Research Paper

In-field proximal sensing of septoria tritici blotch, stripe rust and brown rust in winter wheat by means of reflectance and textural features from multispectral imagery

Romain Bebronne [a,1], Alexis Carlier [a,1], Rémi Meurs [b], Vincent Leemans [a], Philippe Vermeulen [c], Benjamin Dumont [b], Benoît Mercatoris [a,*]

[a] Biosystems Dynamics and Exchanges, TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés, 2, 5030, Gembloux, Belgium
[b] Plant Sciences, TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés, 2, 5030, Gembloux, Belgium
[c] Walloon Agricultural Research Centre, Knowledge and Valorization of Agricultural Products Department, Quality and Authentication of Agricultural Products, Chaussée de Namur, 24, 5030, Gembloux, Belgium

ARTICLE INFO

During its growth, winter wheat (*Triticum aestivum* L.) can be impacted by multiple stresses involving fungal diseases that are responsible for high yield losses. Enhancing the breeding and the identification of resistant cultivars could be achieved by collecting automated and reliable information at the plant level. This study aims to estimate the severity of stripe rust (SR), brown rust (BR) and septoria tritici blotch (STB) in natural conditions and to highlight wavebands of interest, based on images acquired through a multispectral camera embedded on a ground-based platform. The severity of the three diseases has been assessed visually in an agronomic trial involving five wheat cultivars with or without fungicide treatment. An acquisition system using multispectral imagery covering the visible and near-infrared range has been set up at the canopy level. Based on spectral and textural features, estimations of area under disease progress curve (AUDPC) were performed by means of artificial neural networks (ANN) and partial least squares regression (PLSR). Supervised classification was also implemented by means of ANN. The ANN performed better at estimating disease severity with R^2 of 0.72, 0.57 and 0.65 for STB, SR and BR respectively. Discrimination in two classes below or above 100 AUDPC reached an accuracy of 81% ($\kappa = 0.60$) for STB. This study, which combined the effect of date, cultivar and multiple disease infections, managed to highlight a few wavebands for each disease and took a step further in the development of a machine vision-based approach for the characterisation of fungal diseases in natural conditions.

© 2020 Published by Elsevier Ltd on behalf of IAgrE.

---

1. Introduction

While winter wheat is the most grown cereal, its yield remains deeply impacted by fungal diseases, such as stripe rust (SR) caused by *Puccinia striiformis*, brown rust (BR) caused by *Puccinia recondita* and septoria tritici blotch (STB) caused by *Zymoseptoria tritici*. These diseases can induce yield reductions of up to 60% on sensitive and untreated wheat cultivars in Belgium (Bodson, De Proft, & Watillon, 2017). In order to ensure high yield, the current agricultural practice involve the selection of resistant cultivars and the uniform application of preventive pesticides. However, breeding resistant cultivars requires plant phenotyping which is conventionally assessed visually by experts. For instance, classical disease notation is based on specific disease symptoms and visual estimations of infected surfaces. Such a method remains time-consuming and induces high variability between observations which highlights the need for automatic and robust measurement tools (Naik et al., 2017).

In the context of precision agriculture, numerical technologies have been used to account for the intra-plot variability of crop parameters (Lindblom, Lindblom, Lundström, Ljung, & Jonsson, 2017) such as leaf area index (Haboudane, Miller, Pattey, Zarco-Tejada, & Stratchan, 2004), nitrogen content (Filella, Serrano, Serra, & Peñuelas, 1995) and disease infection (Mahlein, 2016). These current innovations are considered as promising to monitor disease state in crops such as winter wheat and then adapt the amount of pesticides sprayed to be more cost-effective and with limited impact on the environment (Van Der Werf, 1996). In agronomic trials, the numerical advances would play a significant role in characterising plants resistant to diseases (Kuska et al., 2015). These technologies are however not fully developed yet and require accurate phenotyping tools, especially in natural conditions.

Among the available technologies, imaging spectroscopy has been identified as a mature and adequate way for plant phenotyping (Li, Zhang, & Huang, 2014). The reflectance spectrum of a leaf is influenced in the visible range (400–700 nm) by pigment concentrations (Gitelson, Gritz, & Merzlyak, 2003; Inoue et al., 2016), in the near-infrared (700–1300 nm) by leaf cellular structure (Peñuelas & Filella, 1998) and in the mid infrared (1300–3000 nm) mainly by radiation absorption by water but also by protein, lignin and cellulose (Downing, Carter, Holladay, & Cibula, 1993; Koch, Ammer, Schneider, & Wittmeier, 1990; Zhao et al., 2016). Multispectral imagery in the visible and near-infrared domains carries a lot of information on plant disease state (Sankaran, Mishra, Ehsani, & Davis, 2010). For winter wheat, it has been successfully set up in proximal sensing to detect SR infection at the growth stage of first node (GS 31; Zadoks, Chang, & Konzak, 1974) on an artificially inoculated cultivar (Moshou et al., 2004). Odilbekov, Armonienė, Henriksson, and Chawade (2018) offered the possibility of STB detection on winter wheat grown in greenhouse conditions. Until recently, studies in proximal disease detection on wheat often occur early in plant development when the canopy architecture is still open (Bravo, Moshou, West, McCartney, & Ramon, 2003; Krishna et al., 2014). Furthermore, the experimental setup is often combined with, on one side, healthy wheat and, on the other side, wheat artificially infected by a single disease. Although essential for the identification of wavebands where reflectance is altered by the disease infection, this first step in imaging-based wheat phenotyping needs to be taken further in real conditions. Over time in the field, wheat can reach high canopy densities which induce shadows and overlapping leaves, and can also be affected by multiple diseases at the same time. In-field phenotyping in natural conditions is still challenging but needs to be assessed in order to move forward towards the development of a robust and accurate high-throughput phenotyping machine.

Values of reflected light measured by spectral imagery are commonly exploited in vegetation indices to be correlated to reference phenotypic measurements (Devadas, Lamb, Simpfendorfer, & Backhouse, 2009; Franke & Menz, 2007; Nebiker, Lack, Abächerli, & Läderach, 2016). However, indices created for wheat phenotyping are often specific to the crop studied (Lowe, Harrison, & French, 2017; Verrelst, Koetz, Kneubühler, & Schaepman, 2006). Due to its intrinsic fine spatial resolution, proximal imaging can also provide relevant textural information. Indeed, disease symptoms and physiological modifications alter the leaf reflected light spectrum with different spatial arrangements (Martinelli et al., 2015). SR is characterised by small yellow-orange spores along leaf ridges. BR is quite similar, but spores are orange-brown and are uniformly distributed on leaf surface whereas STB induces the appearance of small black dots name pycnidia in brown-to-grey spots of different shapes (Fig. 1). Moreover, while rusts develop mainly on the upper leaves of plants, STB initially infects lower leaves and makes it harder to detect due to shading and occlusions.

Regarding data processing, several algorithms can be used to correlate measurements to observations. Partial least squares regression (PLSR) is a commonly used statistical analysis for plant phenotyping and in particular disease detection (Krishna et al., 2014; Zhang et al., 2012). With the technological progress, machine learning has proven to be fast and efficient when it comes to image classification (Mohanty, Hughes, & Salathé, 2016). Especially, artificial neural networks (ANN) and support vector machines have been successful in disease detection (Mahlein, Oerke, Steiner, & Dehne, 2012; Mewes, Franke, & Menz, 2011; Moshou et al., 2004; Rumpf et al., 2010). With image acquisition in natural conditions, the dataset will involve winter wheat at different growth stages and infected by multiple diseases which complicates plant phenotyping and requires the use of efficient algorithms of machine learning. To find the most suitable tool for winter wheat phenotyping, different approaches must be used and compared when predicting disease severity.

This study aims to investigate whether the severity of the three main diseases affecting winter wheat, namely SR, BR
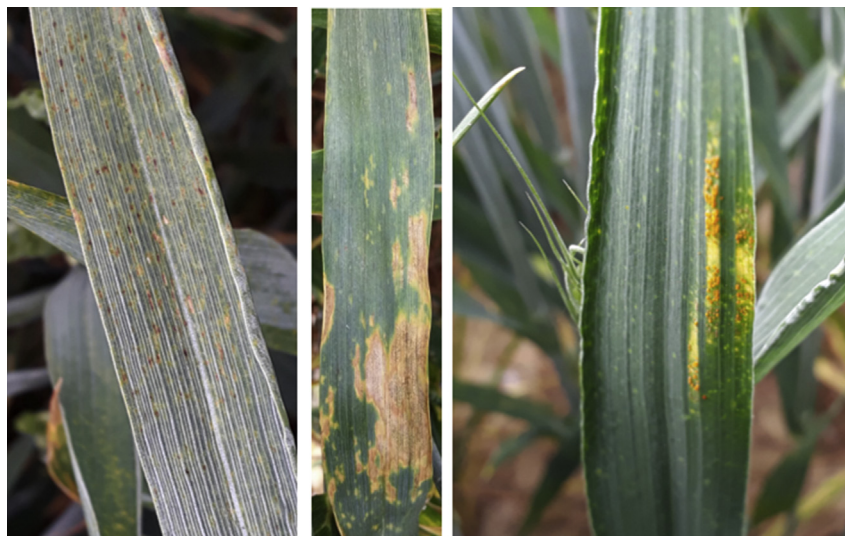
Fig. 1 – Typical disease symptoms for BR (left, photo taken by Sébastien Dandrifosse), STB (middle, photo taken by Charlotte Bataille, Walloon Agricultural Research Centre) and SR (right, photo taken by Sébastien Dandrifosse).

and STB, can be predicted using reflectance and pixel arrangement in specific wavelength bands in the visible and near-infrared domains. This study compares the performance of PLSR and ANN for disease severity regression based on textural analysis and pixel group reflectance from multispectral images. This research also intends to embrace every aspect of field conditions such as wind, natural disease inoculation and canopy architecture densification by observing five varieties of winter wheat with contrasting diseases sensitivities over five acquisition dates.

2. Materials and methods

2.1. Experimental field

The experiment took place in an agronomic trial studying the effect of cultivar and number of fungicide treatments on winter wheat yield. The field is located in Lonzée in Belgium (50°32′N; 4°44′E). Winter wheat was sown on 16th November 2017 with a density of 350 grains m^{-2}. Nitrogen fertilisation was applied at three specific growth stages: start of tillering (GS 21 according to the Zadok's scale), stem elongation (GS 31) and flag leaf (GS 39). This study focused on five cultivars of winter wheat (Alcides, Benchmark, Edgar, RGT Reform and Triomph) presenting contrasting degrees of sensitivity to SR, BR and STB. The cultivars have been exposed to no fungal treatment (0T) or two treatments on 8th and 24th May 2018 (2T) respectively with Opus team 1.5 l ha^{-1} + Bravo 1 l ha^{-1} and Aviator XPRO 1.25 l ha^{-1}. There were four replications. The five cultivars and the applied treatments were chosen in the interest of observing a large range of disease severity. The field was divided in micro-plots (8 × 2.05 m^2) including a small band on each side to limit border effects. Fungal inoculation occurred naturally and relied on the selection of cultivars of winter wheat sensitive to SR, BR and STB.

2.2. Disease score assessment

Disease severity for SR, BR and STB was assessed weekly from 2nd May (GS 31) to 3rd July 2018 (GS 83) by an expert in plant pathology. The assessment was performed on individual leaves randomly sampled. The severity was then averaged at the micro-plot scale. Rust severity represents a degree of leaf infection according to the modified Cobb scale which ranges from 0 to 100 (Peterson, Campbell, & Hannah, 1948). STB severity represents the percentage of the leaf surface presenting symptomatic lesions. In order to properly assess the disease infection throughout the growing season and account for the time evolution of the disease severity, the disease score of a plot at a given date is given by the area under the disease progress curve (AUDPC) as explained by Arora, Venkatesh, Sharma, and Saharan (2014). In order to analyse the contrast in infection between the varieties for each disease, the Tukey's honestly significant difference (HSD) test was used as a *post hoc* mean separation test (p < 0.05). This comparison test was performed for each date of image acquisition by interpolating the observed disease scores.

2.3. Multispectral acquisition system

A multispectral imagery system has been set up by means of a monochromatic 12-bit 1.3-megapixel BCi$_5$ CMOS camera (C-Cam technologies, Leuven, Belgium). For a given scene, this system acquired a set of 11 images through band-pass filters placed on a motorised wheel which successively presented the filters under the camera (Fig. 2). The wavebands of the filters were centred at wavelengths from 450 nm to 950 nm in steps of 50 nm with a bandwidth of 40 or 50 nm. The camera and wheel controllers were coded in C++. In order to take into account the variations of the natural illumination conditions, a reference plate made of a diffuse white coated plastic material was placed in the field of view of the camera at the
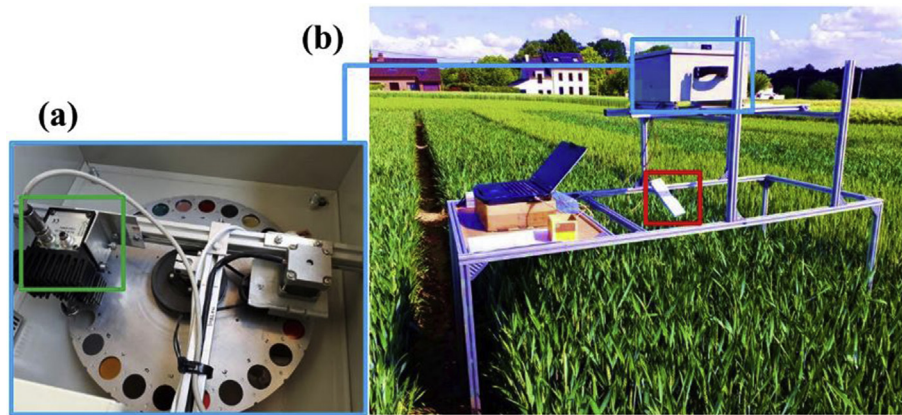
Fig. 2 – Experimental set up: (a) box containing the camera (green rectangle) and the motorised filter wheel, (b) platform composed of the box (blue rectangle), the white reference (red rectangle), the computer, the battery, and the gantry placed above a micro-plot. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1 – Dates of image acquisition and corresponding wheat growth stages according to the Zadok's scale.				
Date	25th May	30th May	6th June	13th June	21st June
Growth stage	GS 39	GS 59	GS 65	GS 71	GS 75
	Flag leaf	End of heading	Medium flowering	Watery ripe	Medium milk

canopy level. The integration time of the camera was adjusted to match the average reflectance of the white reference with a grey level of 0.95 ± 5% for each considered waveband. This allowed benefit to be taken of the colour depth resolution while avoiding pixel saturation and normalisation of the monochromatic images between acquisition dates for a given waveband. The vision system was placed at 1 m above the canopy which results in a spatial resolution of 1 mm at the top level of the winter wheat cover.

2.4. Image acquisition

The raw dataset consists of 630 multispectral images, each of them composed of 11 single-waveband monochromatic layers. These images were acquired on five different dates from 25th May to 21st June 2018 (Table 1). In order to account for the intra-plot variability, four and three marked scenes were shot for 0T and 2T, respectively.

2.5. Image treatment

An image pre-treatment was performed in order to extract a unique leaf segmentation mask for all the layers of a multispectral acquisition. Due to the time sequence of the optical filters, the wind could induce leaf displacement between the layers. This would make it difficult to extract a unique vegetation mask through all the wavebands and then disturb the disease detection algorithms. To assess the wind effect during a multispectral shot, the images obtained with the first, the fifth and the eleventh wavebands in the acquisition sequence

were compared. Since, these wavebands were centred in the near-infrared (i.e. 950 nm, 900 nm and 850 nm, respectively), leaf segmentation could be performed on each layer independently due to the high soil-plant reflectance contrast. The distributions of segmentation difference compared two by two are depicted in Fig. 3. The difference minimum is at 4% and the distribution peaks occurs at around 15% which means that no multispectral image presented perfect leaf-scale superimposition between the layers.

The wind blew strongly during the whole day of 21st June 2018. Images from this acquisition day are mostly on the right part of the graphs of Fig. 3 with 81% of images above a segmentation difference of 30%. This validates that segmentation difference is a good indicator of wind effects on acquisition quality. Images of poor quality were removed by identifying outliers by means of the method proposed by Leys, Ley, Klein, Bernard, & Licata (2013). The outlier detection rule is given by

$$x_i > M(x) + D_l \, b \, M(|x_i - M(x)|)$$

where x is the considered distribution, $M(\cdot)$ is the median operator, x_i is the tested value of the distribution, D_l is the level of decision and b is a constant equal to 1.4826 (Rousseeuw & Croux, 1993). This method is used for non-normal distribution as the median is barely sensitive to outliers. The level of decision D_l is arbitrary and was set to 3, which was motivated in this study by the need to have a large dataset for regression. This criterion leads to a segmentation difference threshold of 34%, above which the multispectral images were removed from the data set. This appears to be a good compromise
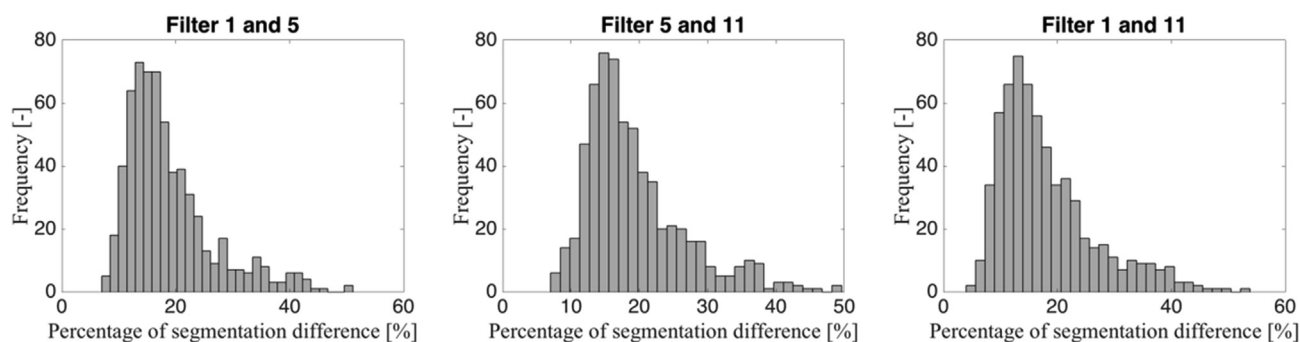
Fig. 3 − Histogram of percentage leaf segmentation difference between the first (950 nm), fifth (900 nm) and eleventh (850 nm) filters of the acquisition sequence.

between high quality data and a sufficient quantity of data for algorithm training. As a consequence, the number of images was narrowed down to 562.

Neural network-based leaf segmentation was operated on these multispectral images using the 11 waveband layers. In order to take into account the natural lighting contrast, pixels belonging to leaves were separated into three groups depending on the lighting conditions they had been exposed to, i.e. sunny, shady or cloudy conditions (Fig. 4). Moreover, since STB infection starts from the bottom of the canopy, the associated symptoms might be preferably located in the shady zones. A fourth group named 'total' gathered every plant pixel.

A first set of layer features was extracted by averaging the reflectance values of the four aforementioned groups of plant pixel (sunny, shady, cloudy and total). In addition, statistical textural features were computed for the group of all plant pixels (total) of each waveband layer by means of an associated grey level co-occurrence matrix. Such a matrix represents the frequency at which any combination of pixel intensity values occurs in the image by considering adjacent pixel pairs. Four scalar properties, defined as contrast, correlation, energy and homogeneity, are directly deduced from the grey level co-occurrence matrix and quantify how the pixel intensity values are arranged in the whole image (Baraldi & Parmiggiani (1995)). These eight features, i.e. the average reflectance value of the four leaf groups and the four textural parameters of the total leaf group, were computed for each waveband resulting in 88 features per observed scene for disease detection. The
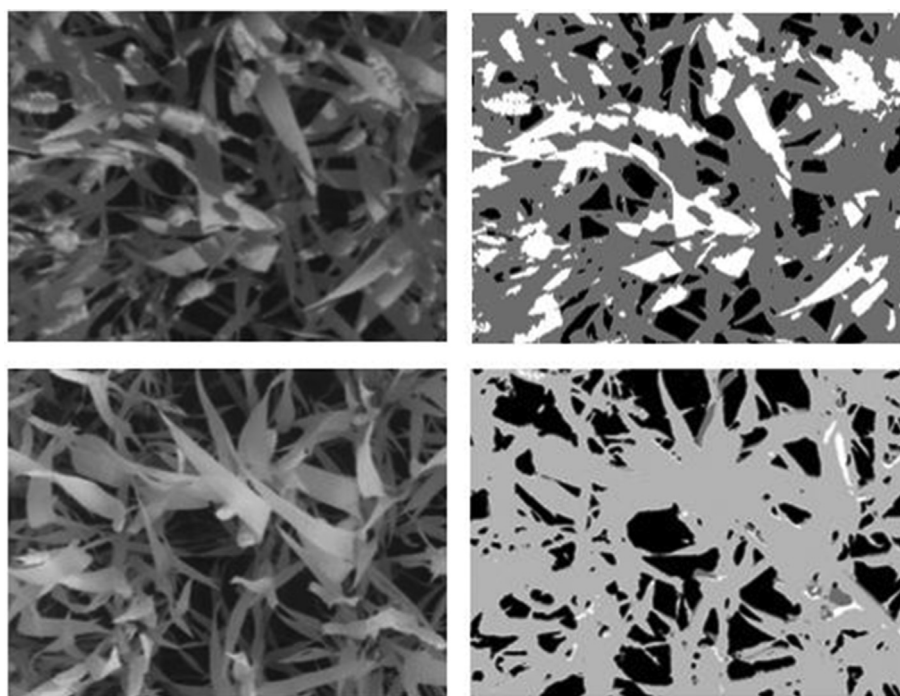

Fig. 4 − Example of a raw canopy image (left) and a segmented image (right) in contrasting illumination conditions, sunny condition (top) and cloudy condition (bottom). Pixels are classified as soil (black) or plant. Plant pixels are separated in three classes: sunny (white), shady (dark grey) and cloudy (light grey).
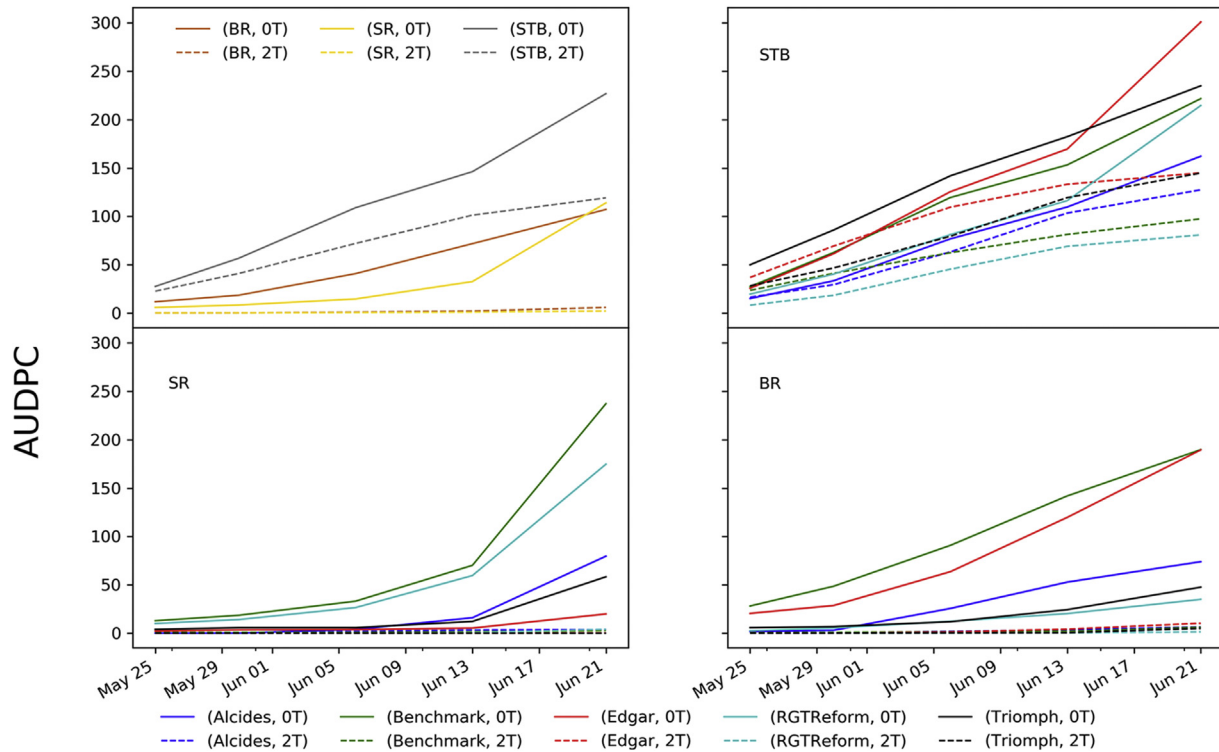
Fig. 5 — Time evolution of the observed infection score based on the area under the disease progress curve (AUDPC) during the 2018 season: average score over the five varieties (top-left), STB score (top-right), SR score (bottom-left), BR score (bottom-right).

proposed image treatment was implemented in Matlab by using precompiled routines.

2.6. Regression algorithms

Based on the spectral and textural features extracted from the multispectral images, disease detection models were established by means of PLSR and ANN. PLSR is specifically adapted when number of features is high (Tenenhaus, Gauchi, & Ménardo, 1995). This algorithm constructs a regression model based on a small number of uncorrelated latent variables built from the independent features. The optimal model was selected by minimising the root mean square error (RMSE). ANN were chosen for their capacity to resolve complex problems and to generalise a small amount of data (Rafiq, Bugmann, & Easterbrook, 2001). ANN require however a smaller amount of features. Based on a bidirectional stepwise regression (StepReg), variables were selected if their p-value was lower than a given threshold. Each time a new variable enters the model, already selected variables were tested and taken out of the model if their p-value was superior to another given threshold. These thresholds were fixed at 0.3 and 0.7, respectively, in order to have a sufficient number of features. This allows removing correlated variables. Remaining variables are ranked by order of increasing p-value to identify the most significant. Multiple ANN-based models were built by means of the most significant features by successively considering one to 15 input features. ANN architecture was composed of one hidden layer of two to nine neurons. The

optimal model for disease severity prediction was picked by minimising the RMSE. The model hyper-parameter tuning was set by splitting the dataset into 80% for training and 20% for testing for both regression methods. Image processing and algorithms were implemented in Matlab using the neural network toolbox for ANN and the statistics and machine learning toolbox for PLSR.

2.7. Image classification

Classification between healthy and diseased wheat crops is usually performed in the literature between treated plots and untreated inoculated plots. In this contribution, natural inoculation and the use of cultivars with contrasting disease sensitivities did not allow the direct identification of two distinct classes. In order to investigate the potential for health status discrimination, images were classified into two classes according to an arbitrarily chosen threshold value of AUDPC disease score. As the AUDPC values range from 0 to 300 for each disease, three infection threshold values were successively considered as 20, 50 and 100. These will be further referred as low, intermediate and high levels of infection threshold values, respectively. To evaluate the quality of classification, sensitivity or true positive rate (TPR), specificity or true negative rate (TNR) and accuracy (Acc) were deduced as follows:

$$TPR = \frac{TP}{TP + FN} \quad TNR = \frac{TN}{TN + FP} \quad Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP is the true positive number, TN is the true negative

Table 2 – Tukey's HSD multiple comparison test for the levels of infection of each fungal disease at each image acquisition date. The letters indicate the groups of cultivars combined with (2T) or without (0T) fungal treatment which are statistically different per date and per disease at the p < 0.05 level.

	STB					SR					BR				
	May 25	May 30	June 6	June 13	June 21	May 25	May 30	June 6	June 13	June 21	May 25	May 30	June 6	June 13	June 21
Alcides, 0T	b	bc	abc	ab	bcd	b	b	b	b	c	b	c	b	b	b
Alcides, 2T	b	bc	bc	ab	cd	b	b	b	b	e	b	c	b	b	b
Benchmark, 0T	ab	abc	abc	ab	abc	a	a	a	a	a	a	a	a	a	a
Benchmark, 2T	ab	abc	bc	b	d	b	b	b	b	e	b	c	b	b	b
Edgar, 0T	ab	abc	ab	a	a	ab	ab	b	b	de	a	b	a	a	a
Edgar, 2T	ab	ab	abc	ab	bcd	b	b	b	b	e	b	c	b	b	b
RGT Reform, 0T	ab	abc	abc	ab	abc	ab	ab	a	a	b	b	c	b	b	b
RGT Reform, 2T	b	c	c	b	d	b	b	b	b	e	b	c	b	b	b
Triomph, 0T	a	a	a	a	ab	ab	ab	b	b	cd	b	c	b	b	b
Triomph, 2T	ab	abc	abc	ab	bcd	b	b	b	b	e	b	c	b	b	b

number, FP is the false positive number and FN is the false negative number. A predicted disease score is counted as positive if higher than the considered threshold value, whereas a lower predicted disease score is classified as negative. The true and false qualifications refer to an agreement and a disagreement with the observed score, respectively. The disease classifications were also assessed by means of the Cohen's coefficient κ which reads

$$\kappa = \frac{p_a - p_e}{1 - p_e}$$

where p_a is the probability of agreement between both disease assessment methods, i.e. $p_a = Acc$, and p_e is the probability of random agreement. The Cohen's coefficient is known to be more robust than the accuracy parameter since it accounts for the probability of true values occurring randomly.

3. Results and discussion

3.1. Disease severity

During the 2018 season, the weather conditions led to a high level of natural STB infection for all cultivars. As depicted in Fig. 5, the applied treatment did not allow maintaining healthy micro-plots. However, Table 2 shows that the score of STB can hardly be statistically differentiated across cultivars. Regarding BR and SR, the infection was nicely controlled by the fungicides. The monitored wheat cultivars with contrasting disease sensitivities gave expected results and offer a dataset with a wide range of disease severity (Fig. 5). However, both rusts were characterised by some heterogeneities in the level of infection between the cultivars. SR began late in the season but was quite virulent, especially on cultivar Benchmark and RGT Reform which can be statistically separated from the other cultivars from 6th June and even from each other for the last acquisition date, as explained in Table 2. In contrast, Benchmark and Edgar were significantly more attacked by BR than the other cultivars with a relative regular time evolution since 25th May.

3.2. Cultivar reflectance

Cultivars of wheat can differ from each other in terms of, for instance, number of leaves, leaf size, leaf tilting and the presence of bearded ears. These morphological differences can induce changes in reflectance from one cultivar to another. Figure 6 shows the mean values of reflectance for treated (2T) micro-plots of the five studied cultivars. The treated cultivars were unaffected by the rusts and slightly but evenly infected by STB. Among the considered cultivars, Benchmark has higher standard deviation than the other cultivars in the visible domain and slightly increased reflectance at 650 and 700 nm. In the red edge and in NIR, however, RGT Reform differs from the other cultivars due to its higher standard deviation and lower reflectance. Benchmark has also lower reflectance than Alcides, Edgar and Triomph at 750 nm. These inter-cultivar differences are rather small compared to standard deviation but may confuse regression algorithms. The study of several cultivars at different growth stages should allow the construction of robust models.

3.3. Inter-temporal reflectance

The multi-temporal assessment of the disease score has to take into account reflectance variations due to plant growth. Figure 7 shows the values of reflectance for treated (2T) micro-plots of cultivars at the monitored growth stages (Table 1). On 25th May (GS 39) the flag leaf was fully developed, which explains the high level of NIR reflectance. It is also observed that the reflectance in the red band increases with time while NIR reflectance decreases. On 21st June (GS 75), reflectance dropped at every wavelength. The variations in reflectance between growth stages are higher than the inter-cultivar variations especially for the last date. This may hide variations due to a disease infection, the senescence stage onset or a photosynthetic activity decrease.

3.4. Disease score regression

The regression results are shown in Figs. 8–10 for STB, SR and BR, respectively. Corresponding R^2 and RMSE are listed in
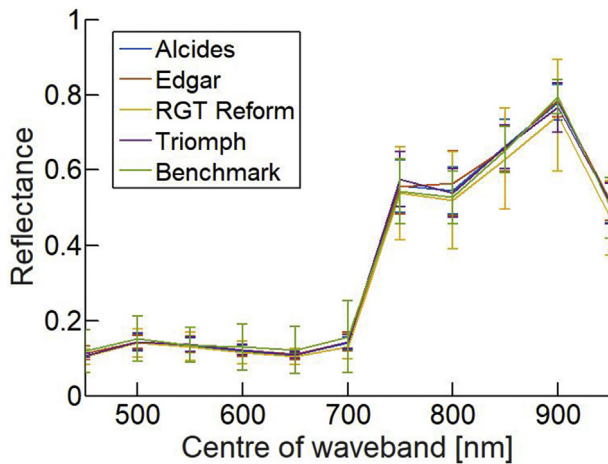
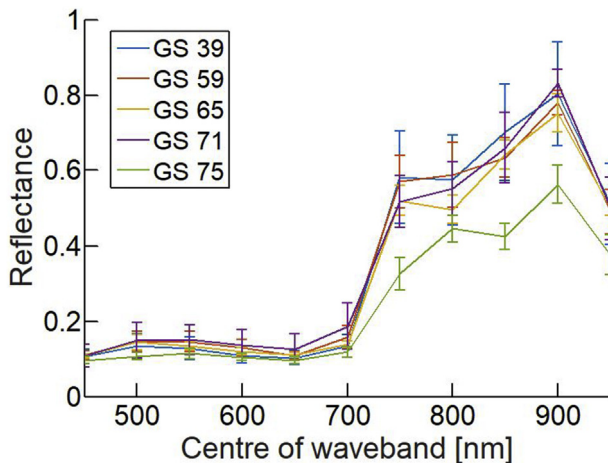**Fig. 6 – Averaged reflectance ± standard deviation of treated wheat for each cultivar.**

Fig. 7 – Averaged reflectance ± standard deviation of treated wheat at each observed growth stage (GS 39, flag leaf; GS 59, end of healing; GS 65, medium flowering, GS 71, watery ripe; GS 75, medium milk).

Table 3. The regression of STB disease score shows an increasing trend for PLSR and ANN (Fig. 8). The ANN gives better predictions, especially in terms of RMSE, with R^2 predicted and RMSE predicted of 0.72 and 34.9 respectively for ANN and 0.69 and 37.1 respectively for PLSR (Table 3).

For SR, ANN gives more accurate predictions in terms of R^2 (0.57) and RMSE (23.7) than PLSR which gives R^2 of 0.19 and RMSE of 32.6 (Table 3). In comparison with ANN, PLSR overestimates points observed at AUDPC values close to zero and greatly underestimates points above observed values of 100. Figure 9 displays clearly two groups of observations. For AUDPC values below 100, regression is poor in quality and SR is often undetected. However, for AUDPC observed above 100, all observations were predicted around 100 by ANN which reveals potential for discrimination in two classes. The number of observations above 100 was however rather small due to a low natural infection level during the 2018 season.

BR had slightly infected the winter wheat field during the acquisition period. Benchmark and Edgar cultivars were slightly but steadily infected, which provides over time observations at high AUDPC. Severity prediction is once again better with ANN which reached predicted R^2 of 0.65 and RMSE of 27.2 (Table 3). However, the R^2 and RMSE were boosted by many points at low AUDPC and the lack of observations at high AUDPC values (Fig. 10).

The optimal model depends on the disease being estimated. The number of uncorrelated variables of PLSR optimising RMSE for STB, SR and BR, respectively is presented in Table 4. This also shows, for each disease, the number of predictive features obtained from StepReg and the number of neurons used in the hidden layer of ANN. STB requires more features for both models, which is related to the broader range of disease score with respect to SR and BR. For better performance, the numbers of predictive features for ANN are equivalent to the PLSR models, or even less for the case of SR. The superiority of ANN could be partially explained by its intrinsic ability to model non-linear relationships (Barmpalexis, Karagianni, Nikolakakis, & Kachrimanis, 2018; Lin, Groves, Freivalds, Lee, & Harper, 2012). In addition, whereas PLSR uses rebuilt uncorrelated variables, ANN works directly on spectral and textural features from raw images. As a consequence, the latter has the ability to provide easily interpretable results with the aim of eventually designing dedicated in-field acquisition systems appropriate for
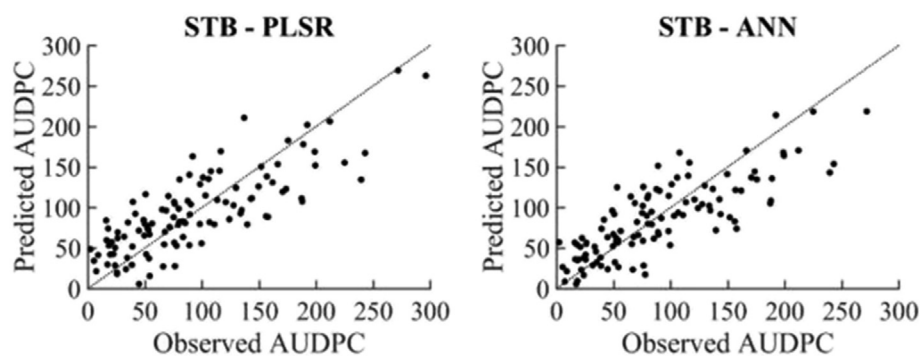

Fig. 8 – Regression results between observed AUDPC of STB and predicted AUDPC by partial least squares regression (left) and artificial neural networks (right).

Fig. 9 – **Regression results between observed AUDPC of SR and predicted AUDPC by partial least squares regression (left) and artificial neural networks (right).**

Fig. 10 – **Regression results between observed AUDPC of BR and predicted AUDPC by partial least squares regression (left) and artificial neural networks (right).**
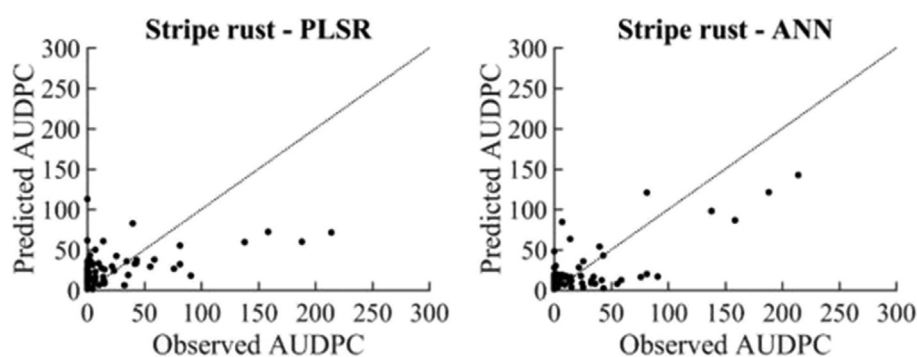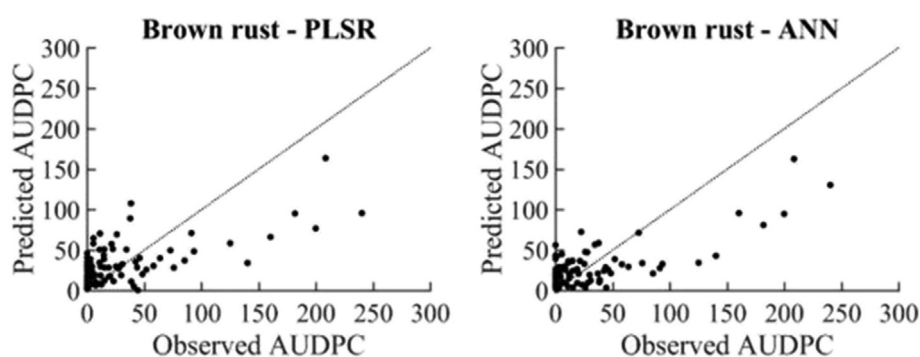
precision agriculture. For these reasons, the sequel of this discussion will be based on ANN results.

For STB, the ANN-based predictions have been computed using 13 input features ranked by increasing p-values in Table 5. Physiological changes induced by STB are known to reduce reflectance in the red edge and NIR wavelengths (Moshou et al., 2011). The waveband at 750 nm, which covers the red edge-NIR transition, is present three times in the top six features. The lack of other red edge and NIR bands in the most discriminating features is explained by the use of the stepwise selection. Red edge selection conforms to the recent article by Yu et al. (2018), which also highlighted such a region (730 nm) for STB detection. However, it is shown in Table 5 that the most important features in this study are 600 nm and 650 nm.

Table 3 – Summary of R^2 and RMSE of the training and prediction sets for PLSR and ANN regression of STB, SR and BR disease score.

	STB		SR		BR	
	PLSR	ANN	PLSR	ANN	PLSR	ANN
Predicted R^2	0.69	0.72	0.19	0.57	0.41	0.65
Predicted RMSE	37.1	34.9	32.6	23.7	35.3	27.2
Trained adjusted R^2	0.63	0.79	0.37	0.55	0.46	0.68
Trained RMSE	38.1	30.4	30.5	26.7	31.2	25.3

Table 4 – Number of parameters (predictive features/hidden neurons/uncorrelated variables) used in the optimal ANN-based and PLSR-based models.

Disease	Predictive features number (StepReg)	Hidden neurons number (ANN)	PLSR uncorrelated variables number
STB	13	6	12
SR	5	3	10
BR	8	4	9

Table 5 – Features selected by stepwise regression for STB disease score assessment, ranked by increasing p-values.

Waveband [nm]	Feature
600	Mean cloudy
600	Energy
650	Homogeneity
750	Contrast
750	Mean cloudy
750	Homogeneity
650	Mean sunny
650	Energy
550	Homogeneity
700	Energy
800	Correlation
850	Homogeneity
900	Contrast

As has been discussed before, reflectance in the red edge and NIR depends greatly on the growth stage (Fig. 5) and may therefore bring biased information for STB regression. Instead, selected features cover the wavelengths corresponding partially to yellow which is the colour of blotches of STB and dried leaves due to senescence. With ten variables in the model, textural features showed that they are crucial for disease score regression.

In addition, cloudy conditions brought poor contrast in the image due to the lack of shade. This helped to compare between leaves and gave features of good quality for disease regression (Table 5). Unexpectedly, no feature derived from shady conditions was selected even though STB infection usually develops in lower leaves, shaded by upper leaves. This could be explained by the fact that, at advanced growth stages, almost all the leaves are shaded except the flag leaves. Moreover, as the winter wheat grows, fewer lower leaves remain visible by the camera due to covering by upper leaves.

SR is known to have an increase in reflectance in the green-red bands (500–690 nm) and in the NIR bands (750–1300 nm) with a maximum difference at 620–670 nm (Yuan et al., 2014). By means of a hyperspectral vision system, Bravo, Moshou, West, MacCartney, & Ramon (2003) identified 543, 630, 750 and 861 nm as key wavebands. However, in this study the selected features only cover the red edge (700, 750 nm) and NIR (900 nm) (Table 6). The regression has been influenced by STB infection which was dominant in the agronomic trial, especially for the Benchmark cultivar. The combined effects of both diseases can disturb AUDPC regressions and predictions of SR, which explains the lack of wavebands in the yellow range and the poor regression at low AUDPC. As for STB, textural features were the prevailing type in the stepwise selection. As shown previously in Fig. 7, the growth stage highly influences reflectance, especially in the red edge and in NIR, which can hide modifications of reflectance in these bands induced by disease infection. Textural features are more reliable for disease regression with multi-temporal data because local modifications due to diseases symptoms on the leaf can be compared with the healthy neighbouring pixels.

The typical brown-orange colour of BR spores is located around 600 nm (Wahabzada et al., 2015), which is partially covered by the most discriminating feature at 650 nm waveband (Table 7). Other features include wavelength in the NIR (950 nm) and in the red edge (750 nm). The prevailing type of feature is textural with homogeneity and contrast in the top three selected features. Textural analysis is adapted to the typical symptoms of BR, consisting of small dots uniformly distributed on the leaf. The use of features in the NIR and red

Table 6 – Features selected by stepwise regression for SR disease score assessment, ranked by increasing p-values.

Waveband [nm]	Feature
750	Contrast
750	Correlation
900	Mean total
900	Contrast
700	Energy

Table 7 – Features selected by stepwise regression for BR disease score assessment, ranked by increasing p-values.

Waveband [nm]	Feature
650	Homogeneity
950	Contrast
750	Contrast
450	Mean cloudy
800	Mean cloudy
650	Energy
700	Homogeneity
850	Mean cloudy

edge bands, where BR hardly induces any changes in reflectance (Wahabzada et al., 2015), is due to the need to predict AUDPC. A high AUDPC means that wheat had previously been strongly infected, which should have caused changes in its health and physiological status. Besides, STB infection induces changes in the red edge and NIR bands and has also BR infected wheat. These bands may not be directly used to detect BR, but to differentiate BR from STB infection.

As discussed above, the waveband selection globally conforms to other studies which generally measure the canopy reflectance in natural conditions by means of spectroradiometer. Such devices offer a higher spectral resolution but do not allow distinguishing elements of the scene such as leaves. Hyperspectral imagery can take full advantage of both spectral and spatial resolution but remains complex and costly to implement in field conditions.

3.5. Image classification based on AUDPC

For STB, Table 8 shows the confusion matrices of AUDPC two-class discriminations considering separately three arbitrarily chosen threshold values. At low and intermediate threshold, STB was not well classified due to a lack of observations below 20 resulting in a low specificity. At high threshold, sensitivity and specificity had moderately high and comparable values. STB discrimination reached good accuracy at high threshold considering the combination of factors affecting wheat in natural conditions. Predictions and observation were in moderate agreement for intermediate and high infection in terms of Cohen's coefficient.

Although specificity and accuracy were good, SR discrimination suffered from low sensitivity at every infection level due to too many points observed above the threshold values but predicted below (Table 9). As most points have AUDPC values below 20, the algorithm tends to optimise its accuracy by predicting systematically low AUDPC values. The moderate agreement for intermediate and high infection level ($\kappa = 0.56$ and 0.60) shows however potential for discrimination in two classes and discrimination should be assessed in future years when conditions are suitable for significant SR infection.

BR classifications were very similar to SR with low sensitivity (Table 10). At low threshold, however, sensitivity and specificity had similar values but the accuracy was only at 68% with a fair agreement of 0.35. The slightly lower quality for BR classifications with respect to SR could be partially

Table 8 – Confusion matrices of the classification with respect to a low (20), intermediate (50) and high (100) AUDPC threshold value for STB.

	AUDPC threshold 20		AUDPC threshold 50		AUDPC threshold 100	
	Observed below	Observed above	Observed below	Observed above	Observed below	Observed above
Predicted below	3	2	19	5	60	10
Predicted above	8	100	14	75	11	32
	Sensitivity	98%	Sensitivity	94%	Sensitivity	76%
	Specificity	27%	Specificity	58%	Specificity	85%
	Accuracy	91%	Accuracy	83%	Accuracy	81%
	Kappa	0.33	Kappa	0.56	Kappa	0.60

Table 9 – Confusion matrices of the classification with respect to a low (20), intermediate (50) and high (100) AUDPC threshold value for SR.

	AUDPC threshold 20		AUDPC threshold 50		AUDPC threshold 100	
	Observed below	Observed above	Observed below	Observed above	Observed below	Observed above
Predicted below	82	13	99	5	107	2
Predicted above	7	10	3	5	1	2
	Sensitivity	43%	Sensitivity	50%	Sensitivity	50%
	Specificity	92%	Specificity	97%	Specificity	99%
	Accuracy	82%	Accuracy	93%	Accuracy	97%
	Kappa	0.40	Kappa	0.51	Kappa	0.56

Table 10 – Confusion matrices of the classification with respect to a low (20), intermediate (50) and high (100) AUDPC threshold value for BR.

	AUDPC threshold 20		AUDPC threshold 50		AUDPC threshold 100	
	Observed below	Observed above	Observed below	Observed above	Observed below	Observed above
Predicted below	48	10	93	9	105	5
Predicted above	26	28	4	6	0	2
	Sensitivity	74%	Sensitivity	40%	Sensitivity	29%
	Specificity	65%	Specificity	96%	Specificity	100%
	Accuracy	68%	Accuracy	88%	Accuracy	96%
	Kappa	0.35	Kappa	0.42	Kappa	0.43

explained by the rather progressive time evolution of BR infection.

Currently, AUDPC is not directly used in agricultural practice. Based on the proposed disease assessment, the decision of fungal treatment application could be linked to such a variable. However, for the case of STB the disease has to be treated only if the infection reaches the upper leaves. This spatial dimension is not assessed by the AUDPC variable. The classification results based on AUDPC show nonetheless good potential for high-throughput phenotyping in agronomic trials.

4. Conclusions

This study needed to take into account the difference in reflectance due to the cultivar but also the change in reflectance over time as images from flag leaf emergence to grain development at medium milk were mixed in one dataset. Despite the combination of stresses observed in natural conditions, some wavebands stood out with 750, 600 and 650 nm for STB, 700, 750 and 900 nm for SR and 650, 750 and 950 nm for BR. Moreover, the textural analysis and the segmentation according to the illumination conditions have been crucial for AUDPC regression. For each disease, ANN reached higher performance than PLSR and proved their efficiency in resolving complex problems. The selected wavebands allowed health status discrimination thanks to ANN at high AUDPC (100) with an accuracy of 81% ($\kappa = 0.60$) for STB. SR and BR lack observations at high AUDPC. Better results can be expected from further studies with earlier infection at high levels offering a more significant number of contrasting disease scores.

Regarding agricultural practice, further studies should focus on disease discrimination based on a disease score related to curative action. For cultivar phenotyping, the

assessment method appears truly promising as good accuracy was reached for STB in difficult conditions. Pixel depth measurement and image co-registration should improve the regression results and allow disease discrimination especially between STB and rusts. This study should be repeated in future years to measure different natural disease development scenarios in order to consolidate the relevant wavebands.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

REFERENCES

Arora, A., Venkatesh, K., Sharma, R. K., Saharan, M. S., Dilbaghi, N., Sharma, I., et al. (2014). Evaluating vegetation indices for precision phenotyping of quantitative stripe rust reaction in wheat. *Journal of Wheat Research, 6*(1), 74–80.

Baraldi, A., & Parmiggiani, F. (1995). Investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing, 33*(2), 293–308.

Barmpalexis, P., Karagianni, A., Nikolakakis, I., & Kachrimanis, K. (2018). Artificial neural networks (ANNs) and partial least squares (PLS) regression in the quantitative analysis of cocrystal formulations by Raman and ATR-FTIR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis, 158*, 214–224.

Bodson, B., De Proft, M., & Watillon, B. (2017). *Livre Blanc Céréales. Septembre 2017* (in French).

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.

Devadas, R., Lamb, D. W., Simpfendorfer, S., & Backhouse, D. (2009). Evaluating ten spectral vegetation indices for identifying rust infection in individual wheat leaves. *Precision Agriculture, 10*, 459–470.

Downing, H. G., Carter, G. A., Holladay, K. W., & Cibula, W. G. (1993). The radiative-equivalent water thickness of leaves. *Remote Sensing of Environment, 46*(1), 103–107.

Filella, I., Serrano, L., Serra, J., & Peñuelas, J. (1995). Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. *Crop Science, 35*(5), 1400–1405.

Franke, J., & Menz, G. (2007). Multi-temporal wheat disease detection by multi-spectral remote sensing. *Precision Agriculture, 8*, 161–172.

Gitelson, A. A., Gritz, Y., & Merzlyak, 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.

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment, 90*, 337–352.

Inoue, Y., Guérif, M., Baret, F., Skidmore, A., Gitelson, A. A., Schlerf, M., et al. (2016). Simple and robust methods for remote sensing of canopy chlorophyll content: A comparative analysis of hyperspectral data for different types of vegetation. *Plant, Cell and Environment, 39*, 2609–2623.

Koch, B., Ammer, U., Schneider, T., & Wittmeier, H. (1990). Spectroradiometer measurements in the laboratory and in the field to analyse the influence of different damage symptoms on the reflection spectra of forest trees. *International Journal of Remote Sensing, 11*(7), 1145–1163.

Krishna, G., Sahoo, R. N., Pargal, S., Gupta, V. K., Sinha, P., Bhagat, S., et al. (2014). Assessing wheat yellow rust disease through hyperspectral remote sensing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, XL*(8), 1413–1416.

Kuska, M., Wahabzada, M., Leucker, M., Dehne, H.-W., Kersting, K., Oerke, E.-C., et al. (2015). Hyperspectral phenotyping on the microscopic scale: Towards automated characterization of plant-pathogen interactions. *Plant Methods, 11*, 28.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*, 764–766.

Lindblom, J., Lundström, C., Ljung, M., & Jonsson, A. (2017). Promoting sustainable intensification in precision agriculture: Review of decision support systems development and strategies. *Precision Agriculture, 18*, 309–331.

Lin, M. B., Groves, W. A., Freivalds, A., Lee, E. G., & Harper, M. (2012). Comparison of artificial neural network (ANN) and partial least squares (PLS) regression models for predicting respiratory ventilation: An exploratory study. *European Journal of Applied Physiology, 112*, 1603–1611.

Li, L., Zhang, Q., & Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors (Switzerland), 14*, 20078–20111.

Lowe, A., Harrison, N., & French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods, 13*, 80.

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors - Paralles and specific demands for precision agriculture and plant phenotyping. *Plant Disease, 100*(2), 241–251.

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*, 197–209.

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., et al. (2015). Advanced methods of plant disease detection. A review. *Agronomy for Sustainable Development, 35*, 1–25.

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

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419.

Moshou, D., Bravo, C., Oberti, R., West, J. S., Ramon, H., Vougioukas, S., et al. (2011). Intelligent multi-sensor system for the detection and treatment of fungal diseases in arable crops. *Biosystems Engineering, 108*, 311–321.

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*(3), 173–188.

Naik, H. S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., et al. (2017). A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods, 13*, 23.

Nebiker, S., Lack, N., Abächerli, M., & Läderach, S. (2016). Light-weight multispectral uav sensors and their capabilities for predicting grain yield and detecting plant diseases. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B1*, 963–970.

Odilbekov, F., Armonienė, R., Henriksson, T., & Chawade, A. (2018). Proximal phenotyping and machine learning methods to identify septoria tritici blotch disease symptoms in wheat. *Frontiers in Plant Science, 9*, 685.

Peñuelas, J., & Filella, L. (1998). Technical focus: Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science, 3*(4), 151–156.

Peterson, R. F., Campbell, A. B., & Hannah, A. E. (1948). A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. *Canadian Journal of Research, 26*(5), 496–500.

Rafiq, M. Y., Bugmann, G., & Easterbrook, D. J. (2001). Neural network design for engineering applications. *Computers & Structures, 79*, 1541–1552.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association, 88*(424), 1273–1283.

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dhene, 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.

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture, 72*, 1–13.

Tenenhaus, M., Gauchi, J.-P., & Ménardo, C. (1995). *Revue de statistique appliquée - Régression PLS et applications (PLS regression and applications)* (Vol. 43). Retrieved from http://www.sfds.asso.fr/publicat/rsa.htm.

Van Der Werf, H. M. G. (1996). Assessing the impact of pesticides on the environment. *Agriculture, Ecosystems & Environment, 60*, 81–96.

Verrelst, J., Koetz, B., Kneubühler, M., & Schaepman, M. (2006). Directional sensitivity analysis of vegetation indices from multi-angular CHRIS/PROBA data. *ISPRS commission VII-term symposium*, 677–683.

Wahabzada, M., Mahlein, A.-K., Bauckhage, C., Steiner, U., Oerke, E.-C., & Kersting, K. (2015). Metro maps of plant disease dynamics—Automated mining of differences using hyperspectral images. *PloS One, 10*(1), Article e0116902.

Yu, K., Anderegg, J., Mikaberidze, A., Karisto, P., Mascher, F., McDonald, B. A., et al. (2018). Hyperspectral canopy sensing of wheat septoria tritici blotch disease. *Frontiers in Plant Science, 9*, 1195.

Yuan, L., Huang, Y., Loraamm, R. W., Nie, C., Wang, J., & Zhang, J. (2014). Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects. *Field Crops Research, 156*, 199–207.

Zadoks, J. C., Chang, T. T., & Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Research, 14*(6), 415–421.

Zhang, J., Pu, R., Wang, J., Huang, W., Yuan, L., & Luo, J. (2012). Detecting powdery mildew of winter wheat using leaf level hyperspectral measurements. *Computers and Electronics in Agriculture, 85*, 13–23.

Zhao, S., Wang, Q., Yao, Y., Du, S., Zhang, C., Li, J., et al. (2016). Estimating and validating wheat leaf water content with three MODIS spectral indexes: A case study in Ningxia plain, China. *Journal of Agricultural Science and Technology, 18*, 387–398.