\cdot R704 \quad (3)$$

$$Powdery\ mildew-index\,(PMI):\frac{R520-R584}{R520+R584}+R724 \quad (4)$$

The separation ability of the spectral disease indices is illustrated in Fig. 4, depending on the index value at different days after inoculation for each class. A threshold value was optimized to separate healthy from diseased sugar beet leaves. Classification based on the HI resulted in a classification accuracy of 89%, with a class precision of 93.5% for the classification of diseased sugar beet leaves and 85.5% for healthy sugar beet leaves (Table 1). The classification error for healthy sugar beet leaves was 5.2%, independently of disease severity. The classification error for diseased sugar beet leaves decreased with increasing disease severity. At a disease severity from 1 to 2% classification error was relatively high (51.3%) and decreased constantly to 0% at a disease severity above 16%.

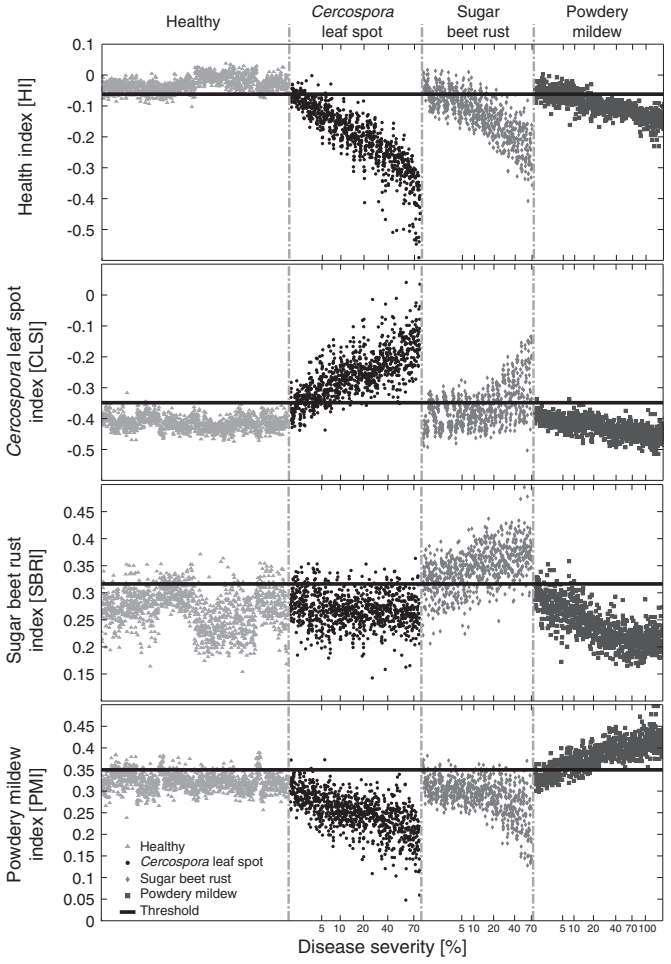


Fig. 4. Classification result based on the health-index (HI) for all disease severities over 21 days after inoculation using an 8-fold cross validation, and scatter plot of the classification based on the (A) *Cercospora* leaf spot-index (CLSI),(B) sugar beet rust-index (SBRI), and (C) powdery mildew-index (PMI) for all disease severities (samples ordered according to the day after inoculation).

Table 1
Classification result for each disease, based on the disease indices (DIs) *Cercospora* leaf spot-index (CLSI) for the classification of *Cercospora* leaf spot, sugar beet rust-index (SBRI) for sugar beet rust detection and the powdery mildew-index (PMI) for powdery mildew detection on sugar beet leaves.

Prediction	Ground truth		Class precision
HI	All other	Healthy	
All other	129	9	93.5%
Healthy	25	148	85.6%
Class recall	83.8%	94.3%	89.0%
CLSI	All other	*Cercospora*	
All other	209	7	96.8%
Cercospora	8	54	87.1%
Class recall	96.3%	88.5%	92.4%
SBRI	All other	Sugar beet rust	
All other	372	15	96.1%
Sugar beet rust	25	61	70.9%
Class recall	93.7%	80.3%	87.0%
PMI	All other	Powdery mildew	
All other	363	16	95.8%
Powdery mildew	18	45	71.4%
Class recall	95.3%	73.8%	84.5%

The classification result based on the CLSI revealed a good separation of all classes (Fig. 4B); however, the separation of highly diseased leaves was challenging during the measuring period. Sugar beet leaves diseased with *Cercospora* leaf spot were detected with a class precision of 92.4% (Table 1). At disease severities of 1 to 2% the classification was complex (classification error of 60.1%), while all other classes were classified with a classification error of 14.1% (Table 1). The classification accuracy for *Cercospora* leaf spot diseased leaves increased with increasing disease severity. For disease severities from 6 to 9% the classification error was only 5.6% for the class *Cercospora* leaf spot, and 5% for all other classes.

Classification based on the SBRI resulted in a separation of the classes sugar beet rust to healthy, *Cercospora* leaf spot and powdery mildew (Fig. 4B). Difficulties remained in the separation of early stages of sugar beet rust infection to the other leaves. With an overall class precision of 87% the classification result was satisfying (Table 1). At low disease severities from 1 to 2% the classification error amounted 52.1% (Table 1). With increasing disease severity the classification error declined, and reached 5.8% for 25 to 30% disease severity.

Classification of the four classes by the PMI resulted in a very good separation of powdery mildew from healthy sugar beet leaves as well

Table 2
Classification error of spectral disease indices (SDIs) at different disease severity levels for each plant disease and the disease specific indices CLSI, SBRI, and PMI, respectively.

Disease severity	Classification error [%]		Classification error [%]		Classification error [%]	
	CLSI		SBRI		PMI	
	All other	CLS	All other	SBR	All other	PM
1–2 [%]	14.1	60.8	11.9	52.1	5.1	97.0
3–5 [%]	8.4	23.3	4.9	42.5	2.3	57.8
6–9 [%]	5.0	5.6	5.8	16.1	0	30.6
10–15 [%]	10.8	3.1	1.7	21.1	2.1	33.9
16–20 [%]	22.0	3.1	1.5	15.5	0.8	12.3
25–30 [%]	27.0	0	2.0	5.8	2.0	0
35–40 [%]	40.6	0	3.9	9.8	0	0
45–50 [%]	29.1	2.5	5.3	5.8	0	0
> 50 [%]	8.5	0	1.2	5.6	0.7	1.6

as from *Cercospora* leaf spot and sugar beet rust diseased sugar beet leaves with an optimized threshold value (Fig. 4). An overall class precision of 84.5% was achieved (Table 1). Difficulties remained at lower powdery mildew disease severities at disease severities of up to 9% (Table 1). Whereas all other classes were definitely detected as non-powdery mildew diseased by this index (classification error: 0% to 5.1%), the classification error for powdery mildew identification was high (classification error: 97.0%to 1.6%). At disease severity above 25%, a perfect classification (classification error 0%) was obtained (Table 2).

3.3. Comparison with common vegetation indices

The new developed SDIs had higher classification accuracy compared to common vegetation indices like the NDVI or pigment specific indices (Table 3). The classification accuracy of the HI was best (89.0%) for the differentiation of healthy and diseased sugar beet leaves, followed by ZM, PRI, PSSRb and ARI with classification accuracies of 85.4%, 83.5%, 83.1%, and 82.7%, respectively (Table 3). Classification of *Cercospora* leaf spot diseased leaves was best using CLSI with a classification accuracy of 92.4%. The mCAI, ARI, and PRI showed acceptable classification accuracies (89.5%, 84.8%, and 82.6%, respectively). However the recall clarified, that all vegetation indices had low classification accuracies for the identification of *Cercospora* diseased leaves, and high accuracies only for the detection of all other leaves (Table 3). In general SVIs are unsuitable to detect sugar beet rust and powdery mildew; the classification accuracy was 50% for all applied SVIs.

Table 3
Comparison of the classification ability of the health-index (HI) and the *Cercospora* leaf spot index (CLSI) and common spectral vegetation indices (SVIs): *NDVI* normalized difference vegetation index, *PRI* Photochemical reflectance index, *ARI* anthocyanin reflectance index, *mCAI* modified chlorophyll absorption integral, *PSSRa* pigment specific simple ratio chlorophyll a, *PSSRb* pigment specific simple ratio chlorophyll b, and *PSSRc* pigment specific simple ratio carotenoids.

Index	Classification accuracy	Recall	
		Healthy	All other
HI	89.0%	94.3%	83.8%
NDVI	80.3%	89.9%	70.8%
PRI	83.5%	88.5%	78.5%
ARI	82.7%	86.7%	78.6%
SIPI	74.5%	86.8%	62.3%
mCAI	77.0%	92.4%	61.5%
PSSRa	78.8%	78.7%	78.9%
PSSRb	83.1%	79.4%	86.7%
PSSRc	57.3%	18.8%	95.9%
GM1	82.7%	77.2%	88.2%
GM2	82.4%	71.5%	93.3%
ZM	85.4%	78.6%	92.2%
TCARI/OSAVI	77.5%	64.2%	90.8%

Index	Classification accuracy	Recall	
		Cercospora	All other
CLSI	92.4%	88.5%	96.3%
NDVI	73.4%	48.4%	98.3%
PRI	82.6%	68.1%	97.2%
ARI	84.8%	70.0%	99.7%
SIPI	74.7%	55.6%	93.9%
mCAI	89.5%	79.7%	99.3%
PSSRa	69.6%	42.1%	97.0%
PSSRb	69.6%	41.1%	98.1%
PSSRc	50.0%	0.0%	100.0%
GM1	50.0%	0.0%	100.0%
GM2	69.5%	41.2%	97.7%
ZM	66.6%	35.7%	97.4%
TCARI/OSAVI	50.7%	1.4%	1.0%

3.4. Application of spectral disease indices on independent data sets

An implementation of the developed SDIs was proven on two different independent data sets. Due to different light conditions and sensor specification it was necessary to calculate the thresholds for disease classification individually for each dataset. In the field experiment powdery mildew and *Cercospora* leaf spot were the predominant diseases during vegetation period (see Mahlein et al. (2009) for details). In most cases both diseases occurred as mixed infections of sugar beet plants. A classification of non-imaging spectroradiometer data from the canopy scale based on the HI was possible with an accuracy of 81.63% (Fig. 5). Due to the high incidence of mixed infection a classification based on a disease specific index was not suitable.

The different SDIs were calculated pixel wise on hyperspectral imaging data of a *Cercospora* leaf spot diseased leaves under controlled conditions (Fig. 6). The HI was suitable to detect all symptomatic necrotic lesions on the sugar beet leaf with an accuracy of 95.91% and all characteristic symptoms were detected as non-healthy. In a next step the disease specific indices were applied on the imaging data. A specific classification of *Cercospora* leaf spots was possible with an accuracy of 95.9% using the CLSI, whereas the PMI and the SBRI were not sensitive to *Cercospora* leaf spot symptoms. Detailed information on the classification result can be found in the supplementary data, Appendix C.

4. Discussion

As shown in previous work, foliar diseases of sugar beet impact the spectral signature of sugar beet leaves in different ways, depending on the biology of the pathogens and the characteristic host–pathogen interaction (Mahlein et al., 2010; Mahlein et al., 2012b; Rumpf et al., 2010). The present study provides evidence that spectral disease indices (SDIs) will improve and simplify plant disease detection based on hyperspectral data. The developed SDIs are based on the most relevant wavelength and normalized wavelength combinations regarding to a plant disease. They are characterized by a high sensitivity and specificity for the detection and identification of the different foliar diseases of sugar beet.

According to Blackburn (2007) recent research for optimizing SVIs resulted in the incorporation of more than two narrow wavebands from the visible region, the red edge or the near infrared. Depending on the measuring scale three band indices (leaf-scale) or four band

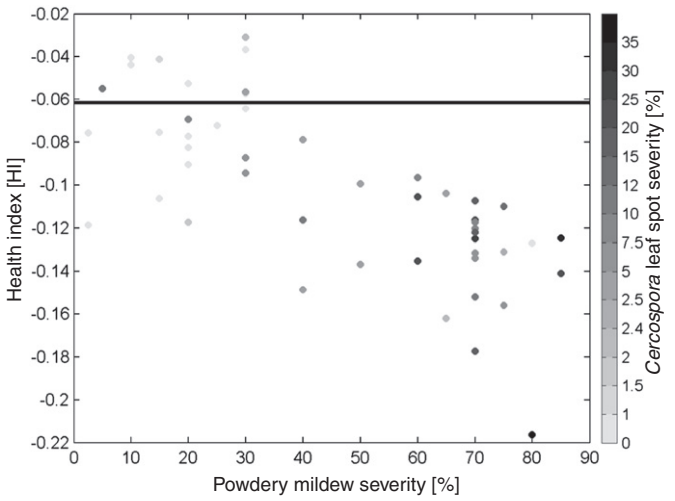


Fig. 5. Classification result based on the application of the health-index (HI) on spectroradiometer data from a field experiment on the canopy scale. Areas of the sugar beet field diseased with powdery mildew and *Cercospora* leaf spot were classified correctly with an accuracy of 81.63%. The gray scale represents disease severity of *Cercospora* leaf spot.

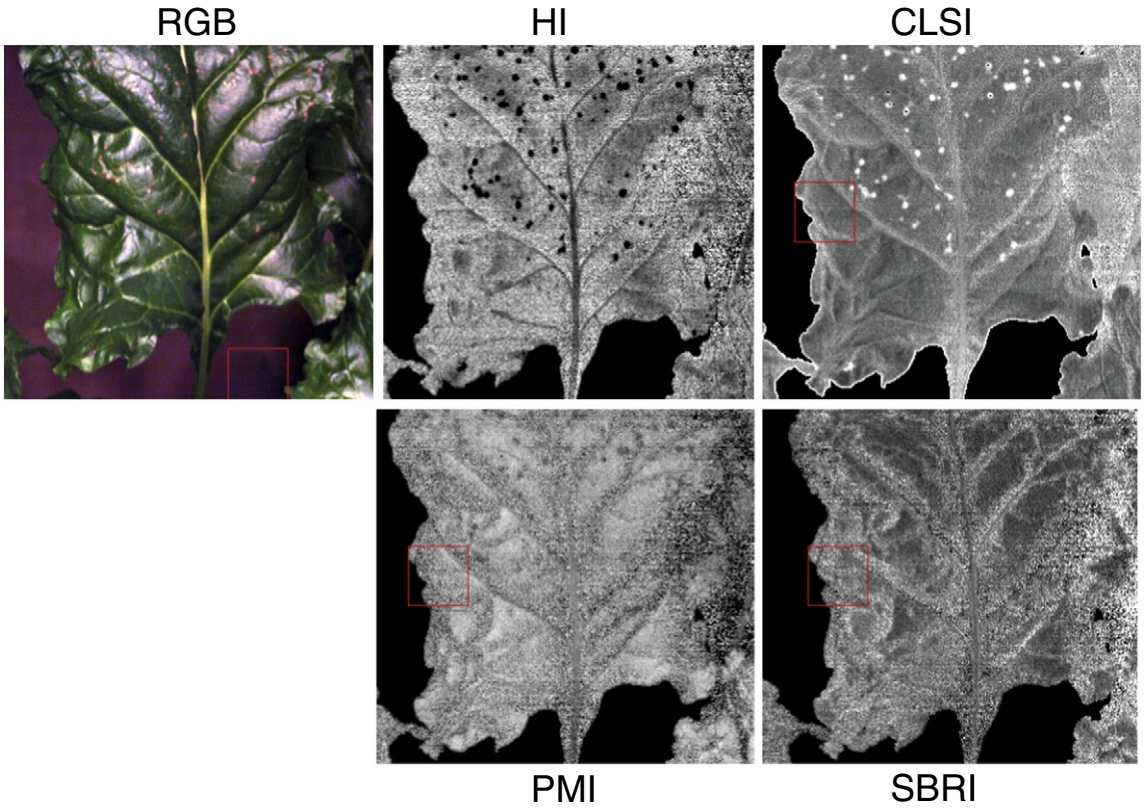


Fig. 6. RGB-image from a hyperspectral image of a *Cercospora* leaf spot diseased sugar beet leaf and gray-scale pixel-wise classification using the developed spectral disease indices (SDIs). Light-colored pixels indicate high index values and dark pixel vice versa. (HI: health-index, CLSI: *Cercospora* leaf spot-index, SBRI: sugar beet rust-index, PMI: powdery mildew-index). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

indices (canopy-scale) are generally applicable (Gitelson et al., 2003; Sims & Gamon, 2002; Thenkabail et al., 2000).

The common way to detect significant wavelength for SVI development is by correlation to a biochemical or biophysical trait (Gitelson & Merzlyak, 1996; Hatfield et al., 2008). For the development of qualitative SDIs the RELIEF-F algorithm offers multiple advantages. Since RELIEF-F is dealing with non-linear classes, a separation of the different classes becomes possible by aggregating all other data in one class. The correlation coefficient would be improper in this case. To reduce the impact of different illumination, topography, crop variety or sensor specific effects, the wavelength differences of all SDIs were normalized (Jackson, 1986; Lillesand & Kiefer, 2000; Lyon et al., 1998). By applying this concept, an improved robustness and generalization ability of the developed SDIs to independent data sets were reached. Moreover, the generalization ability of the developed indices was improved by cross validation by taking the entire data set into account. In this way variances between different experiments under similar conditions have been considered.

With the health-index a first binary classification into healthy and diseased sugar beet plants was possible. In a next step spectral disease indices were developed for an identification of each disease. The indices HI, CLSI and SBRI are based on reflectance at narrow bands centered around 700 nm. As Gitelson and Merzlyak (1996) pointed out, reflectance near 700 nm is a fundamental feature of green vegetation produced by an equilibrium between biochemical and biophysical plant characteristics. The blue shift of the reflectance curve red edge frequently accompanies stress, and could provide early detection of plant stress for most causes of stress (Carter & Miller, 1994; Gitelson & Merzlyak, 1994). Since plant diseases influence the chlorophyll content of crop plants, increased reflectance around 700 nm can be a first but unspecific indicator to detect diseased crops.

Foliar pathogens affect photosynthetic activity of infected leaves by reducing the green leaf area; an effect on photosynthesis in asymptomatic areas is reported as well (Robert et al., 2005). Interestingly, reflectance next to 700 nm is combined in HI with reflectance at 534 nm. According to Gamon et al. (1992) reflectance at 531 nm can detect the interconversion of the xanthophyll cycle pigments. Since xanthophyll cycle pigments are regulatory pigments linked to PSII light use efficiency, reflectance indices incorporating reflectance next to 531 nm could provide an indicator of photosynthetic function (Gamon et al., 1992; Rascher et al., 2010). The identification of diseases was a more complex problem, since different diseases may influence crop plants in a similar way. The main influence of *Cercospora* leaf spot to leaf reflectance was in the VIS from 550 to 700 nm and in the NIR from 700 nm to 850 nm, respectively. These regions are influenced by the chlorophyll and brown pigment content as well as by the water content and structural changes (Peñuelas & Filella, 1998). As a perthotroph pathogen *C. beticola* causes necrotic, coalescing leaf spots due to membrane damage and cell death after the fungus has penetrated the leaf stomata (Daub & Ehrenshaft, 2000). In consequence, the chlorophyll content decreases and leaf tissue structure is impaired. Reflectance in the NIR is mainly influenced by leaf tissue structure, in detail by internal scattering processes due to air spaces, water content and air–water interfaces (Asner, 1998; Ustin & Gamon, 2010; Ustin et al., 2009). Therefore, normalized reflectance difference with 698 nm and 570 nm and constant reflectance at 734 nm was a reliable indicator for *Cercospora* leaf spot infestation.

The SBRI, using three narrow bands in the visible and red edge (i.e. 513 nm, 570 nm and 704 nm) offered a robust solution for the detection of sugar beet rust. According to Merzlyak et al. (1999) reflectance from 510 nm to 520 nm represents the absorption maximum of carotenoids. It is well documented that urediniospores of rust fungi contain a high concentration of carotenoids and melanin-like pigments, causing the characteristic brown-orange color (Hougen et al., 1959; Trocha et al., 1974). Furthermore reflectance at 513 nm, before the

green peak, and 570 nm, behind the green peak, is mainly influenced by the chlorophyll content and can be used as an indicator for the chlorophyll/carotenoid ratio (Blackburn, 1998a; Blackburn, 1998b; Richardson et al., 2002). Reflectance at 704 nm as single wavelength is next to the red-edge inflection point (Carter & Miller, 1994; Gitelson & Merzlyak, 1994). Nevertheless, the small size of rust colonies impeded the detection in early stages or at low disease severities.

Similarly, an unambiguous detection of powdery mildew in early stages is challenging. First symptoms are fluffy white mycelia covering the leaf surface which affects the external reflectance and spectral signature like a dusty coat. The overall reflectance increase with increasing powdery mildew disease severity was the contributing factor in case of the PMI (Mahlein et al., 2010). By using normalized reflectance differences this difficulty has been taken into consideration. Powdery mildew was the only pathogen, causing changes in leaf reflectance in the blue region, and throughout the NIR. This effect, caused by the white powdery mycelium covering the leaf surface is similar to increasing the light intensity. Jackson and Huete (1991) described that ratios are sensitive to increasing solar intensity, which was successfully adopted for the development of PMI by using normalized differences.

For specific classification of plant diseases comparative studies have demonstrated that SDIs are superior to common SVIs. A successful application to independent data sets from the canopy scale and from a hyperspectral imaging sensor is encouraging for precision crop protection applications in the near future. It should be stressed, however, that the applicability of SDIs to datasets, assessed under changing conditions includes several difficulties. Particularly the threshold value has to be optimized prior to classification. Notwithstanding it can be affirmed, that the proposed method to develop SDIs can be transferred to hyperspectral data from different kinds of sensors, from various scales, and for different kinds of biotic and abiotic stress of crops.

5. Conclusion

All four developed spectral disease indices (SDIs) resulted in a high specificity and sensitivity for detecting and identifying plant diseases. Difficulties still remained at early infection stages, due to minor changes in leaf reflectance. For the first time SDIs, developed on disease significant single wavelength and normalized wavelength differences have been applied for non-invasive detection of crop diseases. Using these SDI diseases can be identified and differentiated, which is not possible when using vegetation indices sensitive to abiotic stress conditions, e.g. indices related to chlorophyll content. A first feasibility study demonstrated that SDIs can be used on independent data sets from different kinds of sensors and different measuring setups. Nevertheless we expect, that the use of hyperspectral imaging data for the development of SDIs will further improve the sensitivity and specificity of disease detection in near future. SDIs are of high interest for precision crop protection via reduction of data dimensionality and computationally efficient data analysis and processing. Additionally, the statistical approach for the development of SDIs can be transferred and generalized for other plant pathogen systems. Based on the *a priori* knowledge on required band settings a simple and cheap sensor that only measures reflectance at narrow bands centered at significant wavelength can be configured. This radiometer could instantaneously calculate SDIs, and, by using appropriate calibration functions, give the estimation of disease severity in seconds. Further work will be needed to transfer the developed SDIs into praxis and to test the applicability for a precise, reproducible and time saving disease monitoring on the canopy and field scale with different sensors.

Acknowledgments

This study has been conducted within the Research Training Group 722 'Information Techniques for Precision Crop Protection', funded by the German Research Foundation (DFG). The authors would like to thank the editor and reviewers for their valuable comments and suggestions to improve the quality of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.rse.2012.09.019.

References

Asner, G. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, *64*, 234–253.
Blackburn, G. A. (1998a). Quantifying chlorophylls and carotenoids at leaf and canopy scale: an evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, *66*, 273–285.
Blackburn, G. A. (1998b). Spectral indices for estimating photosynthetic pigment concentrations:a test using senescent tree leaves. *International Journal of Remote Sensing*, *19*(4), 657–675.
Blackburn, G. A. (2007). Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany*, *58*, 844–867.
Blackburn, G. A., & Steele, C. M. (1999). Towards the remote sensing of matorral vegetation physiology: Relationships between spectral reflectance, pigment, and biophysical characteristics of semiarid bushland canopies. *Remote Sensing of Environment*, *70*, 278–292.
Carter, G. A., & Miller, R. L. (1994). Early detection of plant stress by digital imaging with narrow stress-sensitive wavebands. *Remote Sensing of Environment*, *50*, 295–302.
Daub, M. E., & Ehrenshaft, M. (2000). The photoactivated *Cercospora* toxin cercosporin: Contributions to plant disease and fundamental biology. *Annual Review of Phytopathology*, *38*, 461–490.
Delalieux, S., Somers, B., Verstraeten, W. W., van Aardt, A. N. J., Keulemans, W., & Coppin, P. (2009). Hyperspectral indices to diagnose leaf biotic stress of apple plants, considering leaf phenology. *International Journal of Remote Sensing*, *30*(8), 1887–1912.
Doraiswamy, P. C., Moulin, S., Cook, P. W., & Stern, A. (2003). Crop yield assessment for remote sensing. *Photogrammetric Engineering and Remote Sensing*, *69*, 665–674.
Galvao, L. S., Roberts, D. A., Formaggio, A. R., Numata, I., & Breunig, F. M. (2009). View angle effects on the discrimination of soybean varieties and on the relationships between vegetation indices and yield using off-nadir hyperion data. *Remote Sensing of Environment*, *113*, 846–856.
Gamon, J. A., Peñeulas, J., & Field, C. B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, *41*, 35–44.
Gitelson, A. A., Gritz, Y., & Merzylak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, *160*, 271–282.
Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, *80*, 76–87.
Gitelson, A. A., & Merzlyak, M. N. (1994). Spectral reflectance changes associate with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, *143*, 286–292.
Gitelson, A. A., & Merzlyak, M. N. (1996). Remote estimation of chlorophyll content in higher plant leaves. *Journal of Plant Physiology*, *148*, 494–500.
Gitelson, A. A., & Merzlyak, M. N. (1997). Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *International Journal of Remote Sensing*, *18*, 2691–2697.
Gitelson, A. A., Merzlyak, N. M., & Chivkunova, B. O. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, *74*, 38–45.
Haboudane, D., Miller, J. R., Trembaly, N., Zarco-tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, *81*, 416–426.
Hatfield, L. J., Gitelson, A. A., Schepers, S. J., & Walthall, L. C. (2008). Application of spectral remote sensing for agronomic decisions. *Agronomy Journal*, *100*(3), 117–131.
Hillnhütter, C., & Mahlein, A. -K. (2008). Early detection and localisation of sugar beet diseases: New approaches. *Gesunde Pflanzen*, *60*(4), 143–149.
Hillnhütter, C., Mahlein, A. -K., Sikora, R., & Oerke, E. -C. (2011). Remote sensing to detect plant stress induced by heterodera schachtii and rhizoctonia solani in sugar beet fields. *Field Crops Research*, *122*(1), 70–77.
Hougen, F., Craig, B., & Ledingham, D. (1959). The oil of wheat stem rust uredospores. I. the sterol and carotenes of the unsaponifiable matter. *Canadian Journal of Microbiology*, *4*, 521–529.
Jackson, R. D. (1986). Remote sensing of biotic and abiotic plant stress. *Annual Review of Phytopathology*, *24*, 265–287.
Jackson, R. D., & Huete, A. R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine*, *11*, 185–200.
Jacquemoud, S., & Baret, F. (1990). Prospect: a model of leaf optical properties spectra. *Remote Sensing of Environment*, *34*, 75–91.
Kononenko, I. (1994). Estimating attributes: Analysis and extensions of relief. *Proceedings of the European conference on machine learning on machine learning* (pp. 171–182). Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Laudien, R., Bareth, G., & Doluschitz, R. (2003). Analysis of hyperspectral field data for detection of sugar beet diseases. *Proceedings of the EFITA Conference, Debrecen, Hungary* (pp. 375–381).

Lillesand, T. M., & Kiefer, R. W. (2000). *Remote sensing and image interpretation.* New York, NY, USA: John Wiley & Sons, Inc.

Lyon, J., Yuan, D., Lunetta, R., & Elvidge, C. (1998). A change detection experiment using vegetation indices. *Photogrammetric Engineering and Remote Sensing*, *64*(2), 143–150.

Mahlein, A. -K., Hillnhütter, C., Mewes, T., Scholz, C., Steiner, U., Dehne, H. -W., et al. (2009). Disease detection in sugar beet fields: A multi-temporal and multi-sensoral approach on different scales. In C. M. Neale, & A. Maltese (Eds.), *Proceedings of the SPIE Europe Conference on Remote Sensing*, *Vol. 7472*, (747228–747238–10).

Mahlein, A. -K., Oerke, E. -C., Steiner, U., & Dehne, H. -W. (2012). Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology*, *133*(1), 197–209.

Mahlein, A. -K., Steiner, U., Dehne, H. -W., & Oerke, E. -C. (2010). Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. *Precision Agriculture*, *11*(4), 413–431.

Mahlein, A. -K., Steiner, U., Hillnhütter, C., Dehne, H. -W., & Oerke, E. -C. (2012). Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant methods*, *8*(1), 3.

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologica Plantarum*, *106*, 135–141.

Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., & Ramon, H. (2004). Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. *Computers and Electronics in Agriculture*, *44*, 173–188.

Nutter, F. W. J., Littrell, R. H., & Brennemann, T. B. (1990). Utilization of a multispectral radiometer to evaluate fungicide efficacy to control late leaf spot in peanut. *Phytopathology*, *80*, 102–108.

Oerke, E. -C., Steiner, U., Dehne, H. -W., & Lindenthal, M. (2006). Thermal imaging of cucumber leaves affected by downy mildew and environmental conditions. *Journal of Experimental Botany*, *57*, 2121–2132.

Peñuelas, J., Baret, F., & Filella, I. (1995). Semiempirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, *31*, 221–230.

Peñuelas, J., & Filella, I. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science*, *3*, 151–156.

Penuelas, J., Filella, I., Biel, C., Serrano, L., & Save, R. (1993). The reflectance at the 950–970 nm region as an indicator of plant water status. *International Journal of Remote Sensing*, *14*(10), 1887–1905.

Rascher, U., Damm, A., van der Linden, S., Okujeni, A., Pieruschka, R., Schickling, A., et al. (2010). Sensing of photosynthetic activity of crops. In E. -C. Oerke, R. Gerhards, G. Menz, & R. A. Sikora (Eds.), *Precision crop protection — The challenge and use of heterogeneity* (pp. 87–100). Dordrecht, Netherlands: Springer.

Richardson, A., Duigan, S., & Berlyn, G. (2002). An evaluation of noninvasive methods to estimate foliar chlorophyll content. *The New Phytologist*, *153*(1), 185–194.

Robert, C., Bancal, M., Ney, B., & Lannou, C. (2005). Wheat leaf photosynthesis loss due to leaf rust, with respect to lesion development and leaf nitrogen status. *The New Phytologist*, *165*(1), 227–241.

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of relieff and rrelieff. *Machine Learning*, *53*(1), 23–69.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Proc. 3rd Earth Resources Technology Satellite-1 Symposium. Greenbelt, MD NASA* (pp. 301–317).

Rumpf, T., Mahlein, A. -K., Steiner, U., Oerke, E. -C., Dehne, H. -W., & Plümer, L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, *74*, 91–99.

Sims, D., & Gamon, J. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, *81*(2), 337–354.

Steddom, K., Bredehoeft, W. M., Khan, M., & Rush, M. C. (2005). Comparison of visual and multispectral radiometric disease evaluations of *Cercospora* leaf spot of sugar beet. *Plant Disease*, *89*(2), 153–158.

Steiner, U., Bürling, K., & Oerke, E. -C. (2008). Sensor use in plant protection. *Gesunde Pflanzen*, *60*(4), 131–141.

Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationship with agricultural crop characteristics. *Remote Sensing of Environment*, *71*, 158–182.

Trocha, P., Daly, J., & Langenbach, R. (1974). Cell walls of germinating uredospores: I. Amino acid and carbohydrate constituents 1. *Plant Physiology*, *53*, 519.

Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. *The New Phytologist*, *186*, 795–816.

Ustin, S. L., Gitelson, A. A., Jaquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., et al. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment*, *113*, 67–77.

Wolf, P. F. J., & Verreet, A. J. (2002). The IPM sugar beet model, an integrated pest management system in Germany for the control of fungal leaf diseases in sugar beet. *Plant Disease*, *86*(4), 336–344.

Zarco-Tejada, P. J., Miller, J. R., Mohammed, G. H., Notlamd, T. L. L., & Sampson, P. H. (2001). Scaling-up and model inversion methods with narrow-band optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 1491–1507.

A.4 Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with support vector machines

Römer, C., Bürling, K., Hunsche, M., Rumpf, T., Noga, G., Plümer, L., 2011. Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with support vector machines. Computers and Electronics in Agriculture 79 (2), 180–188.

Abstract

Early recognition of pathogen infection is of great relevance in precision plant protection. Pre-symptomatic disease detection is of particular interest. By use of a laserfluoroscope, UV-light induced fluorescence data were collected from healthy and with leaf rust inoculated wheat leaves of the susceptible cultivar Ritmo 2-4 days after inoculation under controlled conditions. In order to evaluate pathogen impact on fluorescence spectra 215 wavelengths in the range of 370-800 nm were recorded. The medians of fluorescence signatures suggest that inoculated leaves may be separated from healthy ones, but high-frequency oscillations and individual reactions of leaves indicate that separability is difficult to achieve. The misbalance between the high number of measured wavelengths and the low number of training examples induces a high overfitting risk. For a pre-symptomatic pathogen identification a small number of robust features was desired which comprise most of the information relevant for the given classification task. Instead of choosing only the most relevant wavelengths, the coefficients of polynomials fitting the spectra were used for classification. They specify the global curve characteristics. Piecewise fitting by polynomials of fourth order led to high classification accuracy. Support Vector Machines were used for classification. Cross validation demonstrated that the achieved classification accuracy reached 93%. This result could be attained on the second day after inoculation, before any visible symptoms appeared. The described method is of general interest for pre-symptomatic pathogen detection based on fluorescence spectra.

Computers and Electronics in Agriculture 79 (2011) 180–188

Contents lists available at SciVerse ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with Support Vector Machines

Christoph Römer [a,*], Kathrin Bürling [b], Mauricio Hunsche [b], Till Rumpf [a], Georg Noga [b], Lutz Plümer [a]

[a] Institute of Geodesy and Geoinformation, Department of Agriculture, University of Bonn, Meckenheimer Allee 172, 53115 Bonn, Germany
[b] University of Bonn, Institute of Crop Science and Resource Conservation, Horticultural Science, Auf dem Huegel 6, 53121 Bonn, Germany

ARTICLE INFO

Article history:
Received 26 January 2011
Received in revised form 21 July 2011
Accepted 19 September 2011

Keywords:
Machine learning
Laser-induced fluorescence
Support Vector Machines
Feature extraction
Precision plant protection
Cereals' foliar diseases

ABSTRACT

Early recognition of pathogen infection is of great relevance in precision plant protection. Pre-symptomatic disease detection is of particular interest. By use of a laserfluoroscope, UV-light induced fluorescence data were collected from healthy and with leaf rust inoculated wheat leaves of the susceptible cultivar Ritmo 2–4 days after inoculation under controlled conditions. In order to evaluate pathogen impact on fluorescence spectra 215 wavelengths in the range of 370–800 nm were recorded. The medians of fluorescence signatures suggest that inoculated leaves may be separated from healthy ones, but high-frequency oscillations and individual reactions of leaves indicate that separability is difficult to achieve. The misbalance between the high number of measured wavelengths and the low number of training examples induces a high overfitting risk. For a pre-symptomatic pathogen identification a small number of robust features was desired which comprise most of the information relevant for the given classification task. Instead of choosing only the most relevant wavelengths, the coefficients of polynomials fitting the spectra were used for classification. They specify the global curve characteristics. Piecewise fitting by polynomials of fourth order led to high classification accuracy. Support Vector Machines were used for classification. Cross validation demonstrated that the achieved classification accuracy reached 93%. This result could be attained on the second day after inoculation, before any visible symptoms appeared. The described method is of general interest for pre-symptomatic pathogen detection based on fluorescence spectra.

1. Introduction

One of the most important limiting factors for crop production is the occurrence of foliar diseases (Oerke and Dehne, 2004). In the specific case of wheat, one of the three most cultivated cereal species worldwide, significant yield and quality losses are induced by the leaf rust pathogen *Puccinia triticina*. In recent years, the development of technologies for the precise and site-specific pesticide application has contributed decisively for a more environment-friendly food production. In this context, detection and differentiation of diseases as early as possible plays an important role for the precise crop protection (West et al., 2003). In addition to reflectance and hyperspectral measurements, fluorescence approaches have been developed under laboratory and field conditions for the fast and non-destructive early detection and identification of diseases (Kuckenberg et al., 2009; Bürling et al., 2010, 2011; Lüdeker et al., 1996; Chaerle et al., 2004) or weed (Panneton et al., 2010). In this context the spectral resolved fluorescence signature, which is still used under laboratory conditions, provides valuable information for the pre-symptomatic detection of diseases and the differentiation between susceptible and resistant cultivars (Bürling et al., 2011).

As generally known, excitation of green leaves with ultraviolet light (UV) induces two types of fluorescence signals: in the range of 400–630 nm the blue–green (BGF) and yellow fluorescence which originates from a large number of fluorophores, mainly polyphenols (Cerovic et al., 1999; Lang et al., 1991), and in the range of 630–800 nm, the chlorophyll fluorescence (ChlF). Fluorescence signals depend on plant species and genotypes, and the occurrence of stresses might induce modifications in the amount and composition of fluorophores in the leaf tissue, impacting the typical fluorescence signature in direct and indirect ways. Especially in terms of biotic stresses, plants might accumulate salicylic acid and phenylpropanoid compounds, which are key substances in plant disease resistance (Vermerris and Nicholson, 2006). Furthermore, several preformed defense substances, such as phenol and its derivates, lactones, saponines and stilbenes, are involved in various host–pathogen interactions (Agrios, 2005). Moreover, fungi itself might show a specific auto-fluorescence as indicated

* Corresponding author. Tel.: +49 228 731759; fax: +49 228 731753.
E-mail addresses: roemer@igg.uni-bonn.de (C. Römer), MHunsche@uni-bonn.de (M. Hunsche).

0168-1699/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2011.09.011

by a UV-light induced bright blue fluorescence from rust fungal structures (Zhang and Dickinson, 2001).

Until now, the use of laser-induced fluorescence for detecting pathogens on agricultural crops focused on measurements at red (R) and far-red (FR) wavelengths of the ChlF, being the changes of R/FR fluorescence ratios related to the chlorophyll content i.e. chlorophyll breakdown (Kuckenberg et al., 2009). However, only limited information on leaf rust triggered changes in the blue–green range of the visible spectra of wheat leaves is available. In this context, experiments have been conducted to evaluate changes of the fluorescence induced by the pathogens *P. triticina* (Bürling et al., 2011). As a common and recommended procedure, ratios of amplitudes and half-bandwidth of pre-defined peaks are calculated for a reliable comparison of treatments (Cerovic et al., 1999; Buschmann, 2007). However, the question remains if it is possible to extract more information from a high resolved spectral dataset in order to give more precise information about disease occurrence and development, as already novel algorithms were applied on spatially resolved fluorescence signatures to reveal early symptoms of plant response (Berger et al., 2007). Potential applications of improved evaluation models rely on the quantification of the impact of single stresses on plants as well as the discrimination between multiple stress factors. Finally, a precise data evaluation for future applications in the context of precision agriculture should support a reliable and site-specific detection and differentiation of stress factors. Hence, the approach of the present study was to exploit the suitability of a different data evaluation method for a pre-symptomatic and accurate differentiation between healthy and leaf rust infected wheat leaves.

Support Vector Machines (SVMs) is a sophisticated machine learning method (Vapnik, 1998; Schölkopf and Smola, 2002) which is very robust against overfitting (i.e. the description of noise and random deviations instead of underlying pattern characterizing different classes), and able to learn accurate classification models with a small number of training data. SVMs are commonly adopted for the classification of spectral images in remote sensing (Bruzzone and Camps-Valls, 2009; Pal and Mather, 2005) since they have superior ability to deal with the so called 'curse of dimensionality'. In the context of plant sciences, Ferreiro-Arman et al. (2006) applied SVMs to hyperspectral airborne data in order to distinguish between grape varieties. Camargo and Smith (2009) identified visual symptoms of plant diseases with Support Vector Machines in images. Furthermore, Zheng et al. (2010) processed chlorophyll fluorescence recordings and other nutritional parameters in order to improve the classification of jujube according to its postharvest quality and storage potential by means of SVMs.

The potential of SVMs for precision agriculture was recently reviewed (Mucherino et al., 2009; Huang et al., 2010). Based on reflectance measurements, SVMs have been used for the automatic classification of weeds and levels of nitrogen fertilization of corn (Karimi et al., 2006). Moreover, when applied to hyperspectral data for discrimination of foliar diseases, the superiority of this classification method as compared to Decision Tree and Artificial Neural Networks, was demonstrated (Rumpf et al., 2010).

The importance of the pre-symptomatic detection of diseases for precise crop protection is evident. On the other hand, the minor signal differences between healthy and inoculated classes at this early stage impose a challenge for a precise classification. As a rule, processing of spectrally resolved fluorescence measurements results in a large number of features. In case of a small number of training data, a misbalance to the number of features (i.e. the curse of dimensionality) and the associated high risk of overfitting, might occur (Hughes, 1968).

Besides the classifier, the features used to describe the samples have a great impact on the classification accuracy. Well-established ratios use the fluorescence intensity at specific wavelengths of the spectrum as basis for interpretation of stress-induced physiological modifications of plant tissues. However, there may be still unknown structure which enables to separate classes, e.g. healthy from infected leaves.

Spectrally resolved fluorescence measurements give the opportunity for calculation of many more ratios as those usually adopted in such studies. However, overlapping classes and the limited size of the dataset make it difficult to use only single wavelengths for discrimination. It would therefore be preferable to use the complete information of the spectrum while reducing the number of features at the same time. For this, polynomials are used to fit the curve whereupon the coefficients are more robust against random deviations and might contribute for a more reliable pathogen detection. We raise he hypotheses that (1) a differentiation between healthy and leaf rust inoculated leaves is possible at a pre-symptomatic stage with a high classification accuracy when using sophisticated machine learning methods like Support Vector Machines and (2) reducing the dimension of feature space to a small number of features which comprise the information about the whole spectrum yields more robust and accurate classification results.

2. Material and methods

2.1. Plant material and growth conditions

Experiments were conducted in a controlled-environment cabinet simulating a 14-h photoperiod with 200 μmol m^{-2} s^{-1} photosynthetic active radiation (PAR; Philips PL-L 36W, Hamburg, Germany), temperature of 20/15 ± 2 °C, and relative humidity of 75/80 ± 10%. Winter wheat (*Triticum aestivum* L. emend. Fiori. et Paol.) seeds (n = 5 per pot) of the rust-susceptible cultivar Ritmo (susceptibility degree according to the descriptive variety list of the German Federal Plant Variety Office, 2008, SD = 8) were sown in pots (0.44 L) filled with perlite. Emerging plants were provided with Hoagland nutrient solution. Twenty days after sowing, when plants had reached the second-leaf stage, twelve pots (six for inoculation, six as control) were selected for the experiment. Leaf rust inoculation was done on the youngest fully expanded leaf of two plants per pot. The experiments were repeated three times under similar conditions, and the results were pooled combined to a data pool for the statistics.

2.2. Inoculation of P. triticina

Inoculation was accomplished with a non-specific mixture of *P. triticina* spores produced on wheat plants without known resistance genes (Department of Phytomedicine, University of Bonn). Before each experiment, fresh *P. triticina* spores were suspended in a solution of distilled water + Tween 20 (0.01% w/v; Merck-Schuchardt, Hohenbrunn, Germany). The spore concentration was estimated microscopically with a Fuchs–Rosenthal counting chamber and adjusted to 1×10^4 spores per ml. On each leaf, the middle of the leaf length was marked on the adaxial side with a felt tip pen, and seven 6-μl droplets of spore suspension were applied in a row on one leaf half. Prior to the application, leaves were fixed horizontally on a sample holder to prevent droplet run-off. Control plants were handled similarly but treated with droplets of distilled water + Tween 20 (0.01%). During the inoculation period (22 h), plants were maintained in the climate chamber at almost saturated atmosphere under a plastic cover. Thereafter, the plastic cover was removed and the leaves were released from their horizontal fixation.

2.3. Fluorescence measurements

Fluorescence measurements were carried out using a compact fiber-optic fluorescence spectrometer with nanosecond time

resolution by the boxcar technique (IOM GmbH, Berlin, Germany). A pulsed N_2-laser (MNL100, LTB Lasertechnik Berlin GmbH, Berlin, Germany) with emission wavelength of 337 nm and a repetition rate of 20 Hz served as excitation source. The fiber-optic probe for detection of fluorescence signals consisted of a central excitation fiber and six surrounding emission fibers, each with a 200-μm diameter. The pulse energy at the probe exit was adjusted to be 1–2 μJ with a pulse length of approximately 1 ns resulting in a density of 5.5×10^{15} photons per cm^2 and impulse. As observed in preliminary tests (unpublished data), these settings did not cause any significant change in the blue–green and chlorophyll fluorescence intensity and lifetime of wheat leaves even when samples were sequentially analyzed for more than 40 min. Fluorescence was recorded with an acousto-optic tunable filter (AOTF) monochromator, which enabled a minimal step width of 1 nm between 370 and 800 mm. A photomultiplier (PMT, H5783-01, Hamamatsu, Hamamatsu City, Japan) was used as detector. The sensitivity of the PMT was adjusted in order to optimize the signal intensity during the spectral measurements. Time resolution was accomplished using a gated integrator with a 2-ns half-width; the position of the gate could be set with an accuracy of 0.1 ns.

Detection of fluorescence spectra was done on leaves fixed horizontally on a plate with integrated sample holder. The fiber-optic probe was positioned in a 90° angle to the leaf. Thereby, with help of a laser-based rangefinder (OptoNCDT 1300; Micro-Epsilon Messtechnik GmbH & Co. KG, Ortenburg, Germany) fixed beside the probe, the distance between leaf and probe surface was adjusted. The standard distance enabled fluorescence intensities in a narrow data range, providing a minimum of signal intensity and avoiding signal saturation. Spectra were measured at 21–23 °C under ambient light conditions (about 18 μM $m^{-2}s^{-1}$ PAR) during the pre-symptomatic stage from 2 to 4 days after pathogen inoculation (dai). Only on the 4th dai first few loom chlorotic spots could be observed at the inoculation site. Before fluorescence measurements plants were adapted for 0.5 h to room conditions. For optimization of fluorescence signals, equipment settings were adjusted as follows. The pulse count was set on 32, which is the number of laser pulses averaged for each single data point. The gate position was at 3 ns (in the temporal signal maximum) and the wavelength interval was of 2 nm. Measurements were accomplished at a distance of 3.95 mm between the probe and sample, and a PMT gain of 600 V.

2.4. Characteristics of the fluorescence dataset

The described experimental design lead to a dataset of 36 leaves inoculated with leaf rust and 36 healthy leaves which were observed on day 2, 3 and 4 after inoculation, resulting in three datasets per leaf. In consequence of the equipment settings with measurements between 370 and 800 nm and step width of 2 nm, 215 observations at different wavelengths were produced.

Firstly, the medians of both classes were analyzed, which, in contrast to the mean, is robust against outliers and noise. Comparison of both medians in Fig. 1 suggests that it is feasible to differentiate between healthy and inoculated leaves at the peak near the wavelength of 450 nm and the local maxima near 700 and 750 nm. In a next step, the deviations of the classes were taken into account. Because of high-frequency oscillations, the natural variability of the leaves, and differences in reaction intensity to the pathogen inoculation, both classes have a high standard deviation, which is even higher than the difference between the medians (Fig. 1, bottom picture).

Besides the apparently large difference between both classes in the wavelength range between 430 and 470 nm, this is also the area with the biggest variance. In contrast, the area between 540 and 630 nm has a small variance, while the medians of both classes are very close to each other. This boundary condition is especially strong for dai two and gets weaker till the fourth dai.

In addition to the demanding ratio between class variance and the difference of the medians, the dataset is highly prone to overfitting due to its misbalance between the number of features and the number of samples.

Therefore, a classification method which enables to deal with those challenging boundary conditions and significantly reduces the risk of overfitting was necessary. Support Vector Machines are a classification method with statistical properties which enable to fulfill this task.

2.5. Support Vector Machines

Under the assumption that both classes are linearly separable, SVMs use hyperplanes (Fig. 2, left) of the form

$$\langle \boldsymbol{\omega}, \mathbf{x} \rangle + b = 1$$

for classification. The normal vector $\boldsymbol{\omega}$ denotes the orientation of the hyperplane and b the distance from the origin. The hyperplane divides feature space into two subspaces. Samples $\mathbf{x}$ are classified depending on the subspace they are in, i.e. on which side of the hyperplane the sample is:

$$f(\mathbf{x}) = \operatorname{sgn}(\langle \boldsymbol{\omega}, \mathbf{x} \rangle + b), \tag{1}$$

It is necessary to exploit which hyperplane leads to the best classification performance. Therefore, the training data of both classes is needed. Following the concept of structural risk minimization, Vapnik (1998) proofs that the hyperplane leading to zero classification errors in the training set which at the same time has the maximal margin is best. The margin is the distance between the hyperplane and its closest training sample. The distance is given by $1/||\boldsymbol{\omega}||$.

The training samples are given by an N-dimensional vector $\mathbf{x}_i$, $i = 1 \ldots m$. Each sample of the training dataset $\mathbf{x}_i$ has a corresponding class label $y_i \in \{\pm 1\}$. The optimal hyperplane is constructed by solving the following optimization problem:

$$\min_{\omega} h(\mathbf{x}) = \frac{1}{2}|\boldsymbol{\omega}|^2, \tag{2}$$

$$\text{subject to } y_i(\langle \boldsymbol{\omega}, \mathbf{x} \rangle + b) \geq 1, \forall i. \tag{3}$$

Constrained optimization problems (Boyd and Vanderberghe, 2004) like this are solved with the help of the Karush–Kuhn–Tucker conditions (Rockafellar, 1993), leading to the dual Lagrangian problem:

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \tag{4}$$

$$\text{subject to } \alpha_i \geq 0 \tag{5}$$

$$\text{and} \sum_{i=1}^{m} \alpha_i y_i = 0. \tag{6}$$

Here α denotes the dual Lagrangian variable. During the optimization process only specific points, called support vectors, have an assigned α that is greater than zero. With the help of the dual Lagrangian the maximal margin hyperplane is constructed with

$$\hat{\boldsymbol{\omega}} = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i \tag{7}$$

$$\text{and } \hat{b} = -\frac{1}{2} \langle \hat{\boldsymbol{\omega}}, \mathbf{x}_a + \mathbf{x}_b \rangle, \tag{8}$$

where $\mathbf{x}_a$ and $\mathbf{x}_b$ are any support vectors of both classes. As only support vectors have α greater than zero, no other vector has any influence on the calculation of the normal vector $\boldsymbol{\omega}$.

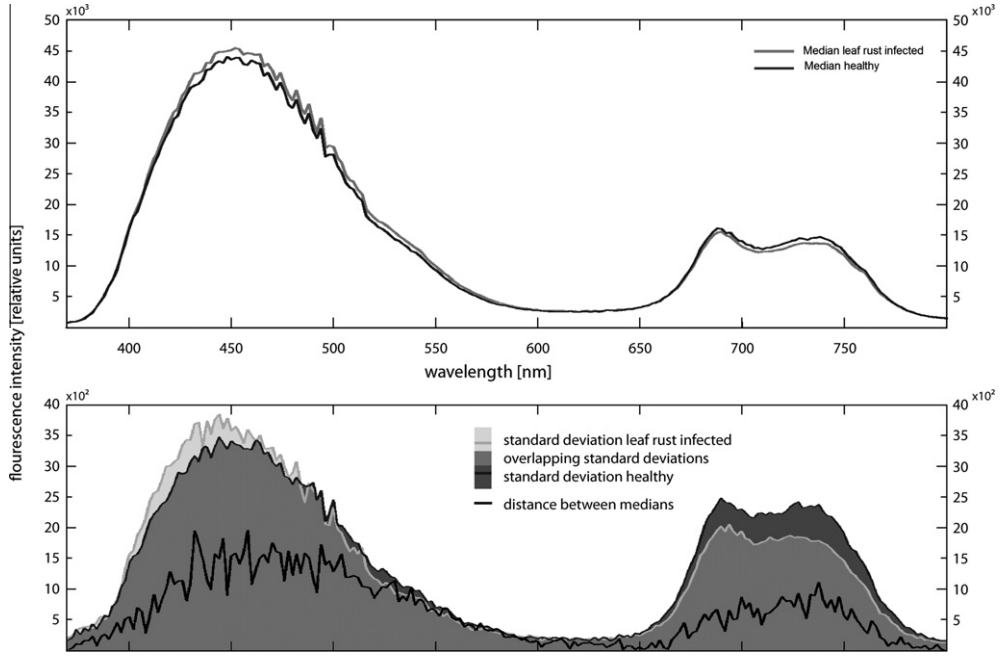


Fig. 1. Medians of healthy and inoculated wheat leaves (top), and comparison of median difference with standard deviations at dai 2 (bottom).

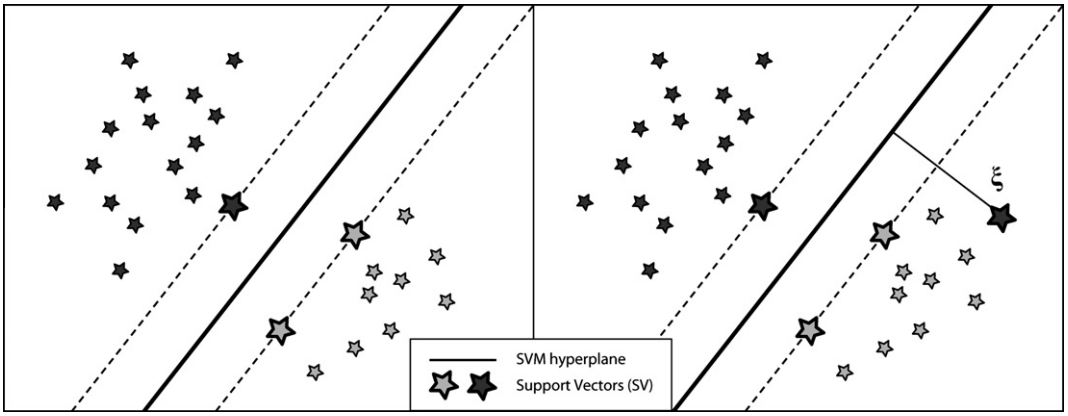


Fig. 2. Maximal margin hyperplane separating the class of dark stars from the class of bright stars on the left. Soft Margin SVM with one wrongly classified dark star on the right.

So far it is assumed that the problem is linearly separable, though this in fact is not the case for the fluorescence dataset. Both classes are strongly overlapping. Therefore, classification errors in the primal optimization problem (2) have to be allowed as no hyperplane exists which leads to zero misclassifications (Fig. 2, right). Hence, the optimization problem is modified to:

$$\text{Maximize } h(\boldsymbol{\omega}) = \frac{1}{2}|\boldsymbol{\omega}|^2 + C\sum_{i=1}^{m}\xi_i, \tag{9}$$

$$\text{subject to } y_i(\langle \boldsymbol{\omega}, \mathbf{x}\rangle + b) \geqslant 1 - \xi_i, \forall i \tag{10}$$

$$\text{and } \xi_i \geq 0, \forall i. \tag{11}$$

C weights the influence of misclassified samples on the optimization process. Apart from an additional boundary on constraint (5) to $C \geq \alpha_i \geq 0$, the dual optimization problem and the calculation of the hyperplane remains unchanged.

Hence, the only parameter to be optimized is C which weights the slack variable ξ in Eq. (9). The parameter C leading to the best classification performance is a priori unknown. Therefore, different parameters have to be tested and the one resulting in the best prediction accuracy should be chosen. Evolutionary algorithms (Goldberg, 1989) are well suited for this task.

The LibSVM tool from Chang and Lin (2001) was used to train the SVM model evaluated and presented in this paper.

3. Results

3.1. Classification on all wavelengths

In a first step the features had to be normalized to a mean of zero with a standard deviation of one. This is important as otherwise features with large fluorescence values have more weight in Eq. (7) and on the indicator function in Eq. (1) as features with a smaller scale.

A linear soft margin SVM was trained on the original fluorescence spectrum. In the following, results were evaluated using a 10-fold cross validation (Kohavi, 1995). All parameters (in this

Table 1
Classification accuracy of an SVM using all wavelengths as features, 2, 3, and 4 days after inoculation (dai).

Dai	SVM (%)
2	73.6
3	79.2
4	80.7

and in following sections) were optimized using an evolutionary optimization approach (Goldberg, 1989).

The classification performance of correctly classified leaves was 73.6% for dai 2, 79.2% for dai 3 and 80.7% for dai 4 (Table 1). These results clearly indicate that both classes are separable using an SVM classifier, even though the accuracy is still not sufficient.

However, if many features do not separate well, their sum may outvote the good features in Eq. (1), which has a negative effect on the classification accuracy. A classical method to solve this problem would be to reduce the number of features with feature ranking algorithms (Kohavi and John, 1997). Those evaluate the relevance of each feature for the classifier. The better both classes can be distinguished in a single feature, the higher it will be ranked.

3.2. Relevant wavelength evaluation

Feature evaluation was done with a statistical variance analysis of the wavelengths. The separability is measured with a hypothesis test (Koch, 1999), where the null hypothesis is that the medians of both classes belong to a different distribution. The exponent of the *p*-value is a measure for the quality of the separation between classes. As only 1.63% of the features are normally distributed, according to a Lilliefors test (Lilliefors, 1967) with 95% significance value, the non-parametric Wilcoxon ranksum test (Wilcoxon, 1945) was used for the hypothesis test.

The test reveals that the wavelengths between 550 and 630 nm are feasible for a separation between healthy and inoculated leaves, whereas the area from 650 to 800 nm has only very limited impact (Fig. 3). The peak between 450 and 500 nm has only a small influence on dai 2, but its importance raise at dai 3 and dai 4.

Nevertheless, due to the misbalance between the number of features and the number of samples the features may only randomly work well because of the high frequency oscillations and the narrow classes. Therefore, the results obtained could be a coincidence in this special case. Instead of getting a robust prediction the risk of overfitting would consist.

3.3. Feature construction by robust curve fitting

The number of features is too high for the available amount of data. Therefore the question arises if the information from all 120 wavelengths could not be expressed with a handful of features which contain most of the information of the spectrum. In addition, the data is very noisy and hence should be smoothed for more reliable observations.

The starting point of our approach is the observation that the spectrum is compounded by several parabolas, which can be described by polynomials of the third order. The polynomial coefficients describe the characteristics of a certain part of the curve and at the same time they smooth the spectrum, reducing the influence of noise. The polynomials were calculated by a least square fitting approach.

Therefore, in a further step, polynomial coefficients which describe the characteristics of a certain part of the curve were used (as a side note: the polynomials were derived on the original datasets, not the normalized ones from Section 3.1., therefore they reflect relative amplitudes. Only the coefficients for the SVM model (Section 3.4.) need to be normalized again).

As outcome, the number of features is reduced to the order of the polynomial. At the same time, the coefficients describe the characteristics of the curve in a more robust way with regard to single outliers and noise, as they have only a limited impact on the coefficients. With these coefficients as features the classifier differentiates between healthy and inoculated leaves.

The problem of oscillating coefficients at polynomials of a high order is avoided by using splines for the approximation of the curve. Fig. 4 shows the approximation with a B-spline (De Boor, 1978) of third order. The important point to be addressed was the question of where the knots between the polynomials of the spline had to be placed.

In a first step, four parabolas were derived with the knots at the wavelengths of 392, 514, 664 and 706 nm. However, as shown in Fig. 3 the wavelengths from 380 to 400 nm have a significant relevance for the classifier; although the error in the approximation would be very high in this area when only fitting the parabolas. Hence, an additional knot was included at 392 nm which significantly reduced the approximation error. A more accurate approximation would be possible by using more knots, even though the aim was to describe areas which are as large as possible with just one piecewise polynomial in order to get coefficients that are robust to noise.

However, the approximation error increases at the knots of the spline even if only the form parameters of the piecewise polynomial are needed. Therefore, instead of using a B-spline for the whole curve, it is split into several smaller curves and the

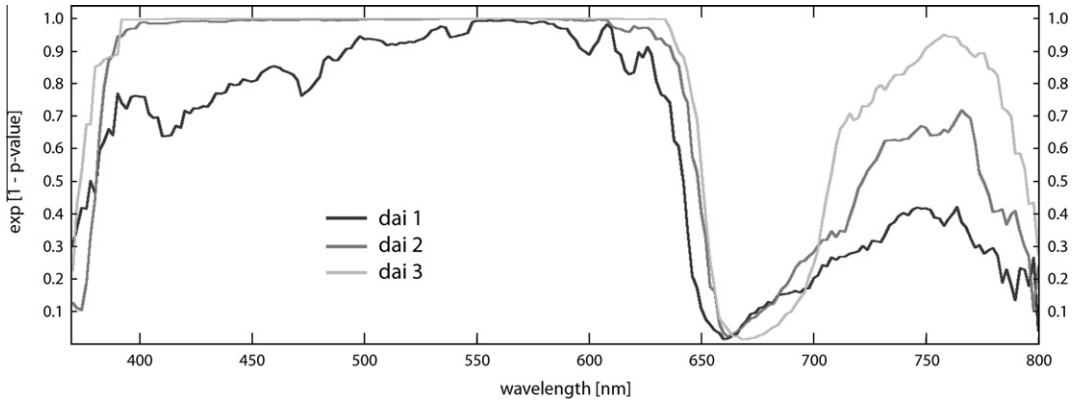


Fig. 3. Calculation of relevant wavelength using the Wilcoxon test.

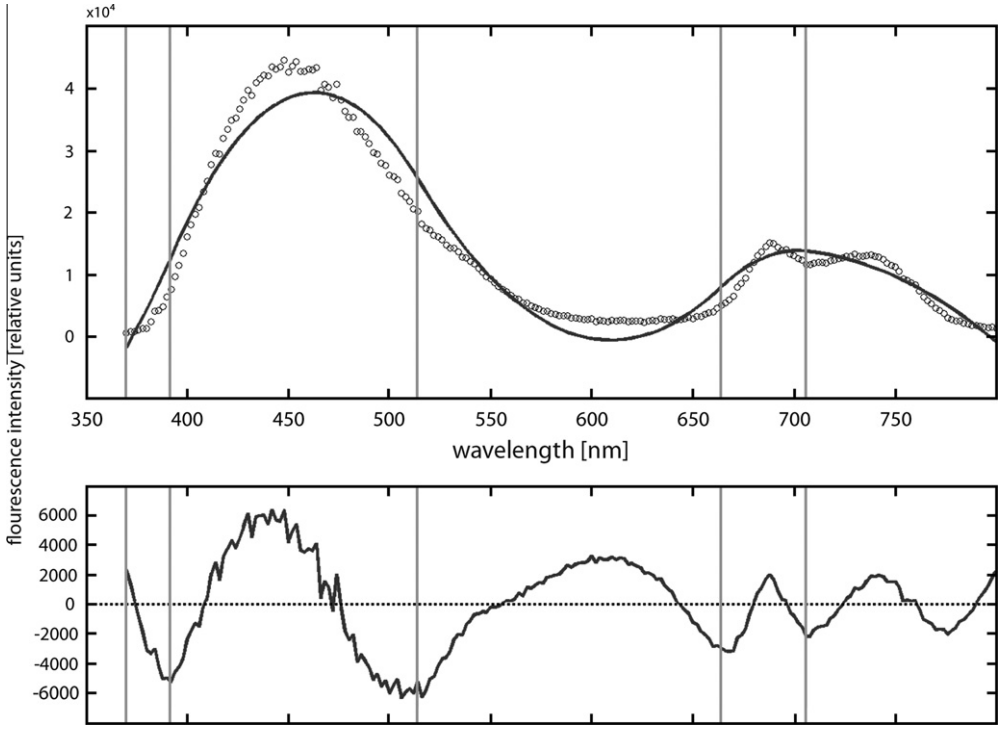


Fig. 4. Quadratic B-spline through the whole fluorescence spectra (top) and the corresponding approximation error (bottom).

polynomials are then calculated only for that part of the curve without the condition of smoothness at the knots. It is expected that the largest errors will occur at the ends of the piecewise polynomials. Once more the knots were placed such that the most relevant wavelengths calculated in the previous section are approximated with the smallest possible error (hence even though the Wilcoxon test applied in Section 3.2. may not be optimal for datasets with high redundancies, it still gives us a very good impression about the areas of interest). Additionally the knots should not be at the chlorophyll fluorescence peaks at about 450, 680 and 740 nm. If points at the end of the curve pieces will not fit into a parabola and are not of high relevance, those points are ignored and not fitted at all in order to reduce the approximation error of the more important wavelengths. The piecewise approximation reduced the standard deviation to roughly one third as compared to the whole approximation, as displayed in Fig. 5.

3.4. Results for Support Vector Machines with piecewise polynomials

The use of third order piecewise polynomials as features for an SVM classifier resulted in 79.2% correctly classified samples for dai two, 77.8% on dai three, and 87.5% on the fourth dai (Table 2). Even if the results dropped by 1.2% on dai three, they improved by 5.55% on dai two and even 6.78% on the fourth dai as compared to the results trained with all features (Table 1).

Further improvements can be reached by giving more weight to features with a high relevance for the classifier. With this aim the algorithm developed by Guyon et al. (2002) is used to rescale the normalized features according to their corresponding weight, leading to a greater gap between features with a high and low relevance.

Comparisons between unweighted and weighted SVMs indicate a significantly improved classification for the weighted data, ranging between 4.2% (dai 4) and 9.7% (dai 3), as indicated in Table 2. Polynomials of higher order approximate the data more accurately but on the other hand they are more sensitive to noise. In order to evaluate this relation, the results were compared with polynomials of higher, up to the 6th order.

For the second dai a 4th order polynomial proved to yield the best results. The classification performance was of 93.1%, which was approximately 9% higher as compared to the 3rd order polynomial (Table 3). Furthermore, both the 5th and 6th order polynomials revealed higher performance than the 3rd order polynomial. The highest classification at dai three (90.3%) was achieved with a polynomial of the 6th order (Table 3). Processed fluorescence data of dai 4 revealed the highest classification performance with the polynomial of third (91.7%) order, whereas the percentage of correctly classified leaves dropped with polynomials of higher order. Overall, the 4th order polynomial leads to the best results for all 3 days after inoculation.

4. Discussion

The aim of our study was to evaluate the potential of machine learning algorithms applied to spectrally resolved fluorescence signatures for improved, pre-symptomatic leaf rust detection on wheat, as a representative pathosystem for foliar diseases of cereals. For this purpose fluorescence signals were recorded from 370 to 800 nm with 2 nm resolution on healthy and inoculated leaves. High noise, high-frequency oscillations and particularly great variability between the leaf samples of the same class resulted in a challenging scenario. Therefore an advanced classification algorithm was needed, while the curse of dimensionality called for extraction of robust, highly informative features.

The use of SVMs on the fluorescence data enabled a classification accuracy of 79%, 78%, and 87% at 2, 3, and 4 dai, respectively (Table 2). Weighted SVMs increased the classification accuracy in a range of 4–9%, depending on the evaluation day. Especially in the context of precision crop protection the early and precise detection of stressors plays a key role. In previous studies, ratios between chlorophyll fluorescence data measured at 690 and 740 nm on wheat leaves infected with leaf rust or powdery mildew

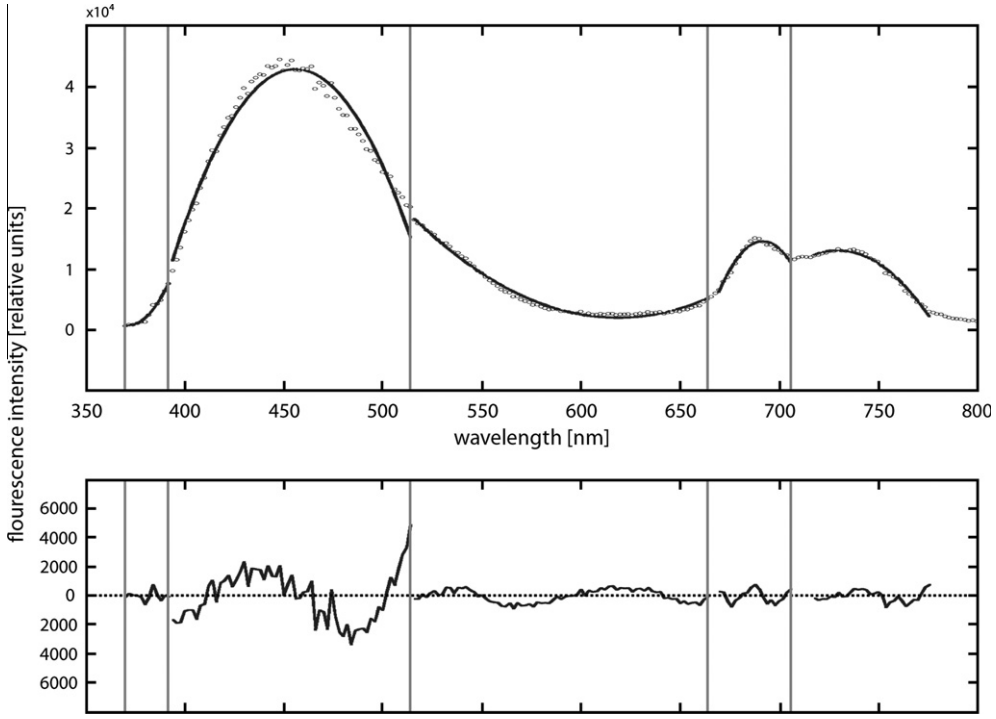


Fig. 5. Polynomials of 3rd order calculated only on certain ranges of the fluorescence spectra.

Table 2
Comparison between a weighted and an unweighted SVM for day 2, 3, and 4 after inoculation (dai) using 3rd order, piecewise polynomials.

Dai	SVM (%)	Weighted SVM (%)	Difference (%)
2	79.16	84.22	+5.06
3	77.78	87.50	+9.72
4	87.50	91.67	+4.17

Table 3
Influence of the order of the polynomials on the classification accuracy 2–4 days after inoculation (dai).

Dai	Weighted SVM 3rd order (%)	Weighted SVM 4th order (%)	Weighted SVM 5th order (%)	Weighted SVM 6th order (%)
2	84.22	**93.05**	88.89	91.67
3	87.50	88.89	84.72	**90.28**
4	**91.67**	90.28	87.50	79.17

yielded a classification of 25% and 50% at 7 and 8 dai, respectively (Kuckenberg et al., 2009). Aiming to improve even more the quality of the proposed method, piecewise approximation by polynomials of 3rd, 4th, 5th, and 6th order were compared. As outcome a classification accuracy of 93% at 2 dai by using weighted SVM and polynomials of 4th order was achieved (Table 3).

As a tool for quality control of the methodological approach, SVMs were compared with Decision Trees and Artificial Neural Networks (ANN), two machine learning algorithms commonly adopted in agriculture (Huang et al., 2010). The ANN was a feed-forward neural network trained by a backpropagation algorithm. For this purpose the dataset with all wavelengths and the dataset of polynomial coefficients of 4th order, recorded 2 days after inoculation were used. Both, Decision Trees and ANN, classification algorithms train the decision function with respect to an empirical loss function in contrast to the structural risk minimization which are used by SVMs (Vapnik, 1998). While empirical risk minimization converges to the optimum for infinite samples (Schölkopf and Smola, 2002) it may be prone to overfitting on small training datasets and therefore result in worse performance as compared to SVMs. This is an important factor for the classification task on this dataset because even if the dimension of feature space was reduced to 20 for 4th order polynomials, this is still a huge number of features compared to the number of training samples. Hughes (1968), for example, suggested using more than 500 samples for 20 features. Hence, the number of samples can be considered as very small. On the other hand a strong restriction of the number of used features might reduce the robustness and accuracy of the classifier, as the remaining features may not reveal enough information to describe the underlying patterns discriminating both classes.

In order to calculate Decision Trees and ANNs, the RapidMiner (Mierswa et al., 2006) environment was used. As shown in Table 4, the SVMs yield a more precise classification than Decision Trees and Neural Networks on both the original dataset and the pre-processed dataset using only the polynomial coefficients of 4th order. This underlines the theoretical considerations that SVMs are well suited for small training datasets. The superiority with respect to the curse of dimensionality triggered the use of SVMs for improved evaluation of spectral data in many remote sensing applications (Bruzzone and Camps-Valls, 2009).

Apart from the theoretical decision to choose polynomial coefficients to avoid overfitting, the results were compared with two other classical methods to reduce the dimensionality of feature space. For this purpose a Principal Component Analysis (Bishop, 2006) and a feature selection algorithm like Relief (Kononenko,

Table 4
SVMs compared with Decision Trees and Artificial Neural.

Dataset	SVMs (%)	Decision Trees (%)	Artificial Neural Network (%)
All wavelengths	73.61	70.83	63.89
Polynomials of 4th order	**93.05**	61.96	82.50

Table 5
Results for SVMs trained on different, 20 dimensional features spaces.

	Polynomial (%)	Principal components (%)	Best 20 features (%)
Dai 2	**93.05**	72.22	72.22

1994), from which the most relevant features were selected for the classifier, were chosen. Both PCA and Relief was done for 2 dai. In case of the PCA the optimization process was the same as for the polynomials, i.e. the principal components were weighted with an SVM weighting before learning the model and parameter optimization was done with an evolutionary algorithm, except that the number of principal components was an additional parameter, ranging from 1 to 20. The optimal cross-validation accuracy was achieved for 4 principal components (reconstruction accuracy of 89%). In case of Relief the features were weighted with the corresponding Relief weights. For evaluation the k best features were selected before training the classification model, where k was ranging from 1 to 110. The best result was achieved for 24 features with 83% accuracy. Using more features resulted in a slightly weaker performance, even though the impact was small. Apparently more features do not contain additional information.

Apart from theoretical considerations, the outcome of the comparison show that the polynomial representation yields higher classification accuracy than the other methods (Table 5).

Another possible option is the choice of the kernel for the SVM. We have consciously applied a linear kernel. The obvious alternative would be an RBF kernel, the most popular non-linear kernel. In fact we compared both options and found that the results achieved with an RBF kernel were weaker (82% classification accuracy for 4th order polynomials on dai 2) than with the linear case (93%). Above, the RBF kernel model is more complex than the linear model, and the risk of overfitting increases with the size of the model. For these reasons we adopted the linear kernel.

The impact of different curve fitting algorithms and their suitability for the used classifier might have further potential for the improvement of the classification accuracy. In addition, the residuals of the approximations could yield additional information. As displayed in Fig. 5, residuals between the wavelengths of 520–660 nm still have significant structure, though ideally only white noise should remain after the approximation. Hence, the question arises, whether the distribution of residuals exhibits high frequency patterns relevant for class separation.

In order to exploit the residuals for a further accuracy improvement, it would be necessary to approximate different parts of the spectral signature with the optimal polynomial order for the specific part. Hence, computational efficient algorithms which evaluate the quality of the approximation with respect to the classification performance instead of just the approximation error are needed. In most cases, classification algorithms derived by machine learning approaches are based on observable features and do not incorporate knowledge on physiological modifications due to pathogen–plant interactions.

In response to pathogen attack, the synthesis and accumulation of phenolic compounds at the infection site is stimulated. Such alterations in plant secondary metabolism are time-dependent and may result in changes of specific fluorescence intensities. As shown before, the establishment and development of fungus on plant tissues leads to higher fluorescence intensities in the blue–green spectral region (Lüdeker et al., 1996; Bélanger et al., 2008). In the present study, the determination of relevant wavelengths with the Wilcoxon test indicate a higher relevance of the blue–green spectral region for the classification accuracy as compared to the red and far-red region (Fig. 3). This might be one reason for the higher classification accuracy in our study using UV-induced spectral data as compared to red-light induced chlorophyll fluorescence data (Kuckenberg et al., 2009). Results of the Wilcoxon test at 2 dai indicate the highest relevance of the green as compared to the blue spectral range for data classification (Fig. 3). Accordingly, a more pronounced raise of green fluorescence as compared to the increase in blue fluorescence was observed early after inoculation (Bürling et al., 2011). Moreover, in the spectral region between 560 and 620 nm longer mean lifetime was measured as result of pathogen infection (Bürling et al., 2011). In the time course from 2 to 4 dai the importance of all wavelengths increased and thereby the relevance of blue aligned to the relevance of green (Fig. 3).

Despite the discussed reasons for changes in fluorescence signals due to fungal attack as well as the suitability of SVMs for detection of fungal infection, further investigations on host–pathogen interaction by means of HPLC or microscopic analysis should be initialized. It is well known that the environmental conditions such as the absolute values as well as the oscillations in temperature and radiation intensity might influence the composition and the spectral characteristics of plant tissues (Buschmann et al., 2008). Thus, with regard to robustness and classification accuracy, several questions still remain. Is there an influence of the measurement set-up as well as between laboratory and field conditions? How does classification accuracy depend on the leaf structure and external factors influencing plant physiology and pathogen development? Is the described method robust with regard to such disturbances? In the present study wheat plants of the cultivar Ritmo were grown under controlled conditions and inoculated with *P. triticina* according to a predefined inoculation protocol. Under other circumstances (e.g. different cultivar, growth conditions, pathogen strain, etc.) the fluorescence signature of both healthy and inoculated plants might differ from those observed in this experimental series. However, the evaluation tools presented here for plants cultivated under well-defined conditions establish the primary basis for more complex systems. Moreover, the stepwise development of the approach by adding further parameters enables to better understand the relations step by step for a better interpretation of field data.

The method described in this paper is not specific for the experimental conditions described and therefore adaptable to further scientific purposes in plant sciences. A special application of the method in the context of precision agriculture relays on the objective and precise pre-symptomatic detection of pathogens, or to differentiate between changed fluorescence signatures induced by different pathogens. Similarly, the impact of additional stress factors such as nutrient deficiency or drought stress might be exploited. Finally, the proposed data evaluation might be used to differentiate new genotypes in plant breeding lines according to their susceptibility to single or multiple stress events.

5. Conclusion

As a result, polynomial fitting of fluorescence signatures together with Support Vector Machines revealed to be a powerful method to improve the early detection of plant diseases. The results achieved demonstrate a superior potential for handling a demanding signal to noise ratio and a high level of robustness. The method described in this paper is independent from the specific instance studied here and is adaptable to any spectral signature. Further studies should explore the applicability of this method to additional stress factors on economically relevant plants species.

References

Agrios, G.N., 2005. Plant Pathology. Elsevier Academic Press, New Delhi, pp. 922.
Bélanger, M.-C., Roger, J.-M., Cartolaro, P., Viau, A.A., Bellon-Maurel, V., 2008. Detection of powdery mildew in grapevine using remotely sensed UV-induced fluorescence. Int. J. Remote Sens. 29 (6), 1707–1724.

Berger, S., Benediktyová, Z., Matous, K., Bonfig, K., Mueller, M.J., Nedbal, L., Roitsch, T., 2007. Visualization of dynamics of plant–pathogen interaction by novel combination of chlorophyll fluorescence imaging and statistical analysis: differential effects of virulent and avirulent strains of *P. syringae* and of oxylipins on *A. thaliana*. J. Exp. Bot. 58 (4), 797–806.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York. USA.

Boyd, S., Vanderberghe, L., 2004. Convex Optimization, Seventh ed. Cambridge University Press, Cambridge, UK.

Bruzzone, L., Camps-Valls, G., 2009. Kernel Methods for Remote Sensing Data Analysis. Wiley, Chichester, UK.

Bürling, K., Hunsche, M., Noga, G., 2010. Quantum yield of non-regulated energy dissipation in PSII (Y(NO)) for early detection of leaf rust (*Puccinia triticina*) infection in susceptible and resistant wheat (*Triticum aestivum* L.) cultivars. Precis. Agric. 11 (6), 703–716.

Bürling, K., Hunsche, M., Noga, G., 2011. UV-induced fluorescence spectra and lifetime determination for detection of leaf rust (*Puccinia triticina*) in susceptible and resistant wheat (*Triticum aestivum*) cultivars. Funct. Plant Biol. 38, 337–345.

Buschmann, C., 2007. Variability and application of the chlorophyll fluorescence emission ratio red/far-red of leaves. Photosynth. Res. 92, 261–271.

Buschmann, C., Langsdorf, G., Lichtenthaler, H.K., 2008. Blue, green, red and far-red fluorescence signatures of plant tissues, their multicolor fluorescence imaging and application for agrofood assessment. In: Zude, M. (Ed.), Optical Monitoring of Fresh and Processed Agricultural Crops - Basics and Applications for a Better Understanding of Non-Destructive Sensing. CRC Press, (Francis & Taylor Group), Boca Raton, pp. 272–319.

Camargo, A., Smith, J.S., 2009. Image pattern classification for the identification of disease causing agents in plants. Comput. Electron. Agric. 66, 121–125.

Cerovic, Z.G., Samson, G., Morales, F., Tremblay, N., Moya, I., 1999. Ultraviolet-induced fluorescence for plant monitoring: present state and prospects. Agronomie 19, 543–578.

Chaerle, L., Hagenbeek, D., De Bruyne, E., Valcke, R., Van der Straeten, D., 2004. Thermal and chlorophyll-fluorescence imaging distinguish plant–pathogen interactions at an early stage. Plant Cell Physiol. 45 (7), 887–896.

Chang, C.-C., Lin, C.-J. 2001. A library for support vector machines. Available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

De Boor, C., 1978. A practical Guide to Splines. Springer, New York, USA.

Ferreiro-Arman, M., Da Costa, J., Homayouni, S., Martn-Herrero, J., 2006. Hyperspectral image analysis for precision viticulture. In: Campilho, A., Kamel, M. (Eds.), Image Analysis and Recognition. Springer-Verlag, pp. 730–741.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Boston, USA.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46 (1–3), 389–422.

Huang, Y., Lan, Y., Thomson, S.J., Fang, A., Hoffmann, W.C., Lacey, R.E., 2010. Development of soft computing and application in agricultural and biological engineering. Comput. Electron. Agric. 71, 107–127.

Hughes, G.F., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theory 14, 55–63.

Karimi, Y., Prasher, O.S., Patel, R.M., Kim, M.R.H.S., 2006. Application of support vector machine technology for weed and nitrogen stress detection in corn. Comput. Electron. Agric. 51 (1–2), 99–109.

Koch, K.R., 1999. Parameter Estimation and Hypothesis Testing in Linear Models, seconnd ed. Springer, Verlag, Bonn.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, vol. 2(12), pp. 1137–1143.

Kohavi, R., John, G., 1997. Wrappers for feature subset selection. Artif. Intell. 97 (1–2), 273–324.

Kononenko, I. 1994. Estimating Attributes: Analysis and Extensions of RELIEF. ECML, pp. 171–182.

Kuckenberg, J., Tartachnyk, I., Noga, G., 2009. Temporal and spatial changes of chlorophyll fluorescence as a basis for early and precise detection of leaf rust and powdery mildew infections in wheat leaves. Precis. Agric. 10, 34–44.

Lang, M., Stober, F., Lichtenthaler, H.K., 1991. Fluorescence emission spectra of plant leaves and plant constituents. Radiat. Environ. Biophys. 30, 333–347.

Lilliefors, H., 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. J. Am. Stat. Assoc. 62, 339–402.

Lüdeker, W., Dahn, H.-G., Günther, K.P., 1996. Detection of fungal infection of plants by laser-induced fluorescence: an attempt to use remote sensing. J. Plant Physiol. 148, 579–585.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T., 2006. YALE: rapid prototyping for complex data mining tasks. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 935–940.

Mucherino, A., Papajorgij, P., Paradolos, M.P., 2009. A survey of data mining techniques applied to agriculture. Oper. Res. Int. Journal 9 (2), 121–140.

Oerke, E., Dehne, H., 2004. Safeguarding production-losses in major crops and the role of crop protection. Crop. Prot. 23 (4), 275–285.

German Federal Plant Variety Office [Bundessortenamt], 2008. Beschreibende Sortenliste – Getreide, Mais, Ölfrüchte, Leguminosen und Hackfrüchte außer Kartoffeln. Deutscher Landwirtschaftsverlag GmbH, Hannover. (in German).

Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. Int. J. Remote Sens. 26 (5), 1007–1011.

Panneton, B., Guillaume, S., Roger, J.-M., Samon, G., 2010. Improved discrimination between monocotyledonus and dicotyledonous plants for weed control based on the blue–green region of ultraviolet-induced fluorescence spectra. Appl. Spectrosc. 64 (1), 30–36.

Rockafellar, R.T., 1993. Lagrange multipliers and optimality. SIAM Rev. 35, 183–238.

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Plümer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Comput. Electron. Agric. 74 (1), 91–99.

Schölkopf, B., Smola, A., 2002. Learning with Kernels – Support Vector Machines, Regularization and Beyond. MIT Press, Cambridge.

Vapnik, V., 1998. Statistical Learning Theory. Wiley-Interscience, New York, USA.

Vermerris, W., Nicholson, R., 2006. The role of phenols in plant defense. In: Vermerris, W., Nicholson, R. (Eds.), Phenolic Compound Biochemistry. Springer, Dordrecht, The Netherlands, pp. 222–234.

West, J., Bravo, C., Oberti, R., Lemaire, D., Moshou, D., McCartney, H.A., 2003. The potential of optical canopy measurement for targeted control of field crop diseases. Annu. Rev. Phytopathol. 41, 593–614.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bull. 1, 80–83.

Zhang, L., Dickinson, L., 2001. Fluorescence from rust fungi: a simple and effective method to monitor the dynamics of fungal growth in plants. Physiol. Mol. Plant Pathol. 59 (137), 141.

Zheng, H., Lu, H., Zheng, Y., Lou, H., Chen, C., 2010. Automatic sorting of Chinese jujube (*Zizyphus jujube* Mill. Cv. 'hongxing') using chlorophyll fluorescence and support vector machines. J. Food Eng. 101 (4), 402–408.

A.5 Identification of combined vegetation indices for the early detection of plant diseases

Rumpf, T., Mahlein, A.-K., Dörschlag, D., Plümer, L., 2009b. Identification of combined vegetation indices for the early detection of plant diseases. In: Neale, M. C., Maltese, A. (Eds.), Proceedings of the SPIE Conference on Sensing for Agriculture, Ecosystems and Hydrology. Vol. 7472. Berlin (Germany).

Abstract

The aim of this research is the early detection of plant diseases based on the combination of vegetation indices. We have seen that an individual index such as the most popular one, namely NDVI, does not discriminate adequately between healthy and diseased plants, e.g. *Cercospora beticola*, *Erysiphe betae*, and *Uromyces betae*. However, by combining vegetation indices, which are usually called features in classification, very reliable results can be achieved. We use Support Vector Machines for classification. By this we receive a classification accuracy of almost 95% for *Cercospora beticola* and *Uromyces betae* and still over 92% for *Erysiphe betae*. Depending on the different plant diseases we have found that different vegetation indices are important, too. Consequently, the question how to find the best index for every plant disease and the choice of the best subset arise. Both questions are not the same, because different indices contain similar information which can already be seen from the formula of the calculation of the vegetation index. These dependencies do not have to be linear. In order to identify optimal subsets of features for the different pathogens already at an early stage of infestation, we have found that entropy and mutual information are adequate concepts. Accordingly we use the minimum redundancy - maximum relevance (mRMR) criterion to evaluate the features. We have found that we need different indices and feature subsets of different sizes for different diseases.

Identification of combined vegetation indices for the early detection of plant diseases

T. Rumpf[1], A. Mahlein[2], D. Dörschlag[1] and L. Plümer[1]
[1]Dep. of Geoinformation, Institute of Geodesy and Geoinformation, University of Bonn, Germany
[2]Institute of Crop Science and Resource Conservation, (INRES) – Phytomedicine, University of Bonn, Germany

ABSTRACT

The aim of this research is the early detection of plant diseases based on the combination of vegetation indices. We have seen that an individual index such as the most popular one, namely NDVI, does not discriminate adequately between healthy and diseased plants, e.g. *Cercospora beticola*, *Erysiphe betae*, and *Uromyces betae*. However, by combining vegetation indices, which are usually called features in classification, very reliable results can be achieved. We use Support Vector Machines for classification. By this we receive a classification accuracy of almost 95% for *Cercospora beticola* and *Uromyces betae* and still over 92% for *Erysiphe betae*. Depending on the different plant diseases we have found that different vegetation indices are important, too. Consequently, the question how to find the best index for every plant disease and the choice of the best subset arise. Both questions are not the same, because different indices contain similar information which can already be seen from the formula of the calculation of the vegetation index. These dependencies do not have to be linear. In order to identify optimal subsets of features for the different pathogens already at an early stage of infestation, we have found that entropy and mutual information are adequate concepts. Accordingly we use the minimum redundancy – maximum relevance (mRMR) criterion to evaluate the features. We have found that we need different indices and feature subsets of different sizes for different diseases.

Keywords: plant diseases**,** feature selection, correlation, entropy, mutual information, minimum redundancy – maximum relevance criterion, combination of vegetation indices, classification, Support Vector Machines

1. INTRODUCTION

The automatic classification of plant diseases at an early stage is vital for precision crop protection. Spectral reflectance has proven its suitability as a method to detect changes in plant vitality [9, 22]. Several studies have shown a high potential in discriminating between healthy and stressed plants [3, 5, 13, 23]. However, more research is still required. Especially, more research on the effects of the diseases on spectral characteristics is needed, because little attention has been paid to this area up to now. In addition to the discovery in [14] we use combinations of vegetation indices to discriminate between healthy and diseased plants. In this study the following plant diseases *Cercospora beticola, Erysiphe betae,* and *Uromyces betae* on sugar beet leaves are considered. Based on hyperspectral information nine different vegetation indices such as NDVI, ARI, etc., which refer to different wavelengths of the visible and near infrared spectrum, were calculated. A detailed description of the vegetation indices is given in [14]. Support Vector Machines (SVMs) are used as classification algorithm. SVMs determine an optimal separating hyperplane by quadratic optimisation aiming at maximising the margin between two different classes, therefore leading to suitable generalisation properties. It is not possible to use one vegetation index separately to differentiate adequately between inoculated and

non inoculated plants. For the classification on the basis of the most popular index NDVI of plants that are inoculated with *Uromyces betae* and non inoculated plants, for example, we have only achieved a classification accuracy of less than 71%. We have taken into account inoculated plants from the day after inoculation until first specific symptoms are visible. Thus, we take a new approach which combines multiple vegetation indices. As a result by combining all calculated vegetation indices we have received a classification accuracy of nearly 95%. Despite the good results we have obtained so far we may have used irrelevant features for classification. In order to find an optimal subset of features, their relevance has to be specified. One suitable criterion is based on the correlation with the class label, viz. the intensity of infestation. Then we select the top features. Against this background the difficulty consists in only using correlation to the class label that possible dependencies between the features themselves are not considered. In this study we describe a procedure which rectifies this mistake.

In general, two different feature selection approaches are distinguished, i.e. the filter approach and the wrapper approach. The filter approach assesses the relevance of features and filters out irrelevant ones before an induction occurs. Accordingly, it can be seen as a pre processing step. By contrast the wrapper approach considers the dependence of the selected feature subset on the performance of the induction algorithm. One advantage of the filter approach is its short run time. In order to find the optimal subset several search strategies can be used, which are not discussed in this study.

This paper is organized as follows: In section 2 we introduce the notion of relevance and discuss feature subset selection by using the correlation to the label class only. In section 3 we describe the filter approach for finding optimal subsets and discuss several criterions with different search methods for rectifying the mistake of ignoring the dependencies between the features. The idea of an entropy-based discretization is presented in section 4. The classification algorithm Support Vector Machine, which is subsequently used, is presented in section 5. In section 6 the results are compared by applying the two different procedures for finding the optimal feature subset. We conclude with the discussion of the results in section 7.

2. FEATURE RELEVANCE AND FEATURE RANKING

In supervised learning each instance (sample) is associated with a label class. In our case the whole process from the appearance of the plant diseases after inoculation to a high degree of specific symptoms on the sugar beet leaves is subsumed under inoculated plants. The relevance of different features is often unknown a priori. Thus, in order to get all potential information about the label class Y, a large number of features is introduced. However, some of the features are either partially or completely irrelevant or redundant for the classification. In this context [1, 11, 12] give a definition of relevant features. Generally, the set of all features X can be distinguished into relevant and irrelevant ones. Furthermore, two degrees of relevance are defined, viz. strong relevance and weak relevance. An individual feature X_i is strong relevant if $P(Y|X - X_i) \neq P(Y|X)$. Strong relevance implies that the feature cannot be removed without losing prediction accuracy. Weak relevant features can sometimes contribute to prediction accuracy. This depends on the existence of a subset of features S so that $P(Y|S - X_i) \neq P(Y|S)$. Features are either strongly or weakly relevant, otherwise they are irrelevant.

The necessary probability density function is unknown so that at this point we have to select a suitable criterion to measure the relevance. Typical criteria are, for instance, correlation, consistency or information gain. The most simple approach to determine the relevance of a feature is based on one single feature. Every feature is considered individually with regard to the label class. Afterwards the feature subset is built by selecting the top k of the ranking.

In our study we calculate the correlation coefficient between vegetation index-values and disease severity of *Cercospora beticola, Uromyces betae* and *Erysiphe betae*. The top five vegetation indices with the best correlation to the label class *Erysiphe betae* are SIPI, NDVI, PSSRb, PSSRa and SR which to a large extend are highly correlated themselves (see figures 1).

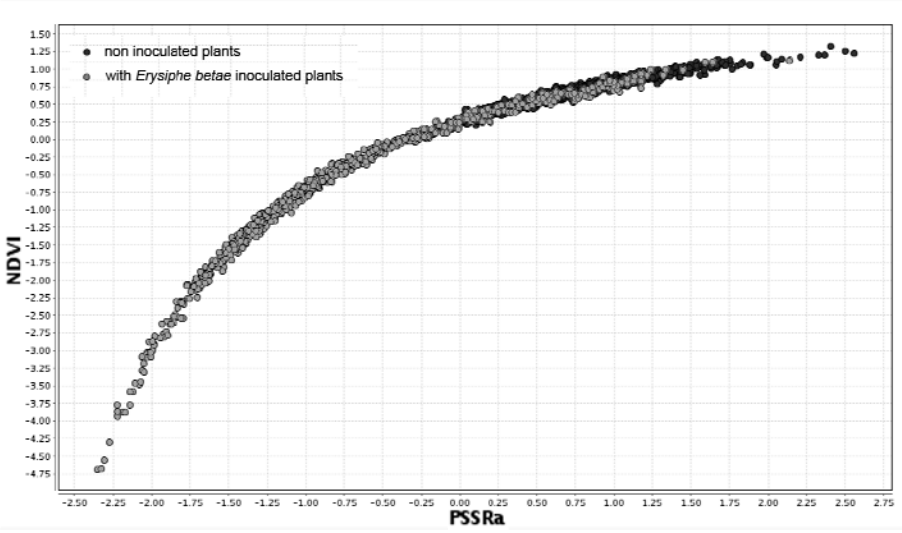


Figure 1: Differentiation of the two classes based on the vegetation indices NDVI and PSSRa: inoculated with *Erysiphe betae* and non inoculated plants. The indices nearly contain the same information because of a function dependency where the two classes overlap each other. As the function is not linear, the dependency is underestimated by the linear correlation. Hence, the additional information of the features is overestimated.

In this case we have shown that an individual measure against the correlation to the label class for every feature is not sufficient, because the dependency with other features is not taken into account. Consequently, weak relevant features remain unconsidered.

3. FEATURE SELECTION

Two different approaches of feature selection are distinguished, i.e. the 'filter approach' and the 'wrapper approach' [12]. The filter approach assesses the relevance of features already before induction occurs and then filters out all irrelevant features. In the wrapper approach, the feature selection algorithm exists as a wrapper around the induction algorithm. It evaluates alternative subsets of features by applying some induction algorithm to the training data and by using the evaluation of the resulting classifier as metric. We focus on the filter approach which can be seen as a preprocessing step with the advantage of a short runtime. With regard to the class labels, which in our case are the inoculated and non inoculated plants, we are interested in finding the optimal feature subset.

In this study we apply a minimum redundancy – maximum relevance (mRMR) criterion for feature selection in order to find the optimal subset [6, 17]. This mRMR criterion is based on mutual information which in turn can be calculated by the difference between entropy and conditional entropy. Entropy [20] quantifies the unpredictability of a feature X and can be formulated as follows

$$H(X) = -\sum_{x \in X} P(x) \log_2 P(x). \tag{3.1}$$

The higher the entropy, the less reliable are our predictions about X . In other words, $H(X)$ can be understood as the amount of uncertainty about X as estimated from its probability distribution.
The conditional entropy is a measure of the remaining uncertainty about X_1 with the knowledge of X_2 . It can be determined by subtracting the entropy of X_2 from the joint entropy of X_1 and X_2 .

$$H(X_1|X_2) = -\sum_{x_1,x_2} P(x_1,x_2)\log_2 P(x_1|x_2) = H(X_1,X_2) - H(X_2) \tag{3.2}$$

Accordingly, the dependence between two features can be quantified by the 'mutual information'

$$I(X_1;X_2) = \sum_{x_1,x_2} P(x_1,x_2)\log_2 \frac{P(x_1,x_2)}{P(x_1)P(x_2)} = H(X_1) + H(X_2) - H(X_1,X_2). \tag{3.3}$$

In essence, the mutual information $I(X_1;X_2)$ can be seen as a measure of dependency between the features. The mutual information is not negative. It is zero if the features are independent, i.e. $P(x_1,x_2) = P(x_1)P(x_2)$. Mutual information between a feature and a label class measures the amount of information provided by feature X_i about label Y . In this case it is often called 'information gain' $I(X_i;Y)$.
Now we are able to describe the mRMR criterion. Let $S \subseteq X$ which consists of m features. The basic idea of the mRMR criterion is to find a discrimination of a subset of features which maximizes the information gain and minimizes the mutual information to each other selected feature. In detail the maximal relevance of a feature is approximated by the mean value of all information gain values between the individual feature X_i and the label class Y .

$$\max \; D = \frac{1}{|S|}\sum_{x_i \in S} I(x_i;y) \tag{3.4}$$

Frequently, these features have a high redundancy so that the loss for discrimination is not significant if one of the features is removed. Thus, the redundancy has to be minimized

$$\min \; R = \frac{1}{|S|^2}\sum_{x_i,x_j \in S} I(x_i,x_j). \tag{3.5}$$

The constraints (3.5) and (3.6) are combined in the mRMR criterion which optimizes D and R simultaneously

$$\max \; \Phi(D,R), \Phi = D - R. \tag{3.6}$$

In practice, incremental search methods are used to find the optimal feature subset by maximizing the defined $\Phi(D,R)$ in (3.6).

4. DICRETIZATION

The original data consists of continuous features, but the criterion in order to find the optimal subset of features only works with discretized data. In general, two different discretization methods are distinguished, namely supervised and unsupervised methods, in analogy to supervised and unsupervised learning methods.

The difference between the two methods is that supervised discretizations take into account class labels whereas unsupervised do not. In our work we apply a recursive entropy minimization algorithm [7, 8], a well established discretization algorithm which uses a very fine partition where the probability density function of the individual features are overlapping. With regard to the development of the plant diseases the overlap between the probability density function is located at the point when inoculated plants do not show visible symptoms or when the specific symptoms rarely appear.

This supervised algorithm selects bin boundaries for discretization based on the class information entropy of candidate partitions which can be formulated as follows:

$$H(X,T;D)=\frac{|D_1|}{D}H(D_1)+\frac{|D_2|}{|D|}H(D_2) \tag{4.1}$$

where D represents a set of instances (samples) and T is the cut value which determines the partition boundaries for the subset of the samples. Because of the binary discretization boundary $D_2 = D - D_1$. For a given feature X the minimal boundary $T_{\min}$ which minimizes the entropy function over all possible partition boundaries is selected. This is applied recursively to both subsets induced by $T_{\min}$ until some stopping criterion is achieved. The entropy-based discretization in [8] uses the 'Minimum Description Length' (MDL) criterion [18] to stop the recursive partitioning if

$$I(X_1,T;D)<\frac{\log_2(N-1)}{N}+\frac{\Delta(X_1,T;D)}{N} \tag{4.2}$$

where N is the number of instances in the set S, $\Delta(X_1,T;S)=\log_2\left(3^k-2\right)-[k\cdot H(S)-k_1\cdot H(S_1)-k_2\cdot H(S_2)]$ and k_i is the number of class labels represented in the set S_i. This procedure is applied to every feature. Since the partitioning is carried out in this way and evaluated independently by using this criterion, the following characteristic arise. The discretization partitions very finely in areas with high entropy whereas other areas with low entropy will be roughly divided.

5. CLASSIFICATION WITH SUPPORT VECTOR MACHINES

Classification algorithms aim at finding regularities in the features $\mathbf{x}$ to predict the class label y. In our case the features are vegetation indices based on hyperspectral information. The training data can be formulated as

$$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\ldots,(\mathbf{x}_n,y_n) \tag{5.1}$$

where the bold $\mathbf{x}$ describes a feature vector of the form $x_1,\ldots,x_m$ and where y represents the label. This way of learning is called supervised learning. Then the classifier algorithm is faced with an unseen object which has to be assigned to one of the classes. This results in finding a function f which fits the training data in the best way possible, i.e.

$$y_i = f(\mathbf{x}_i) \quad \text{for all} \quad i,\ldots,n\,. \tag{5.2}$$

However, it is not sufficient to minimise the training error. The classifier is mainly interested in discriminating unseen data. This is also known as generalisation ability.

Finding a function which fits the training data is an 'ill-posed problem' since there is an infinite number of functions which have this property. It is a basic assumption of machine learning that the class of function to be learned has to be restricted. In this respect the selection of the function class is a compromise between an adequate high model complexity and a prevention of overfitting, both of which are carried out optimally by a Support Vector Machine (SVM) [21].
In the following, we first examine the most simple case of classification where the function class is linear and where it can be described immediately in form of scalar product in the input space. In general, SVM belongs to the kernel based margin classifier that draws an optimal separating hyperplane in the feature space. The hyperplane is constructed by its normal vector $\mathbf{w}$ and the bias b. The hyperplane can be specified as follows

$$\langle \mathbf{w},\mathbf{x} \rangle + b = 0 \quad \text{where} \quad \mathbf{w} \in \mathbb{R}^m, b \in \mathbb{R}. \tag{5.3}$$

where $\langle \mathbf{w}, \mathbf{x} \rangle$ denotes the scalar product and means $x_1 \cdot w_1 + \ldots + x_m \cdot w_m$. This yields the corresponding decision function

$$f(\mathbf{x}) = \operatorname{sgn}(\langle \mathbf{w}, \mathbf{x} \rangle + b) \tag{5.4}$$

where the sign function extracts the sign of a real number to build the two different classes. In order to maximize the margin $\mathbf{w}$ and b have to be minimized. This can be formulated as a quadratic optimisation problem

$$\underset{\mathbf{w}\in\mathbb{D},b\in R}{\text{minimize}} \quad \tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 \tag{5.5}$$

$$\text{subject to } y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 \quad \text{for all} \quad i = 1, \ldots, n. \tag{5.6}$$

Up to now only linearly separable classes were considered, but SVMs are able to separate classes with a non-linear discriminant, too. The basic idea of SVMs is to map the data into a new feature space and then solve the constrained optimisation problem. It seems to be very expensive to compute the mapping into a high-dimensional space. For this reason a kernel function that generalizes the scalar product is introduced to make the computation very simple [2].
This is referred to as the 'kernel trick', which causes an implicit mapping in the feature space without explicitly knowing the mapping function Φ. The most popular kernel function is the radial basis function $k\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle = \exp\left(\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2\right)$. Accordingly, $\Phi(\mathbf{x}_i)$ is substituted for $\mathbf{x}_i$ and the condition for optimal classification is

$$y_i((\mathbf{w} \cdot \Phi(\mathbf{x}_i)) + b \geq 1) \quad \text{for all} \quad i = 1, \ldots, n, \tag{5.7}$$

So far we have made the implicit assumption that the datasets are free of noise and may be classified perfectly. In this case SVM gives a 'hard margin'. In practice this assumption does not hold in most cases. This problem, however, may be handled by 'soft margins' which allow and penalize classification errors. Accordingly, a modification was introduced where slack-variables ξ_i are used to relax the so-called hard-margin constraints (5.6) [4], so that some classification errors which depend on ξ_i are allowed. The 'soft margin' is formulated as

$$y_i((\mathbf{w} \cdot \mathbf{x}_i) + b \geq 1 - \xi_i), \qquad \xi_i \geq 0, \; i = 1, \ldots, n. \tag{5.8}$$

Consequently, the SVM requires the solution of this optimization problem

$$\underset{\mathbf{w}\in\mathbb{D},b\in R}{\text{minimize}} \quad \tau(\mathbf{w},\xi_i)=\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{n}\xi_i \ . \tag{5.9}$$

The influence of the classification errors are parameterised with the parameter C . A larger C penalizes a wrong classification more strongly.

6. CLASSIFICATION RESULTS

We need healthy and diseased plants for the training of the classification algorithm [14]. The plants are differentiated into inoculated plants from first day after inoculation (dai) where specific symptoms corresponding to the plant disease are visible and non inoculated plants. The development of diseases and the symptoms caused by *Erysiphe betae* are already visible five days after the inoculation as fluffy white mycelia covering the leaf surface, whereas *Cercospora beticola* causes its typical necrotic leaf spots after six days. At last, eight days after inoculation sugar beet rust pustules caused by *Uromyces betae* are visible. The results by using all vegetation indices as features for classification are displayed in table 1 varied for the three pathogens.

	classification accuracy from non inoculated sugar beet plants and an individual plant diseases
Cercospora beticola: dai ≥ 6	(true non inoculated plants, true *C. beticola*) **94.90%** (97.19%, 92.60%)
Uromyces betae: dai ≥ 8	(true non inoculated plants, true *U. betae*) **94.64%** (95.60%, 93.69%)
Erysiphe betae: dai ≥ 5	(true non inoculated plants, true *E. betae*) **92.37%** (93.33%, 91.35%)

Table 1: Classification results based on all vegetation indices to separate between individual plant diseases and non inoculated sugar beet plants, where the first value presents the accuracy (specificity, sensitivity).

We used the software Rapid Miner [16] in order to define a process chain for an automatic classification which is evaluated by a cross validation. The average classification accuracy of *Erysiphe betae* and non inoculated plants was over 92% whereas otherwise it was even almost 95%. In all cases the classification accuracy of the non inoculated plants (sensitivity) was better than the one of the plant disease. A detailed consideration of the classification results based on the combination of all vegetation indices is given in [19].

In order to determine a relevant feature subset we define a threshold that does not decrease the classification accuracy by more than 2.5%. Accordingly, as mentioned in section 4, we added the entropy-based discretization in the preprocessing step of the process chain. Afterwards, we used a Matlab program for calculating the mRMR criterion to find an optimal feature subset [Peng H. 2003]. In order to classify the different plant diseases and non inoculated plants we used Support Vector Machines. The classification was based on feature selection with the mRMR criterion as well as the selection after the correlation to the label class. Both, the features combination and the number of features depend on the plant diseases (see table 2). With regard to the number of features, we found that *Cercospora beticola* was already classified with two features whereas *Uromyces betae* needed three features and *Erysiphe betae* even five features.

The results of the two different feature selection methods were obviously different. There were different results for the classification accuracy as well as the feature combination. However, the differences were nearly between two and ten percent. One of the large distinctions between the classification results appeared by differentiating between non inoculated plants and plants that were inoculated with *Erysiphe betae*. Beside the clear difference in classification accuracy the feature combination was nearly complementary. The only common vegetation index was the NDVI. The feature combination by discriminating between plants which were inoculated with *Cercospora beticola* and non inoculated leaves was also very different, but here the classification accuracy only differed about two percent.

	classification accuracy from non inoculated plants and an individual plant diseases	**selected vegetation indices**
Cercospora beticola:	(true non inoculated plants, true *C. beticola*)	
correlation with label class:	**91.67%** (98.12%, 85.21%)	NDVI, mCAI
mRMR criterion:	**93.39%** (98.75%, 88.02%)	ARI, SPAD
Uromyces betae:	(true non inoculated plants, true *U. betae*)	
correlation with label class:	**83.69%** (95.24%, 72.14%)	mCAI, NDVI, ARI
mRMR criterion:	**93.93%** (96.07%, 91.79%)	SPAD, REP, ARI
Erysiphe betae:	(true non inoculated plants, true *E. betae*)	
correlation with label class:	**84.65%** (97.75%, 70.73%)	SIPI, NDVI, PSSRb, PSSRa, SR
mRMR criterion:	**90.25%** (93.33%, 86.98%)	REP, SPAD, ARI, mCAI, NDVI

Table 2: Classification results based on the feature subset determined by feature selection to differentiate between different plant diseases and non inoculated sugar beet leaves, where the first value presents the accuracy (specificity, sensitivity).

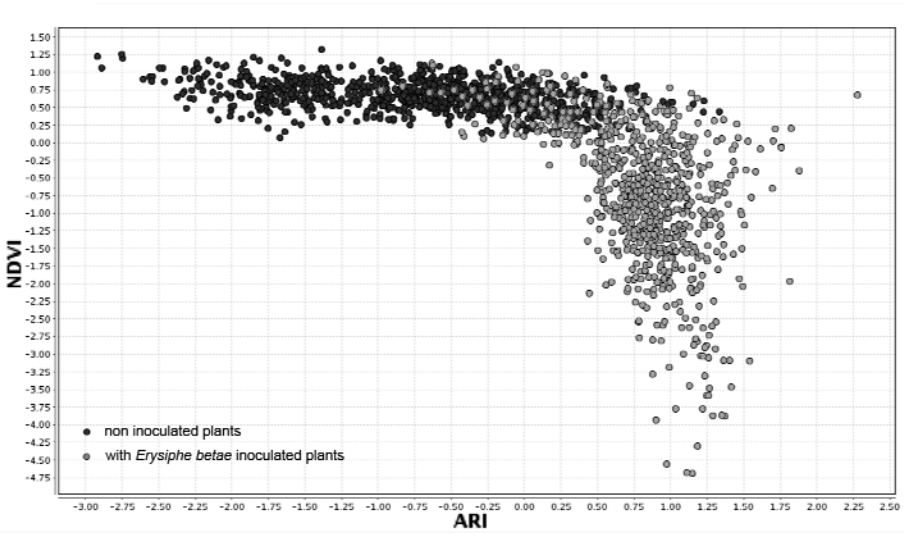


Figure 2: Differentiation of the two classes, viz. inoculated with *Erysiphe betae* and non inoculated plants, based on the vegetation indices NDVI and ARI. The indices are almost independent of each other, with a small overlapping of the two classes.

7. DICUSSION AND CONCLUSION

Combined vegetation indices are able to discriminate adequately between one of afore mentioned plant diseases and non inoculated plants.

We achieved high classification accuracies from almost 95% for *Cercospora beticola* and *Uromyces betae* and over 92% for *Erysiphe betae* already at an early stage of infestation and even before specific symptoms were visible. Regarding the question whether a subset of features can be used for classification the answer is affirmative. Only one feature subset was required, though different with the different plant diseases. In this context we have found out that by using the optimal subset of features the classification accuracy for *Uromyces betae* was even better than using all features. Here some more investigations have to be done. However, the results of the two different feature selection methods vary substantially. The classification with features, which were selected on the basis of correlation to the label class, was always inadequate in comparison to the usage of the mRMR criterion. Correspondingly the dependencies to the other features had to be taken into account during finding an optimal subset for classification. In this context a suitable discretization of the continuous features is very important. If we use an equal width binning discretization, this will lead to a comparatively bad classification which we did not discuss in this paper. One disadvantage of the mRMR criterion is that the number of features that forms the subset had to be determined by defining a loss threshold of classification accuracy, whereby the optimal model complexity is established implicitly. This can easily be rectified by resolving this mRMR optimization problem through using a more sophisticated search method.

ACKNOWLEDGEMENTS

This study has been conducted within the Research Training Group 722 ‘Information Techniques for Precision Crop Protection’, funded by the German Research Foundation (DFG).

REFERENCES

[1] Blum, Avrim L.; Langley, Pat: “Selection of relevant features and examples in machine learning”, Artificial Intelligence: Vol. 97(1-2), 245–271 (1997).

[2] Boser, Bernhard E.; Guyon, Isabelle M.; Vapnik, Vladimir N.: “A training algorithm for optimal margin classifiers”, Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, ACM Press, 144–152 (1992).

[3] Bravo, C., Moshou, D., Oberti, R., West, J., McCartney, A., Bodria, L. and Ramon, H.: “Foliar disease detection in the field using optical sensor fusion” Agricultural Engineering International:Vol.6, 1-14 (2004).

[4] Cortes, Corinna; Vapnik, Vladimir N.: “Support-Vector Networks”, Machine Learning: Vol. 20(3), 273–297 (1995).

[5] Delalieux, S., van Aardt, J., Kelemans, W. and Coppin, P.: “Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral analysis”, Proceedings of the 4th EARSLeL Workshop on Imaging Spectroscopy, Warsaw 2005, 677-689 (2005).

[6] Ding, Chris; Peng, Hanchuan: “Minimum Redundancy Feature Selection from Microarray Gene Expression Data”, Journal of Bioinformatics and Computation Biology: Vol.3(2), 185–205 (2005).

[7] Dougherty, James; Kohavi, Ron; Sahami, Mehran: “Supervised and Unsupervised Discretization of Continuous Features”, Proceedings of the 12th International Conference on Machine Learning, San Francisco, CA., Morgan Kaufmann Publishers, 194–202 (1995).

[8] Fayyad, Usama M.; Irani, Keki B: “Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning”, Proceedings of the 13th International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers, 1022–1029 (1993).

[9] Hatfield, J.L., Gitelson, A.A., Schepers, J.S. and Walthall, C.L.: “Application of spectral remote sensing for agronomic decisions”, Agronomy Journal: Vol. 100, 117-131 (2008).

[10} Jakulin, Aleks; Bratko, Ivan: “Analyzing attribute dependencies”, Knowledge Discovery in Database: PKDD 2003, Berlin Heidelberg, Springer [Lecture Notes in Computer Science, 2003], 229–240 (2003).

[11] John, George H.; Kohavi, Ron; Pfleger, Karl: “Irrelevant Features and the Subset Selection Problem”, Proceedings 11th International Conference on Machine Learning, San Mateo, CA: Morgan Kaufmann Publishers, 121–129 (1994).

[12] Kohavi, Ron; John, George H.: “Wrappers for feature subset selection”, Artificial Intelligence: Vol. 97(1-2), 273–324 (1997).

[13] Laudien, R., Bareth, G. and Doluschitz, R.: “Analysis of hyperspectral field data for detection of sugar beet diseases”, Proceedings of the EFITA Conference, Debrecen 2003, 375-381 (2003).

[14] Mahlein, A-K., Steiner, U., Dehne, H-W., Oerke, E-C.: “Spectral signatures of diseased sugar beet leaves”, Proceeding of the 7th JIAC 2009, Wageningen, Netherlands, 239-246 (2009).

[15] McGill, William J.: “Multivariate information transmission”, Psychometrika: Vol. 19(2), 97–116 (1954).

[16] Mierswa, Ingo; Wurst, Michael; Klinkenberg, Ralf; Scholz, Martin; Euler, Timm: “YALE: Rapid Prototyping for Complex Data Mining Tasks”, Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2006).

[17] Peng, Hanchuan; Long, Fuhui; Ding, Chris: “Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy”, IEEE Transactions on Pattern Analysis and Machine Intelligence: Vol. 27(8), 1226–1238 (2005).

[18] Rissanen, Jorma: “Stochastic Complexity and Modeling”, The Annals of Statistics: Vol. 14(3), 1080–1100 (1986).

[19] Rumpf, T.,Mahlein, A-K. “Support Vector Machines for an early detection and classification of plant diseases based on combined vegetation indices”, (in preparation).

[20] Shannon, C.E.: “A mathematical theory of communication”, The Bell System Technical Journal, Vol. 27, 379-423, 623-656, (1948).

[21] Vapnik, Vladimir N.: “Estimation of dependences based on empirical data”, New York, NY, Springer (1982).

[22] West, J., Bravo, C., Oberti, R., Lemaire, D., Moshou, D., Mc Cartney, H.A., “The potential of optical canopy measurement for targeted control of field crop diseases”, Annual Review of Phytopathology: Vol. 41, 593- 614 (2003).

[23] Zhang, M., Qin, Z. and Liu, X.: “Remote sensed spectral imagery to detect late blight in field tomatoes”, Precision Agriculture: Vol. 6, 489-508 (2005).

A.6 Optimal wavelengths for an early identification of *Cercospora beticola* with support vector machines based on hyperspectral reflection data

Rumpf, T., Mahlein, A.-K., Römer, C., Plümer, L., 2009a. Optimal wavelengths for an early identification of cercospora beticola with support vector machines based on hyperspectral reflection data. In: Institute of Electrical and Electronics Engineers (IEEE) (Ed.), 2010 IEEE International Geoscience and Remote Sensing Symposium. Hononulu (Hawai).

Abstract

Automatic classification of plant diseases at an early stage is vital for precision crop protection. Our aim was to identify sugar beet leaves inoculated with *Cercospora beticola* before symptoms are visible. Therefore hyperspectral reflection between 400 and 1050 nm was observed. Relevant wavelengths have to be found in order to implement practical sensor systems with reduced development costs. The main contribution of this study is to identify a minimal subset which is sufficient for separating healthy and inoculated leaves. The heuristic of Hall which analyses the relevance of a feature subset considering the intercorrelation among the features was applied. In order to select a good subset in a reasonable amount of time a genetic algorithm was used. This way enabled a subset of only seven out of 462 wavelengths, which nevertheless enabled us to identify low disease severity $\leq 5\%$ with a classification accuracy of 84.3%. Disease severity above 5% was classified with 99.8%.

OPTIMAL WAVELENGTHS FOR AN EARLY IDENTIFICATION OF *CERCOSPORA BETICOLA* WITH SUPPORT VECTOR MACHINES BASED ON HYPERSPECTRAL REFLECTION DATA

Till Rumpf *, *Christoph Römer, Lutz Plümer*

University of Bonn
Institute of Geodesy and Geoinformation
Department of Geoinformation

Anne-Katrin Mahlein

University of Bonn
Institute of Crop Science and
Resource Conservation (INRES)
Department of Phytomedicine

ABSTRACT

Automatic classification of plant diseases at an early stage is vital for precision crop protection. Our aim was to identify sugar beet leaves inoculated with *Cercospora beticola* before symptoms are visible. Therefore hyperspectral reflection between 400 and 1050 nm was observed. Relevant wavelengths have to be found in order to implement practical sensor systems with reduced development costs. The main contribution of this study is to identify a minimal subset which is sufficient for separating healthy and inoculated leaves. The heuristic of Hall which analyses the relevance of a feature subset considering the intercorrelation among the features was applied. In order to select a good subset in a reasonable amount of time a genetic algorithm was used. This way enabled a subset of only seven out of 462 wavelengths, which nevertheless enabled us to identify low disease severity $\leq 5\%$ with a classification accuracy of 84.3%. Disease severity above 5% was classified with 99.8%.

Index Terms— hyperspectral data, *Cercospora beticola*, Support Vector Machines, feature selection, genetic algorithm

1. INTRODUCTION

Cercospora beticola, the agent of Cercospora leaf spot, is the major fungal leaf pathogen in sugar beet production. *Cercospora beticola* causes a reduction of yield quantity and quality, with economical losses up to US$ 1500 ha^{-1} [1]. Characteristic symptoms are grey-brown necrotic leaf spots, with a reddish brown boarder (Fig. 1).

At increasing infestation whole leaves are necrotic. The application of fungicides is still one fundamental tool in managing fungal leaf pathogens in sugar beet. To optimize the efficiency of the fungicide application, the detection and differentiation of foliar diseases at early stages of the disease epidemic is essential. As visual monitoring of diseases in the field is time-consuming and expensive, a sensor based alternative is required. Foliar sugar beet diseases influence the characteristic spectral signature of sugar beet leaves in different ways. Mahlein et al. [2] pointed out that sugar beet leaves, diseased with different leave diseases, show disease specific spectral signatures. Each disease had a characteristic impact on the spectral behavior of sugar beet. This is due to physiological and physiochemical changes caused by the pathogen, and subsequently to the characteristic symptoms. Cercospora leaf spot induce changes in the visible spectrum from 400 to 700 nm which is mainly influenced by the content of pigments, and also in the near infrared range from 700 to 1050 nm as an effect of the cellular structure and the water content.

Fig. 1. Typical mature symptoms of Cercospora leaf spot.

Separation, however, is difficult. If we focus on the disease at an early stage, the difference between the medians of the two classes (healthy leaves and inoculated leaves) is smaller than their respective standard deviations (Fig. 2). Hence, it is unfeasible to achieve a high accuracy using simple discriminants or ratios. Therefore, the approach of this study was, whether more advanced data analysis methods will succeed with a more accurate differentiation between healthy and inoculated leaves at an early stage.

*corresponding author, e-mail: rumpf@igg.uni-bonn.de

For reliable classification it is important that the classifier finds hidden patterns discriminating both classes instead of overfitting the training data and relying on single training samples which discriminate randomly. This ability is called generalization.
As Vapnik [3] proved Support Vector Machines (a supervised classification method) have statistical properties which lead to a superior generalization ability. The potential of Support Vector Machines (SVMs) in combination with hyperspectral sensors for precise, reproducible and non destructive disease detection was shown by Rumpf et al. [4].
However, the approach of Rumpf et al. [4] was based on vegetation indices. In order to get more specific features separating healthy and inoculated leaves the hyperspectral signature is used as features. This study aims at finding an optimal subset of only a few wavelengths in order to avoid overfitting and to reduce sensor costs.

2. MATERIAL AND METHODS

2.1. Experiment

Sugar beet plants were cultivated in a greenhouse environment under controlled temperature and light conditions. The experimental plants were watered as necessary and fertilized weekly. The inoculation with the *Cercospora beticola* spore suspension was performed at growth stage 14, when four leaves where fully developed. For inoculation the plants were transferred in a humidity chamber for 48 hours at 25 °C. Spectral data were collected daily till 21 days after inoculation (dai).
Reflectance spectra were obtained using an ASD FieldSpec Pro FR spectroradiometer (Analytic Spectral Devices, Boulder, USA) with a contact probe foreoptic and a leafclip spectra. The contact probe foreoptic has a 10 mm spot size and an integrated halogen light source. The wavelength range of the ASD FieldSpec is from 450 nm to 1050 nm with a spectral resolution of 1.4 nm. Sample reflectance was obtained by inserting a leaf into the leaf clip and comparing leaf reflectance to the reflectance of a dark current as a minimum value and to the reflectance of the white standard (Spectralon Reflectance Standards, Labsphere, North Sutton, USA). Each spectral measurement was an average of 10 scans, in each treatment 15 plants and 2 leaves per plant were sampled.

2.2. Characteristics of the dataset

This resulted in a dataset with 630 samples of both classes (healthy leaves and leaves inoculated with *Cercospora beticola*). Particularly in the early stage, the difference between the medians of the two classes is smaller than their respective standard deviations (2). Thus, the high variance in the samples makes a classification difficult. Only the wavelengths between 570-640 nm and in the rage of 700 nm seem to be suitable for differentiation, but the variance between individual plants was still high. Caused of the interpolation every 1.4 nm adjacent wavelengths are highly correlated. As correlation decreases with increasing distance between two wavelengths, we have to take the relations between different wavelengths into regard as additional information.

2.3. Feature Selection

Using only a subset of wavelengths is important in two different respects. First, in order to implement practical sensor systems with reduced development cost, and second, to reduce the feature space dimension of classification. In particular, too many bad features could worsen the classification result. Thus, a feature selection method is used to identify relevant wavelengths. In general the filter and wrapper approach for feature selection is distinguished [5, 6]. This study focuses on the filter approach, a pre-processing step independent of a specific classifier. The relevance of features can be determined by several information criteria.
First, one possible form of feature selection is feature ranking. Accordingly, one information criterion, like correlation, is applied for feature weighting and afterwards the top k features are selected.
Second, more sophisticated approaches aim to find an optimal subset of features. In our study the correlation-based filter algorithm of Hall [7] turned out to be reliable. This heuristic is based on the hypothesis that good feature subsets contain features highly correlated with the class, yet uncorrelated with each other. The following equation formalizes the heuristic:

$$\text{Merit}_S = \frac{k\overline{r_{cf}}}{\sqrt{k + k\,(k-1)\,\overline{r_{ff}}}} \qquad (1)$$

where Merit_S is the heuristic 'merit' of a feature subset S containing k features, $\overline{r_{cf}}$ the average feature-class correlation, and $\overline{r_{ff}}$ the average feature-feature intercorrelation. However, an exhaustive search of all possible subset combinations of 462 wavelengths is not feasible. In order to find a good subset in reasonable amount of time the use of different search strategies including forward, backward and randomised [5] are necessary. In our study the optimal subset of features is determined by a genetic algorithm. Subsequently the evaluation is concluded by cross-validation.

2.4. Support Vector Machines

As a supervised learning algorithm, Support Vector Machines derive a decision function from given training data. Based on this decision function the class label of a new, unseen data point is predicted. In a binary classification task the training data is given with an M-dimensional feature vector $\vec{x}_i, i = 1..n$ and a corresponding class label $y_i = \pm 1$. As the training data is only a randomly drawn small subset of all possible data points, it is important that the classifier finds

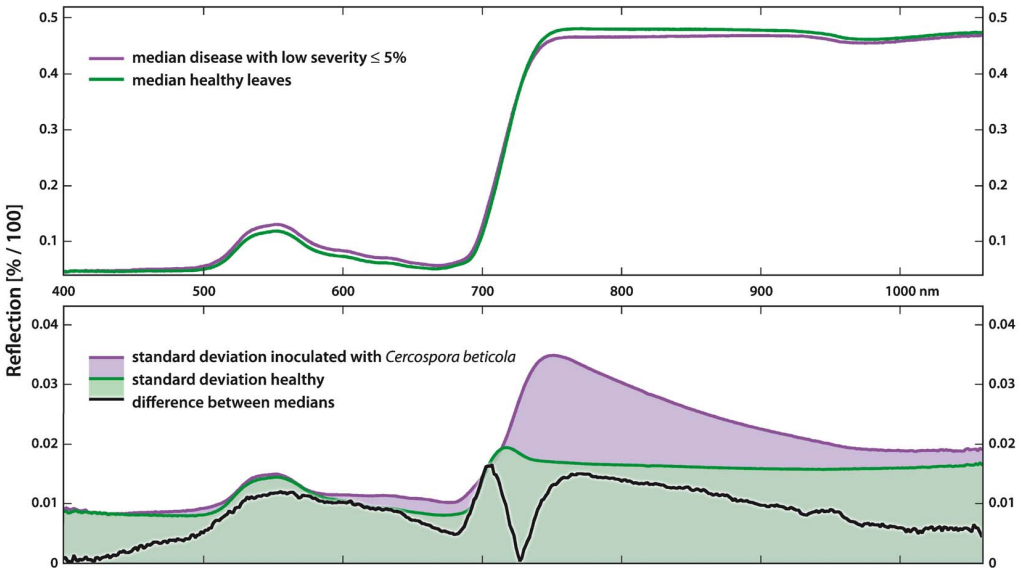


Fig. 2. Reflection spectra healthy versus inoculated leaves: medians and intra-class deviations.

general patterns which distinguish between both classes instead of discriminants which only work for the given samples. Vapnik [3] proved that the generalization ability increases if the complexity of the decision function, measured with the VC-Dimension, decreases. Therefore, assuming both classes are linearly separable, Support Vector Machines restrict their set of possible decision functions to separating hyperplanes of the form: $f(\vec{x}) =< \vec{\omega}, \vec{x} > +b$, where $\vec{\omega}$ are the coefficients of the hyperplane and b describes the offset from the origin.

The optimal hyperplane is the one with maximal margin, i.e. where the distance from the hyperplane to the nearest points of both classes becomes maximal. This, in return, means that we only need those few points on the convex hull of both classes which are nearest to the other class in order to derive the optimal decision function. Those points are called support vectors.

Maximizing the margin is equivalent to minimizing

$$\frac{1}{2}|\vec{\omega}|^2$$

under the constraint that all training data points are correctly classified:

$$y_i\,(< \vec{\omega}, \vec{x} > +b) \geq 1 \quad \forall i.$$

This is a convex optimization problem with inequality constraints. For computational reasons and mathematical advantages those are solved using the dual Lagrangian optimization problem and the KKT conditions [8].

If the classes overlap, it is not possible to meet the constraint that all training data points will be correctly classified without sacrificing the high generalization ability. Therefore, the classifier is allowed to make mistakes during the training step. To allow this, the optimization problem is modified to:

$$\text{minimize } \frac{1}{2}|\vec{\omega}|^2 + C\sum \xi_i,$$

under the constraint:

$$y_i\,(< \vec{\omega}, \vec{x} > +b) \geq 1 - \xi_i \quad \forall i,$$

where the slack variable ξ_i penalizes wrongly classified training samples and C weights the influence of the penalty term on the optimization problem.

Hence, the only parameter needed for the calculation of Support Vector Machines is the parameter C. With the use of kernels [9] nonlinear problems can be solved, too.

3. RESULTS AND DISCUSSION

In this article we showed that for the identification of *Cercospora beticola* the hyperspectral signature can be reduced to seven wavelengths, viz. 611 nm, 651 nm, 660 nm, 795 nm, 858 nm 941 nm, and 948 nm. We applied the correlation-based filter algorithm of Hall [7]. The selected wavelengths are concentrated in three small ranges of the hyperspectral signature. These ranges are important to implement practical sensor systems, because the individual wavelengths may slightly differ when other datasets are used, but the ranges are still the same. Based on the seven selected wavelengths SVMs with linear kernel classify healthy and inoculated leaves at an early stage (from the fourth day after inoculation until an disease severity of 5%) already with an accuracy of

Prediction	Ground truth		*Class precision*
	Healthy	*Cercospora beticola*	
Vegetation indices			
Healthy	204	56	78.46%
Cercospora beticola	36	170	82.52%
Class recall	85.00%	75.22%	**80.25%**
Correlation criterion			
Healthy	216	49	81.51%
Cercospora beticola	24	177	88.06%
Class recall	90.00%	78.32%	**84.31%**

Table 1. Comparison of the classification results between healthy leaves and leaves inoculated with *Cercospora beticola* (disease severity $\leq 5\%$) based on nine vegetation indices and the seven relevant hyperspectral wavelengths identified by the correlation criterion of Hall [7].

84.3% (Table 1). The separation from an disease severity above 5% succeeds nearly without error with an accuracy of 99.8%. In fact the seven selected wavelengths resulted in better classification performance (Table 1) than a comparable approach based on vegetation indices as features on the same dataset [4].

Regarding different levels of disease severity the selected wavelengths by the criterion of Hall always were the same. That indicates that the main changes caused by *Cercospora beticola* influence the same ranges of the hyperspectral signature. With increasing disease severity the classification accuracy rise (Table 2).

Disease severity	Classification accuracy		
	Accuracy	Specificity (healthy leaves)	Sensitivity (diseased leaves)
without symptoms	**68.09%**	76.67%	57.89%
1 - 5%	**93.16%**	98.00%	86.73%
$> 5\%$	**99.81%**	100%	99.63%

Table 2. Classification results between healthy leaves and leaves inoculated with *Cercospora beticola* by different levels of disease severity. Without symptoms means that the inoculated leaf shows no visible symptoms.

Future work to further improve the performance could be a considerably increase of the subset size. With regard to the amount of available samples, however, an oversized subset would possibly result in overfitting. Thus, classification based on selection and combination of individual bands is confined. The next step could be to use aggregated hyperspectral features, as Römer et al. [10] showed. Nevertheless, the necessity of feature selection is not only valid for individual wavelengths, but also by using aggregation.

4. ACKNOWLEDGEMENTS

This study has been conducted within the Research Training Group 722 'Information Techniques for Precision Crop Protection', funded by the German Research Foundation (DFG).

5. REFERENCES

[1] F. J. P. Wolf and A. J. Verreet, "An integrated pest management system in germany for the control of fungal leaf diseases in sugar beet: The ipm sugar beet model," *Plant Disease*, vol. 86, no. 4, pp. 336–344, 2002.

[2] A-K. Mahlein, U. Steiner, H-W. Dehne, and E-C. Oerke, "Spectral signatures of sugar beet leaves for the detection and differentiation of diseases," *Precision Agriculture*, vol. 11, no. 4, pp. 413–431, 2010.

[3] N. Vladimir Vapnik, *The nature of statistical learning theory*, Statistics for engineering and information science. Springer-Verlag, New York, 2nd ed. edition, 2000.

[4] T. Rumpf, A-K. Mahlein, U. Steiner, E-C. Oerke, H-W. Dehne, and L. Plümer, "Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance," *Computers and Electronics in Agriculture*, (accepted).

[5] Ron Kohavi and H. George John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.

[6] L. Avrim Blum and Pat Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 245–271, 1997.

[7] M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," 2000, pp. 359–366, Morgan Kaufmann.

[8] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[9] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines. And Other Kernel-based Learning Methods*, Cambridge University Press, 2000.

[10] C. Römer, T. Rumpf, K. Bürling, M. Hunsche, G. Noga, and L. Plümer, "Early identification of leaf rust on wheat leaves with robust fitting of hyperspectral signatures," in *Proc. of 10th International Conference Precision Agriculture, Denver, USA*, 2010.

Danksagung

"'*Leider läßt sich eine wahrhafte Dankbarkeit mit Worten nicht ausdrücken.*"', Johann Wolfgang von Goethe

Ich möchte es an dieser Stelle aber trotzdem versuchen ...!
Besonderer Dank gilt:

Der Deutschen Forschungsgemeinschaft für die Förderung des Graduiertenkollegs 722 "'Einsatz von Informationstechniken zur Präzisierung des Pflanzenschutzes"', in dessen Rahmen diese Arbeit entstanden ist.

Michael Kneuper und Stefan Teutsch für ihre Hilfe bei der Erstellung von Grafiken.

Meinen ehemaligen Kolleginnen und Kollegen der Professur für Geoinformation für die vielen fachlichen Diskussionen.
Insbesondere danke ich Jan Behmann, Dirk Dörschlag, Jun.-Prof. Dr. Kristian Kersting, Christoph Römer und Dr. Jörg Schmittwilken für die Gespräche, die mich fachlich in meiner Arbeit bestärkt haben.

Dr. Jörg Steinrücken für den letzten Feinschliff.

Sabine Mathey für das Korrekturlesen und die zusätzliche Motivation die Arbeit zu komplettieren.

Pascal Welke für die Umsetzung von Programmideen.

Den Mitautoren meiner Veröffentlichungen für die produktive Zusammenarbeit. Gerade von der intensive Zusammenarbeit mit Dr. Anne-Katrin Mahlein und Dr. Martin Weis hat meine Arbeit besonders profitiert.

PD Erich-Christian Oerke und Prof. Dr. Björn Waske für die Übernahme des Korreferats.

Prof. Dr. Lutz Plümer für meine Förderung, die Betreuung der Arbeit und seine unzähligen Ideen und Anregungen. Darüberhinaus für seinen persönlichen Beistand in einer schwierigen Zeit.

Gaby Mermargen und Herbert Meyer für Kost und Logie, sowie Ihre ermunternden Worte.

Meiner Familie und Freunden für ihre ermutigenden Worte.

Meinen Eltern, Schwiegereltern und besonders meiner Schwester Silke, die sich ganz liebevoll um meine Kinder Timo, Lars und Thorben gekümmert haben während ich an meiner Dissertation gearbeitet habe.

... und Nadine, ohne die die Arbeit gar nicht möglich gewesen wäre.

Acknowledgment

"'*Unfortunately, we can not express a genuine gratitude in words.*"', Johann Wolfgang von Goethe

Anyway, at this point I would try it ...!
Special thanks to:

The Research Training Group 722 'Information Techniques for Precision Crop Protection', funded by the German Research Foundation (DFG) in which context this thesis was developed.

Michael Kneuper and Stefan Teutsch for their support creating figures.

My former colleagues of the department of Geoinformation for many discussions.
Particular thanks to Jan Behmann, Dirk Dörschlag, Jun.-Prof. Dr. Kristian Kersting, Christoph Römer and Dr. Jörg Schmittwilken for the discussions, which have given me solid support and encouragement in my work.

Dr. Jörg Steinrücken for the final touch of details.

Sabine Mathey for proof-reading and the additional motivation for completing my thesis.

Pascal Welke for realising programme ideas.

The co-authors for their very cooperative and productive work in preparing the publications. My thesis especially benefitted from the intensive collaboration with Dr. Anne-Katrin Mahlein and Dr. Martin Weis.

PD Erich-Christian Oerke and Prof. Dr. Björn Waske for agreeing to act as second examiner.

Prof. Dr. Lutz Plümer for my funding, the supervision of the thesis and his uncountable ideas and suggestions. Furthermore, for his personal support during a difficult period.

Gaby Mermargen and Herbert Meyer for housing and food as well as their animating words.

My family and friends for their encouragement.

My parents, parents-in-law and particularly my sister which have lovingly cared for my children Timo, Lars and Thorben at the time when I was working on my thesis.

... and Nadine, without her my thesis would not have been possible.