REVIEW

Open Access

From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy

Clive H. Bock[1*], Jayme G. A. Barbedo[2*], Emerson M. Del Ponte[3], David Bohnenkamp[4] and Anne-Katrin Mahlein[5*]

Abstract

The severity of plant diseases, traditionally the proportion of the plant tissue exhibiting symptoms, is a key quantitative variable to know for many diseases and is prone to error. Good quality disease severity data should be accurate (close to the true value). Earliest quantification of disease severity was by visual estimates. Sensor-based image analysis including visible spectrum and hyperspectral and multispectral sensors are established technologies that promise to substitute, or complement visual ratings. Indeed, these technologies have measured disease severity accurately under controlled conditions but are yet to demonstrate their full potential for accurate measurement under field conditions. Sensor technology is advancing rapidly, and artificial intelligence may help overcome issues for automating severity measurement under hyper-variable field conditions. The adoption of appropriate scales, training, instruction and aids (standard area diagrams) has contributed to improved accuracy of visual estimates. The apogee of accuracy for visual estimation is likely being approached, and any remaining increases in accuracy are likely to be small. Due to automation and rapidity, sensor-based measurement offers potential advantages compared with visual estimates, but the latter will remain important for years to come. Mobile, automated sensor-based systems will become increasingly common in controlled conditions and, eventually, in the field for measuring plant disease severity for the purpose of research and decision making.

Keywords: Disease severity, Assessment, Sensor, Mobile device, Digital technologies, Artificial intelligence, Machine learning, Deep learning, Phenotyping, Precision agriculture, Accuracy, Precision

Background

Plant disease epidemics impact agriculture and forestry by reducing the quantity and quality of the product, and pose a threat to food security and food safety (Strange and Scott 2005; Oerke 2006; Madden et al. 2007; Savary et al. 2012, 2017). Knowledge of the quantity of disease is fundamental to a) determine crop losses; b) conduct disease surveys; c) establish thresholds for decision-making; d) improve knowledge of disease epidemiology, and e) evaluate the effect of treatments (e.g. cultivar, fungicides, etc). Plant disease intensity (a generic term) can be expressed by incidence or severity at the field/plot scale and below. Incidence is the proportion of the plant units that are diseased in a defined population or sample (Madden et al. 2007) while severity is the proportion of the plant unit exhibiting visible disease symptoms, usually expressed as a percentage (Madden et al. 2007). Symptoms of disease on a plant may change in size, shape and color. Disease severity is often the

* Correspondence: clive.bock@usda.gov; jayme.barbedo@embrapa.br; mahlein@ifz-goettingen.de
[1]USDA-ARS Southeastern Fruit and Tree Nut Research Laboratory, Byron, GA 31008, USA
[2]Embrapa Agricultural Informatics, Campinas, SP 13083-886, Brazil
[5]Institute of Sugar Beet Research, Holtenser Landstrasse 77, 37079 Göttingen, Germany
Full list of author information is available at the end of the article

variable that is of most importance or interest in a particular experimental situation (Paul et al. 2005). Quantification of disease severity caused by biotic agents is the focus of this article.

Visual estimation is the action of assigning a value to severity of symptoms perceived by the human eye. A sensor or instrument directly or indirectly measures the amount of disease or stress signal based on remote sensing (Nilsson 1995; Bock et al. 2010a). Thus, an image can be captured in the visible spectrum (VIS) and processed using image analysis (Bock et al. 2010a; Bock and Nutter Jr 2011; Barbedo 2013, 2016a). The amount of disease can also be measured by image capture in the non-VIS spectral range, including by hyperspectral and multispectral imaging (HSI and MSI), and chlorophyll fluorescence or other methods. The latter methods are conceptually different to that estimate or measurement of disease severity based on visible symptoms or the visible spectrum alone (Mahlein et al. 2012a; Mutka and Bart 2015; Simko et al. 2017; Kuska and Mahlein 2018; Mahlein et al. 2018). Visual estimates are based only on the perception of wavelengths of the electromagnetic spectrum in the VIS range (380 to 750 nm), while HSI and MSI systems use wavelengths in the range 250 to 2500 nm (Fig. 1). In general, only part of this range is chosen (usually the near-infrared (NIR) and infrared (IR) bands) - no single system covers the entire range. Raters perceive and learn to discriminate symptomatic from asymptomatic tissue in order to estimate percent diseased tissue. VIS spectrum image analysis bases

measurement on the number of pixels that conform to pre-defined properties of pixels representing a diseased state vs. healthy state, which are identified using a range of statistical procedures. HSI and MSI systems measure signature wavelengths associated with the diseased state. Image acquisition and analysis has additional challenges but also advantages over visual estimates (Mahlein 2016). Similar to visual ratings, image-based systems, depending on the objective, should: (i) detect disease or other stress as early as possible, (ii) differentiate among biotic diseases, (iii) differentiate biotic from abiotic stresses, and (iv) quantify disease severity accurately.

Information on disease severity is needed at various spatial scales from the microscopic to plant organs, whole plants, plots, fields or regions, so scalability is an important criterion to take into account when choosing an assessment method. Furthermore, assessment of severity is needed to complement genomics-scale data and provide timely, appropriate and correct measurements to fulfil the needs of 'phenomics' in plant breeding (Mutka and Bart 2015; Simko et al. 2017). High throughput is an important consideration in the era of phenomics, affecting progress and resource use efficiency.

Optical sensors perform non-invasively and have been developed and used to support disease detection, classification and severity measurement. Precision agriculture and plant phenotyping for resistance breeding already benefit from these technologies (Fiorani and Schurr 2013; Kruse et al. 2014; Stewart et al. 2016; Mahlein et al. 2018). Although other sensor-based methods of
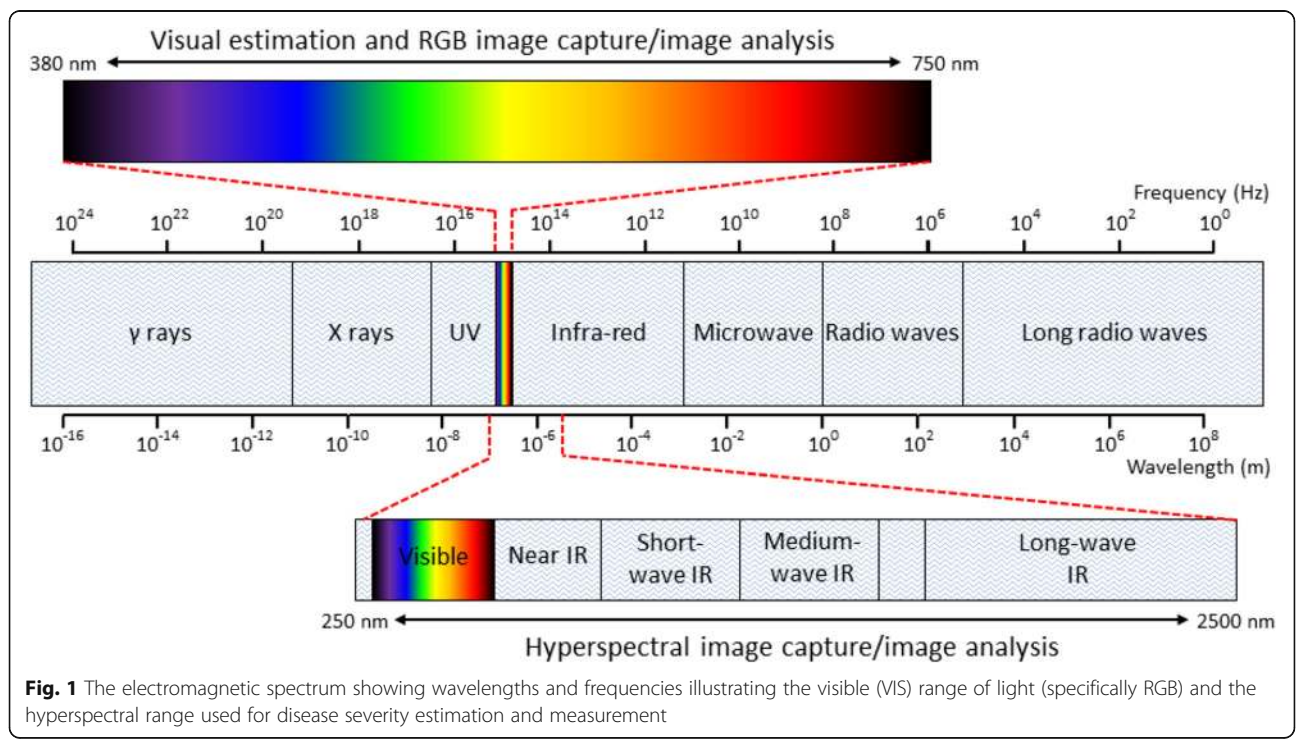

Fig. 1 The electromagnetic spectrum showing wavelengths and frequencies illustrating the visible (VIS) range of light (specifically RGB) and the hyperspectral range used for disease severity estimation and measurement

disease or pathogen quantification exist (thermal imaging, chlorophyll fluorescence and molecular or serological approaches), the reader is recommended to seek out recent publications on these topics elsewhere (Oerke and Steiner 2010; Sankaran et al. 2010; Mutka and Bart 2015; Mahlein 2016). This review will focus primarily on the status and use of visual estimation, VIS spectrum and HSI image analysis as methods to quantify disease severity, paying particular attention to recent developments and challenges to improve accuracy and reliability of the estimates and measurements.

Terms, concepts and the importance of accurate plant disease severity quantification

An accurate estimate or measurement is one that is close to the actual or true value, or 'gold standard' (Nutter Jr et al. 1991; Madden et al. 2007; Bock et al. 2010a; Bock et al. 2016a). In remote sensing, the actual or true values are referred to as 'ground truth' data. Biased estimates or measurements are those that deviate from actual accuracy. Two biases exist: systematic bias (over or underestimation which is related to the magnitude of the actual value) and constant bias (overall tendency to over or underestimate). Precision is the variability of estimates, but in disease severity estimation or measurement accuracy, precision must accommodate closeness to the true value (Madden et al. 2007). By definition, consistently accurate estimates must be reliable (Bock et al. 2016a), where reliability is the tendency for repeated estimates or measurements of the same specimen(s) to be close to one another (Nutter Jr et al. 1991; Madden et al. 2007). Reliability can be described as inter-rater (or method, e.g. various imaging methods) reliability or intra-rater (or method) reliability. Reliability may be less of an issue when measuring disease under controlled conditions using devices like VIS image analysis or HSI compared to estimates by different visual raters, or measurements under field conditions.

Accurate measurements or estimates of severity are important: it ensures that treatment effects are correctly analyzed, yield loss relationships understood, surveys are meaningful, and germplasm phenotypes rated appropriately. Furthermore, severity data might be used as a decision threshold or for disease forecasting purposes and thus the need to spray (or not). Inaccuracy can hamper the research process, waste resources, and could impact grower profitability. The required level of accuracy may vary among situations. Several empirical and simulation-based studies have demonstrated that disease assessment can result in a type II error (a false negative) (Christ 1991; Newton and Hackett 1994; Parker et al. 1995a; Bock et al. 2010b; Chiang et al. 2014; Chiang et al. 2017a, 2017b). A type I error (a false positive) could be as damaging, although this has not been found in disease assessment studies. Accurate estimates or measurements will minimize these two errors.

Visual estimation of disease severity

The evolving status of visual estimates has been punctuated by various reviews and book chapters (Anon 1947; Chester 1950; Large 1966; James 1974; Horsfall and Cowling 1978, Chapter 6; Kranz 1988, Chapter 3; Campbell and Madden 1990, Chapter 6; Chaube and Singh 1991, Chapter 9; Nilsson 1995; Cooke 2006, Chapter 2; Madden et al. 2007, Chapter 2; Bock et al. 2010a). Since 2010 there have been only two reviews, one relating to the issue of accuracy (Bock et al. 2016a), and the other providing a summary of the development and validation of standard area diagrams (SADs, Del Ponte et al. 2017).

Methods of visual estimation and nature of the data

Visual estimates of disease severity are based on various kinds of scales typical of measurement science (Stevens 1946; Baird and Norma 1978). Of the four main scale types, only interval scales are not represented in plant disease severity estimation as they lack a true zero (it is not possible to estimate less than zero disease). Disease severity has been assessed using nominal, ordinal and ratio scales. Their perceived utility, advantages and disadvantages are as follows:

Nominal scales These qualitative (descriptive) scales have been defined and described (Newell and Tysdal 1945; Campbell and Madden 1990; Madden et al. 2007; Bock et al. 2010a; Bock et al. 2016a). Nominal scales are based on brief descriptions such as "no disease", "mild disease", "moderate disease" and "severe disease", or symbols "- "(healthy), "+", "++" and "+++" (various levels of severity). Nominal scales are subjective, and may vary by rater and assessment time. The data may be analyzed using statistical methods based on rank or frequencies.

Ordinal scales (quantitative and qualitative) There remains a lack of clarity on what in this review is termed a 'quantitative ordinal scale', which has a set number of classes describing numeric intervals between 0 and 100%. These have been termed interval scales (Nutter Jr and Esker 2006; Bock et al. 2009a), ordinal scales (Hartung and Piepho 2007), category scales (Chiang et al. 2014) and quantitative ordinal scales (Bock et al. 2016a) in the literature. The American Phytopathology Society in its instruction to authors considers them an ordinal scale (Anon 2020). Qualitative ordinal scales have a clear and significant order of values, but the numeric magnitude of the differences between each class is unknown (for example, the Likert scale, Likert 1932). Quantitative ordinal scales have a clear and significant order of

values, and the magnitude of each ordered number is numerically bounded by a specified range.

Qualitative ordinal scales are valuable for comparing severity of some diseases that do not have easily quantified symptom. Many virus, other systemic diseases and root diseases may fall into this category, for example, cassava mosaic disease (Hahn et al. 1980) and huanglongbing of citrus (Gottwald et al. 2007). These rank data are based on discrete descriptions of symptom types and progression that is almost certainly not linear. It is not statistically appropriate to take means or use mid-points of these scales (Stevens 1946), as the mid-point and mean have little biological relation and violate assumptions of parametric tests. An index based on class frequencies can be calculated for qualitative ordinal scales, which may then be analyzed using parametric statistics, or they can be analyzed using non-parametric statistics suitable for various experiment designs and distribution functions (Shah and Madden 2004; Fu et al. 2012).

Quantitative ordinal scales may have equal or unequal intervals (Horsfall and Heuberger 1942; Horsfall and Barratt 1945; Hunter and Roberts 1978). The Horsfall-Barratt scale (HB, Horsfall and Barratt 1945) has been widely used (Table 1; Haynes et al. 2002; Miyasaka et al. 2012; Jones and Stansly 2014; Rioux et al. 2017; Kutcher et al. 2018; Strayer-Scherer et al. 2018). The US Forestry Service uses it to assess ozone injury (https://www.nrs.fs.fed.us/fia/topics/ozone/methods/). However, it is based on the nonexistent Weber-Fechner law (Nutter Jr and Esker 2006), and the ability of raters to estimate in the broad categories in the middle of the scale is better compared to what the scale indicated (Forbes and Korva 1994; Nutter Jr and Esker 2006; Bock et al. 2009b). Inappropriate scale structure is illustrated in results of

studies in plant breeding (Xie et al. 2012). An improved quantitative ordinal scale has been developed that provides a lower risk of type II error, which is recommended where an ordinal scale is required (Chiang et al. 2014) (Table 2). Analysis of quantitative ordinal scales may be through mid-point conversion (mid-point of the percent interval, not mid-point of the scale itself) and subsequent parametric analysis, or as described above for qualitative ordinal scales, or using a proportional odds model (Chiang et al. 2019).

The frequency of ordinal scores may be used to obtain a disease severity index (DSI) (Chester 1950). Disease severity is estimated on the specimens by a rater using the scale and is used to determine the DSI (%) = [sum (class frequency × score of rating class)] / [(total number of plants) × (maximal disease index)] × 100 (Chester 1950; Hunter and Roberts 1978; Chaube and Singh 1991; Kora et al. 2005; Vieira et al. 2012). Although a relationship may exist between true severity and a severity index, they are intrinsically different and should not be used interchangeably. Recent studies by Chiang et al. (2017a, 2017b) indicate that the DSI can be particularly prone to overestimation when using the above formula if the mid-point values of the rating class are not considered.

Ratio scales Many diseases lend themselves to severity estimation by ratio scales. The percentage scale is a widely applied scale to visually estimate severity (recent examples include Gent et al. 2018; Bock and Chiang 2019; Hamada et al. 2019; Xu et al. 2019). The

Table 1 The Horsfall and Barratt (H-B) quantitative ordinal scale showing the disease severity ranges, midpoints and interval sizes (Horsfall and Barratt 1945)

H-B category	Disease severity range	Midpoint	Interval size
1	0	0	0
2	0$^+$-3	1.5	3
3	3$^+$-6	4.5	3
4	6$^+$-12	9.0	6
5	12$^+$-25	18.5	13
6	25$^+$-50	37.5	25
7	50$^+$-75	62.5	25
8	75$^+$-88	81.5	13
9	88$^+$-94	91.0	6
10	94$^+$-97	95.5	3
11	97$^+$-100	98.5	3
12	100	100	0

Table 2 An improved 16-class quantitative ordinal scale for general assessment of plant disease severity based on the scale developed by Chiang et al. (2014)

Ordinal equivalent	Midpoint	Severity (% range)
0	–	–
1	0.05	0$^+$ to 0.1
2	0.30	0.1$^+$ to 0.5
3	0.75	0.5$^+$ to 1.0
4	1.50	1.0$^+$ to 2.0
5	3.50	2.0$^+$ to 5.0
6	7.50	5.0$^+$ to 10.0
7	15.0	10.0$^+$ to 20.0
8	25.0	20.0$^+$ to 30.0
9	35.0	30.0$^+$ to 40.0
10	45.0	40.0$^+$ to 50.0
11	55.0	50.0$^+$ to 60.0
12	65.0	60.0$^+$ to 70.0
13	75.0	70.0$^+$ to 80.0
14	85.0	80.0$^+$ to 90.0
15	95.0	90.0$^+$ to 100.0

percentage scale ranges from zero to 100% and a rater gauges the proportion of the organ showing symptoms and estimates the severity accordingly. The percentage scale data is amenable to analysis by parametric statistics and means and standard deviations are appropriate measures.

Very few studies have addressed resource use efficiency in visual disease assessment – how to minimize the risk of a type II error while optimizing use of specimen numbers and assessment method (Chiang et al. 2016b). The results of that study indicated that choice of assessment method, optimizing specimen numbers and number of replicate estimates while using a balanced experimental design are important criteria to consider for maximizing the power of hypothesis tests.

Sources of error

Rater variation The earliest study to clearly demonstrate rater variability was that of Nutter Jr et al. (1993), although Sherwood et al. (1983) demonstrated rater effects in their study comparing rater estimates of disease caused by *Stagonospora arenaria* on leaves of *Dactylis glomerata*. Bock et al. (2009a) described rater variability for 28 different raters assessing symptoms of citrus canker on leaves of grapefruit. Some individuals are innately accurate, yet others are inaccurate. Individual raters tend to over or under-estimate and this may extend over the whole scale, or the rater may have variable tendencies over the range of the percentage scale (Hau et al. 1989; Nita et al. 2003; Godoy et al. 2006; Bock et al. 2009a; Bardsley and Ngugi 2013; Yadav et al. 2013; Schwanck and Del Ponte 2014). Where rater bias is concerned, type II error can be exacerbated using quantitative ordinal scales (Chiang et al. 2016a).

Some rater-related characteristics may be associated with cognitive type, gender or other psychological traits, but this is yet to be explored in severity estimation. Inter-rater variability may be problematic although no studies have investigated the impact of different raters in an experiment. Minimizing the number of raters on a specific experiment will help remove potential variability from the data; or deploying raters by block or replicate will help minimize effects of individual raters.

Responses to disease characteristics A common tendency is to overestimate at low disease severities, which is particularly sensitive to the number of lesions and lesion size – the more lesions there are, the greater the tendency to overestimate (Sherwood et al. 1983; Forbes and Jeger 1987; Bock et al. 2008b).
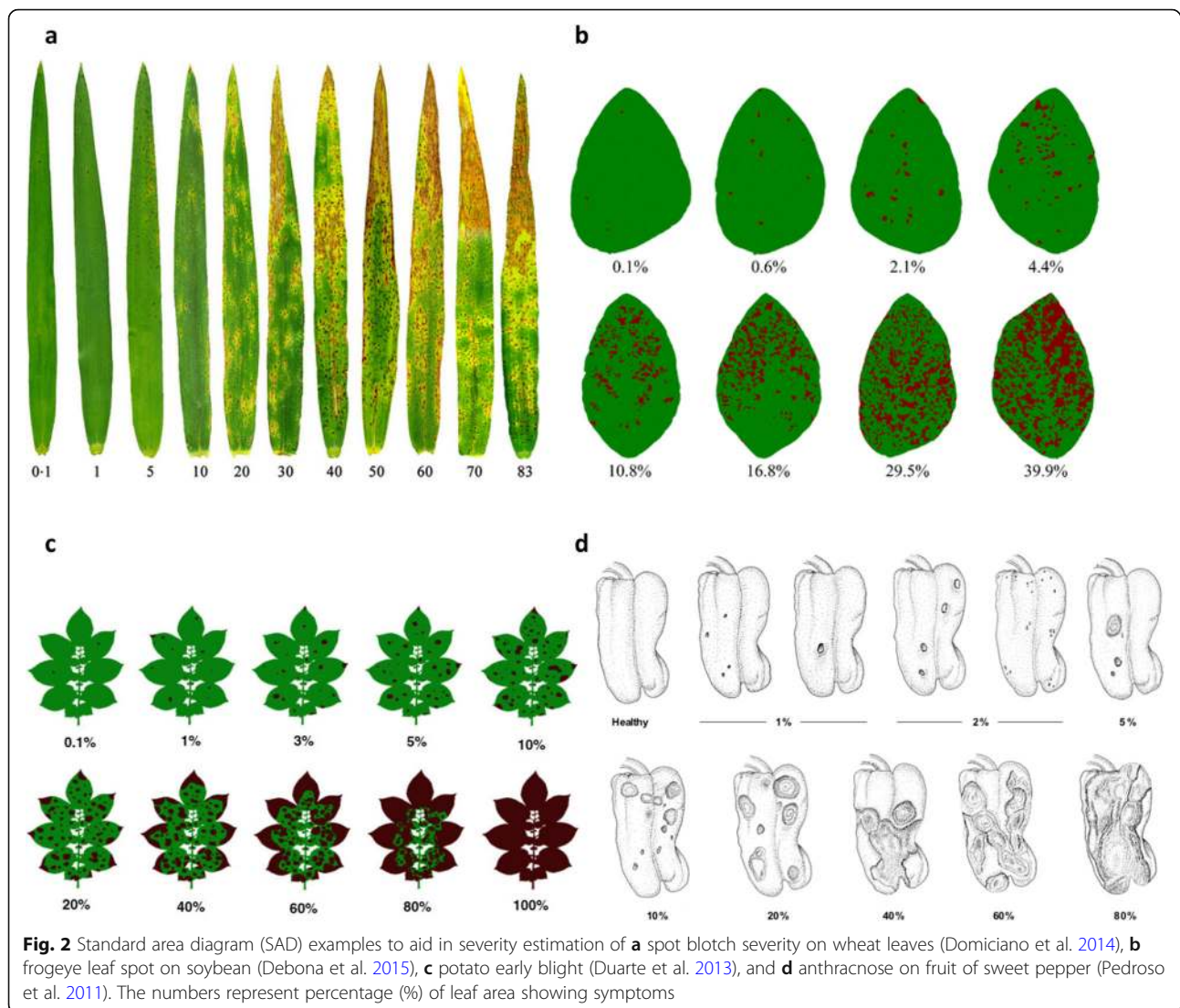
Preferred rating values or "knots" Raters have shown a consistent preference for certain severities at intervals of 5% and particularly 10% at severities > 10%. Thus, raters prefer 10, 15, 20, 25%...95 and 100% (Koch and Hau 1980; Bock et al. 2008b; Schwanck and Del Ponte 2014), which can lead to error.

Host organ characteristics Forbes and Jeger (1987) found that visual assessments of severity on simulated root structures were overestimated. Other organ types were not notably different in terms of accuracy (stems, leaves (various types), panicles, pods, tubers heads and roots). But few studies that have investigated the effect of organ type. Studies on the development and validation of SADs may be useful in this regard, but most diagrams have been developed for foliar diseases (Del Ponte et al. 2017).

Other factors Rating environment: does a rater perform more accurately under certain conditions? What is the effect of noise, heat, exhaustion or time allotted for an assessment? Fast assessments are not necessarily less precise (Parker et al. 1995a). Color blindness may impact disease severity estimation of some pathosystems (Nilsson 1995).

Methods to improve accuracy of estimates

Standard area diagrams (SADs) SADs are a simple and widely used tool to improve accuracy of rater estimates (Fig. 2). The diagrams developed by Cobb (1892) are the oldest assessment aid. James (1971) subsequently developed SADs for several crops. During the last 25 years, research on SAD development and validation has intensified, further demonstrating the value of SADs for improving accuracy (Del Ponte et al. 2017). Gains using SADs are variable among raters and across pathosystems (Spolti et al. 2011; Yadav et al. 2013; Schwanck and Del Ponte 2014), and are generally greatest for those raters who are least accurate (Yadav et al. 2013; Braido et al. 2014; González-Domínguez et al. 2014; Debona et al. 2015; Duan et al. 2015). Increase (Δ) in agreement (based on Lin's concordance correlation, ρ_c) may range from $\Delta > 0.4$ for inexperienced raters, to $\Delta \sim 0$, or possibly a slight loss in agreement for innately accurate raters. Overall, the use of SADs helps standardize raters, improving inter-rater reliability (itself a result of the accuracy of estimates of severity on individual specimens). Agreement (ρ_c) on the 0 to 100% range with actual values from image analysis frequently > 0.90 when using SADs (Spolti et al. 2011; Duarte et al. 2013; Domiciano et al. 2014; González-Domínguez et al. 2014). This can be considered excellent agreement in measurement science (Altman 1991), although others are more conservative (McBride 2005). When SADs are not used,

Fig. 2 Standard area diagram (SAD) examples to aid in severity estimation of **a** spot blotch severity on wheat leaves (Domiciano et al. 2014), **b** frogeye leaf spot on soybean (Debona et al. 2015), **c** potato early blight (Duarte et al. 2013), and **d** anthracnose on fruit of sweet pepper (Pedroso et al. 2011). The numbers represent percentage (%) of leaf area showing symptoms

agreement is often < 0.85. There may be symptomatic patterns where unaided estimates can be quite accurate and so SADs are less useful (Del Ponte et al. unpublished).

A recent, comprehensive review of SADs quantitatively summarizes their characteristics and provides guidelines for additional research (see Table 3 in Del Ponte et al. 2017). Several questions remain to be addressed. Does diagram number in a SADs affect accuracy of the estimates (Bock et al. 2016b)? Recently, an electronic version of interactive SADs was developed for portable devices. The app, called 'Estimate' displays an ordinal quantitative scale (severity intervals in either linear or log increments) accompanied by a SAD representing the mid-point. The severity value is not entered directly as in typical use of a SAD. The rater first selects a main category (specific % interval) and, alternatively, a subcategory in 1 % units (Pethybridge and Nelson 2015).

Table 3 Best practices for maximizing accuracy of visual estimates of severity of plant disease

1. Training in the process of assessing and estimating plant disease severity.

2. Detailed instruction or understanding of the specific symptoms and what constitutes healthy and diseased tissue.

3. Use of SADs

   • Ensure the SAD set is developed and validated appropriately (Del Ponte et al. 2017)

   • Instruction on how to use SADs (generally to interpolate to the nearest percentage point, occasionally as an ordinal scale).

4. Emphasize the risks of estimates falling on preferred, characterized values. Endeavor to estimate to the nearest percent, which is not necessarily at a 5% or 10% step value on the 100% point scale.

5. Appropriate sample sizes – usually 30 (Chiang et al. 2016b), but less or more depending on the errors of the estimates which can vary with actual severity and disease distribution.

Recently, Del Ponte et al. (2019) discerned shortcomings of some scale options in the Estimate app. The study showed the superiority of linear over the log-incremental scale, but only for the two-stage (category and subcategory) assessment process. The delivery of SADs in portable devices may increase in the future, as sophistication improves usability.

Training Nutter Jr and Schultz (1995) demonstrated that computer-based training improved accuracy, but this may be short-lived (Parker et al. 1995b). In a few cases training may reduce accuracy – possibly due to training on pathosystems not related to the one being used in practice (Bardsley and Ngugi 2013). Nutter Jr and Schultz (1995) found that one rater's coefficient of determination (R^2), indicative of precision, changed from 0.825 to 0.933 before and after training. Training software programs were developed for older computer operating systems, for example DISTRAIN (Tomerlin and Howell 1988) and Severity. Pro (Nutter Jr and Litwiller 1998). Neither new nor updated versions of these training programs based on computer-generated images exist; they may have been replaced by training raters with true-color photos of symptoms combined with the use of SADs technology.

Instruction Instruction provides an opportunity for the raters to recognize symptoms and estimate severity accurately. Bardsley and Ngugi (2013) found good instruction of symptoms of bacterial spot on peach and nectarine resulted in the greatest improvement in inter-rater reliability (which could also be tangentially related to improvements in accuracy in that study) by inexperienced raters compared to training. The coefficient of determination (R^2) increased from 0.76 to 0.96 after instruction (and to 0.88 after training).

Experience, general field-based training and other methods Experience in recognizing disease symptoms does have an impact on ability to estimate accurately. Although individual, inexperienced raters may be innately more accurate than some experienced raters, as a group, experienced raters tend to be more accurate (Yadav et al. 2013; González-Domínguez et al. 2014). Grids comprised of squares that overlay a leaf (or other specimen area) were shown to improve accuracy (Parker et al. 1995b) but have never been widely implemented.

Considering these tools available to improve accuracy and reliability (and acknowledging that many questions remain), standardized procedures may be outlined that will provide a basis to maximize accuracy of individual specimen estimates when performing visual assessments (Table 3).

Application in research and practice

Visual assessments are most often applied at the scale of individual organs (leaflets, leaves, fruit, flowers etc.), plants, and occasionally fields. However, these data are used at regional and global levels. Visually estimating severity at the field scale is somewhat archaic. For example, a key was developed during the 1950s to assess late blight of potato in the UK at the field scale (Moore 1943). Such field keys, although a valid method of disease severity assessment, are not considered further as they have been rarely used in recent times.

Visual severity assessment has been applied to compare treatments (for example, fungicide or cultural control methods), assess the effect of disease on yield, for surveys, assess the severity of disease on different genotypes etc.

Summary of how accuracy has been improved for visual estimates

Based on current research, where possible, the percentage scale is demonstrably the most accurate tool on which to base visual estimates of disease severity (Nita et al. 2003; Hartung and Piepho 2007; Bock et al. 2010b; Chiang et al. 2014). Thus, accuracy of disease severity estimation has been improved through a better understanding of error, methods to reduce bias, particularly with the use of SADs, but also through instruction and training.

Visual estimation (with use of the approaches outlined in Table 3) has probably come close to maximizing accuracy of estimates. Appropriate scales, SADs, training and instruction, if correctly implemented can provide remarkably accurate estimates that will minimize the risk of any type II errors.

Measurement of disease severity using visible spectrum image analysis

Assessment based on VIS spectrum image analysis have the potential to be accurate, repeatable and reproducible (Martin and Rybicki 1998; Bock et al. 2008a; Barbedo 2014; Clément et al. 2015). Lindow and Webb (1983) were among the earliest pioneers of digital image analysis of plant disease. Particularly since 2000, more sophisticated algorithms and statistical approaches have advanced the capability of differentiating symptomatic from healthy tissue in digital images (Table 4) (Bock and Nutter Jr 2011; Barbedo 2013, 2016a, 2017, 2019).

Methods of image acquisition

Various cameras or image capturing devices record in the VIS spectrum. Red-green-blue (RGB) sensors are portable and widely available. With the advent of hand-held devices with cameras the possibilities of easily obtaining numerous images is increased many-fold

Table 4 The crop, stress, and analysis technique used to describe severity measurement using visible spectrum (RGB) image analysis with symptom segmentation. The superscript numbers cross-reference the "Reference" with the "Analysis software/technique" and "Symptom measured" for each study. For example, in the first row 'Color Transformations' and 'filtering' were used only by Camargo and Smith, and 'Scion image' only by Wijekoon. Both measured 'Area affected'

Crop[a]	Analysis software/technique	Symptom measured	References
Alfalfa	Color transformations[8], filtering[8], Scion Image[38]	Area affected[8,38]	Camargo and Smith (2009)[8], Wijekoon et al. (2008)[38]
Apple	CNN[37]	Severity[37]	Wang et al. (2017)[37]
Banana	Color transformations[8], filtering[8]	Area affected[8,16]	Camargo and Smith (2009)[8]
Bean	Mathematical morphology[3,4,5], color transformations[5], color analysis[9], thresholding[31], Scion Image[38]	Area affected[4,5,9,31,38], Deformation[9]	Barbedo (2016a, 2016b)[4], Barbedo (2017)[5], Contreras-Medina et al. (2012)[9], Škaloudová et al. (2006)[31], Wijekoon et al. (2008)[38]
Cassava	Mathematical morphology[4,5], color transformations[5], ImageJ[22], support vector classifier[25], k-nearest neighbors[25], extra trees[25]	Area affected[4,5,22], Severity levels[25]	Barbedo (2016a, 2016b)[4], Barbedo (2017)[5], Mutka et al. (2016)[22], Mwebaze and Owomugisha (2016)[25]
Citrus	Mathematical morphology[4,5], color transformations[5], Assess 1.0[6,7]	Area affected[4,5,6,7], Number of lesions[6]	Barbedo (2016a, 2016b)[4], Barbedo (2017)[5], Bock et al. (2008a, 2008b)[6], Bock et al. (2009a, 2009b)[7]
Coconut	Mathematical morphology[4,5], color transformations[5]	Area affected[4,5]	Barbedo (2016a, 2016b)[4], Barbedo (2017)[5]
Coffee	Color transformations[3,5,29], mathematical morphology[3,4,5], thresholding[29], CNN[41]	Area affected[3,4,5,29], Severity levels[41]	Barbedo (2014)[3], Barbedo (2016a, 2016b)[4], Barbedo (2017)[5], Price et al. (1993)[29], Esgario et al. (2019)[41]
Cotton	Mathematical morphology[4,5], color transformations[5]	Area affected[4,5]	Barbedo (2016a, 2016b)[4], Barbedo (2017)[5]
Cucumber	Color comparison[2], self-organizing map + linear perceptron[13,17], Photoshop 6.0 + Matrox Inspector 2.2[18], Superpixel clustering + expectation maximization[39]	Area affected[2,13,17,18,39]	Bakr (2005)[2], Goclawski et al. (2012)[13], Kuźniak et al. (2014)[17], Kwack et al. (2005)[18], Zhang et al. (2019)[39]
Grapevine	Mathematical morphology[4,5], color transformations[5], ImageJ[27]	Area affected[4,5,27]	Barbedo (2016a, 2016b)[4], Barbedo (2017)[5], Peressotti et al. (2011)[27]
Maize	Assess 2.0[1], mathematical morphology[4, 5], color transformations[5,8], filtering[8], thresholding[20], Scion Image[38]	Area affected[1,4,5,20,38], Pustule count[1]	Bade and Carmona (2011)[1], Barbedo (2016a, 2016b)[4], Barbedo (2017)[5], Camargo and Smith (2009)[8], Martin and Rybicki (1998)[20], Wijekoon et al. (2008)[38]
Oat	Color transformations[19], thresholding[19,36]	Area affected[19,36]	Macedo-Cruz et al. (2011)[19], Tucker and Chakraborty (1997)[36]
Passion fruit	Color transformations[3,5], mathematical morphology[3,4,5]	Area affected[3,4,5]	Barbedo (2014)[3], Barbedo (2016a, 2016b)[4], Barbedo (2017)[5]
Pumpkin	Color analysis[9,17]	Area affected[9,17], Deformation[9]	Contreras-Medina et al. (2012)[9], Kuźniak et al. (2014)[17]
Rice	Mathematical morphology[4], fractal dimensions + fuzzy C-means[40]	Area affected[4], Severity[40]	Barbedo (2016a, 2016b)[4], Zhou et al. (2013)[40]
Soybean	Color transformations[3,5,30], mathematical morphology[3,4,5,30], DCNN[12], Linear Discriminant Analysis + Support Vector Machine[23]	Area affected[3,4,5,30], Severity[12,23], number of lesions[30]	Barbedo (2014)[3], Barbedo (2016a, 2016b)[4], Barbedo (2017)[5], Ghosal et al. (2018)[12], Naik et al. (2017)[23], Shrivastava et al. (2015)[30]
Sugar beet	Color transformations[3], mathematical morphology[3], Assess 2.0[10]	Area affected[3,10]	Barbedo (2014)[3], De Coninck et al. (2012)[10]
Sugarcane	Mathematical morphology[4,5], color transformations[5,26], thresholding[26]	Area affected[4,5,26]	Barbedo (2016a, 2016b)[4], Barbedo (2017)[5], Patil and Bodhe (2011)[26]

Table 4 The crop, stress, and analysis technique used to describe severity measurement using visible spectrum (RGB) image analysis with symptom segmentation. The superscript numbers cross-reference the "Reference" with the "Analysis software/technique" and "Symptom measured" for each study. For example, in the first row 'Color Transformations' and 'filtering' were used only by Camargo and Smith, and 'Scion image' only by Wijekoon. Both measured 'Area affected' *(Continued)*

Crop[a]	Analysis software/technique	Symptom measured	References
Tomato	Color transformations[3], mathematical morphology[3], Self-organizing maps + Bayesian classifier[14], thresholding[18,28], Assess 2.24[35]	Area affected[3,14,18,28,35]	Barbedo (2014)[3], Hernández-Rabadán et al. (2014)[14], Lindow and Webb (1983)[18], Pethybridge and Nelson (2015)[28], Sun et al. (2014)[35]
Wheat	Mathematical morphology[4,5], color transformations[5], Assess 2.0[11], Chan-Vese model + PCA[15], Assess 1.0[21,32], ImageJ[33,34], Scion Image[38]	Area affected[4,5,11,15,21,32,38] Number of lesions[33,34]	Barbedo (2016a, 2016b)[4], Barbedo (2017)[5], El Jarroudi et al. (2015)[11], Hu et al. (2017)[15], Mirik et al. (2006)[21], Steddom et al. (2005)[32], Stewart and McDonald (2014)[33], Stewart et al. (2016)[34], Wijekoon et al. (2008)[38]

[a]There are a some crops with only one associated reference: *Arabidopsis thaliana* (Laflamme et al. 2016), avocado (Kerguelen and Hoddle 1999), barley (Kokko et al. 2000), black pepper (Barbedo 2016a, 2016b), bracken fern (Lindow and Webb 1983), California buckeye (Lindow and Webb 1983), cashew (Barbedo 2016a, 2016b), clover (Wijekoon et al. 2008), collards (Pethybridge and Nelson 2015), dahlia (Bakr 2005), goldenrod (Goodwin and Hsiang 2010), green bean (Bakr 2005), kale (Barbedo 2016a, 2016b), lilac (Pethybridge and Nelson 2015), lily-of-the-valley (Goodwin and Hsiang 2010), lima bean (Pethybridge and Nelson 2015), mallow (Pethybridge and Nelson 2015), melon (Barbedo 2016a, 2016b), palm tree (Barbedo 2016a, 2016b), papaya (Barbedo 2016a, 2016b), peanut (Barbedo 2014), pepper (Contreras-Medina et al. 2012), Phalaenopsis seedling (Huang 2007), phlox (Goodwin and Hsiang 2010), plane tree (Clément et al. 2015), potato (Wijekoon et al. 2008), rose (Bakr 2005), squash (Bakr 2005), subterranean clover (Kruse et al. 2014), sunflower (Tucker and Chakraborty 1997), sweet cherry (Olmstead et al. 2001), sycamore (Lindow and Webb 1983), turfgrass (Horvath and Vargas 2005), vigna (Bakr 2005), watermelon (Pethybridge and Nelson 2015), yellow starthistle (Berner and Paxson 2003)

(Pethybridge and Nelson 2015). Analog video cameras (Lindow and Webb 1983; Hetzroni et al. 1994; Martin and Rybicki 1998), digital videos sensors (Lloret et al. 2011; Clément et al. 2015) and flatbed scanners (Olmstead et al. 2001; O'Neal et al. 2002; Berner and Paxson 2003; Kwack et al. 2005; Škaloudová et al. 2006) have also been used.

Methods of image analysis and processing

Segmentation Segmentation (delineation of the area of interest) is a step in many image analysis algorithms (Fig. 3). In testing image analysis, leaf segmentation is generally performed manually, but for practical application segmentation must be automated. The only difference between segmentation and severity measurement is that the latter includes an additional step relating the areas occupied by diseased and healthy tissues. With the rise of artificial intelligence (AI, machine learning, and


Fig. 3 Segmentation steps during image analysis. An image of **a** a pecan leaflet with symptoms of scab (caused by *Venturia effusa*), **b** the same image with the whole leaf segmented from the background, and **c** the leaf with only the diseased areas segmented out. Diseased area on this leaflet is 30.14%

its off-shoot, deep learning) segmentation is less of a requirement.

Software for image analysis Many studies have employed third-party software to measure severity including Assess (Horvath and Vargas 2005; Steddom et al. 2005; Mirik et al. 2006; Bock et al. 2008a, 2008b; Bock et al. 2009a, 2009b, 2009c; De Coninck et al. 2012; Sun et al. 2014; El Jarroudi et al. 2015), launched in 2002 (Lamari 2002). Assess requires the user to predefine segmentation parameters for automation, but this works only if all images were captured under the same conditions (Bock et al. 2009c). Other software include Sigma Pro (Kerguelen and Hoddle 1999; Olmstead et al. 2001; Berner and Paxson 2003), ImageJ (O'Neal et al. 2002; Abramoff et al. 2004; Peressotti et al. 2011; Stewart and McDonald 2014; Laflamme et al. 2016), Adobe Photoshop (Kwack et al. 2005; Cui et al. 2010) and Scion Image Software (Wijekoon et al. 2008; Goodwin and Hsiang 2010). In the review on SADs, 20 programs were reported to obtain actual severity measurements, but Assess and Quant (Vale et al. 2003) were the most commonly used (Del Ponte et al. 2017)

Validation Validation involves comparing the image analyzed measurement to an actual or "gold-standard". The actual value may be based on a visual estimate (Steddom et al. 2005; De Coninck et al. 2012; El Jarroudi et al. 2015) or manually delineated image analysis data (Martin and Rybicki 1998; Bock et al. 2009a; Peressotti et al. 2011). Regression has been widely used to compare accuracy of image analysis systems (Horvath and Vargas 2005; Steddom et al. 2005; Peressotti et al. 2011; El Jarroudi et al. 2015), although other statistical criteria are often used to provide more meaningful insights (Bock et al. 2009a; De Coninck et al. 2012; Stewart and McDonald 2014). Because experimental setups and contexts vary between studies, the results are not always comparable (Horvath and Vargas 2005); reported variabilities based on regressions (R^2) and correlations (r) fall within the 0.70–1.00 range (Martin and Rybicki 1998; Steddom et al. 2005; Peressotti et al. 2011; De Coninck et al. 2012).

Custom systems using color transformations and artificial intelligence Newer methods for severity measurement can be divided in two categories. The first relies on color transformations; the second on AI using machine or deep learning techniques.

i) Color transformation increases the contrast between healthy and diseased areas in images (Hu et al. 2017), often coupled with mathematical morphology operations (Macedo-Cruz et al. 2011;
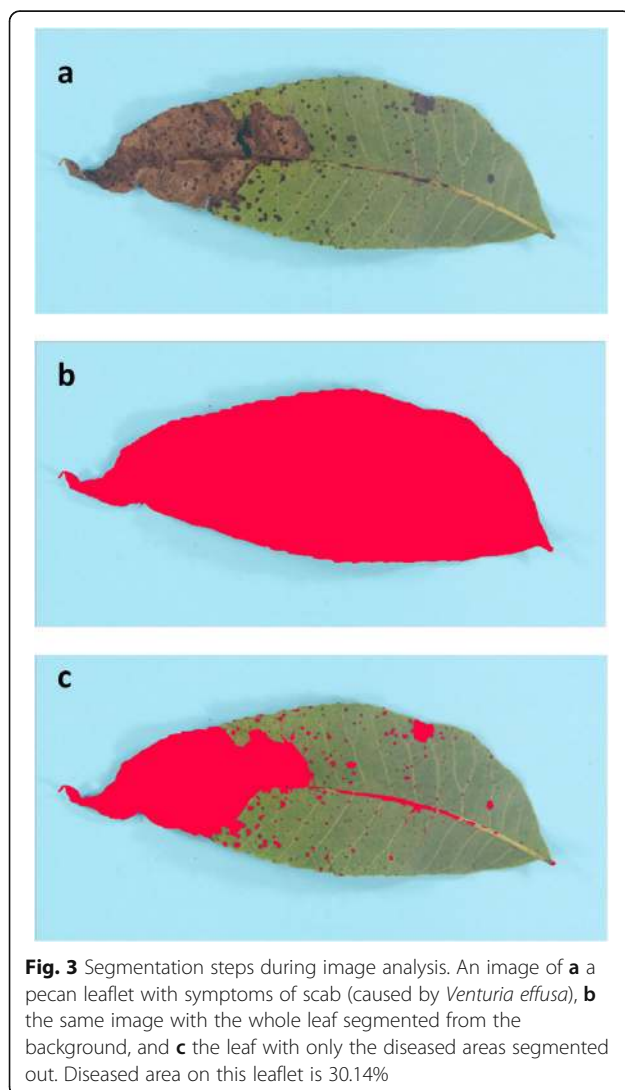
Contreras-Medina et al. 2012; Barbedo 2014; Shrivastava et al. 2015; Barbedo 2016a, 2017), thresholding (Price et al. 1993; Patil and Bodhe 2011; Clément et al. 2015) and filtering (Camargo and Smith 2009), with the objective of isolating the regions of interest. These algorithms are generally quick to develop and simple to implement but may not be suitable for dealing with subtle symptoms.

ii) Many applications of AI for image analysis are based on machine learning, which may be supervised or unsupervised. Supervised learning typically involves methods of classification (including logistic regression, support vector machines and artificial neural networks), while unsupervised learning relies on methods of clustering (including k-means clustering and principal component analysis) that rely on structural patterns in the data. For disease severity measurement the classifiers require the severity to be transformed from continuous data to a discrete scale of values. This is usually accomplished by either labelling each pixel as healthy or diseased, or by defining severity levels based on a nominal or ordinal scale, for example as "low", "medium" and "high". A variety of methods have been tested and reported in the literature, including K-means clustering (Kruse et al. 2014), Fuzzy C-means (Zhou et al. 2013), k-nearest neighbors (Mwebaze and Owomugisha 2016; Naik et al. 2017), linear discriminant analysis (Kruse et al. 2014; Naik et al. 2017), expectation maximization (Zhang et al. 2019) and support vector machines (Mwebaze and Owomugisha 2016; Naik et al. 2017), among others.

Neural networks employing deep learning architectures have become predominant in image-based classification systems (Barbedo 2019). Deep learning efficiently extracts complex features from images without the need for segmentation and is being applied to severity measurement, usually in the form of deep Convolutional Neural Network (CNN) architectures (Fig. 4). Many

images are needed for deep learning systems (Wiesner-Hanks et al. 2018; Ramcharan et al. 2019). The largest database for a disease was reported in 2018 (Wiesner-Hanks et al. 2018); there were 8222 images of corn leaves annotated with 105,705 lesions of northern leaf blight, although all were from a single field in New York. The set was used for detection only. Two other important databases containing images of plant diseases are available: the PlantVillage (Hughes and Salathé 2015), which contains > 50,000 curated images of many crop diseases; and Digipathos (Barbedo et al. 2018, available at https://www.digipathos-rep.cnptia.embrapa.br), also containing > 50,000 images of crop diseases. However, neither has image annotation for sample source location or actual severity. Image libraries are a progress-limiting gap. Data sharing is one solution: globally, plant pathologists working on various pathosystems could capture images to represent the diversity of characteristics and enable image analysis systems (Barbedo 2019).

Many trained deep learning models are lightweight enough for mobile applications, so they can be run directly on the device without the need for connectivity (Ramcharan et al. 2019), important in remote areas.

Accuracy of image analysis

The number of studies employing CNN has increased in the last few years. Ramcharan et al. (2019) used CNN and 2415 leaf samples to automatically detect two severity classes of cassava mosaic disease. Accuracy of low severity detection was 29.4%. Esgario et al. (2019) found that assigning severity of multiple diseases of coffee using deep learning was up to 84.13% accurate. Wang et al. (2017) found accuracy of severity of apple leaf black rot measurements ranged from 83.3 to 100%, depending on class (there were 4 classes of severity). Thus, estimates of accuracy are often being considered at a lower resolution compared to visual estimation using the 0 to 100% scale. Scale type, number of intervals and replication may differ
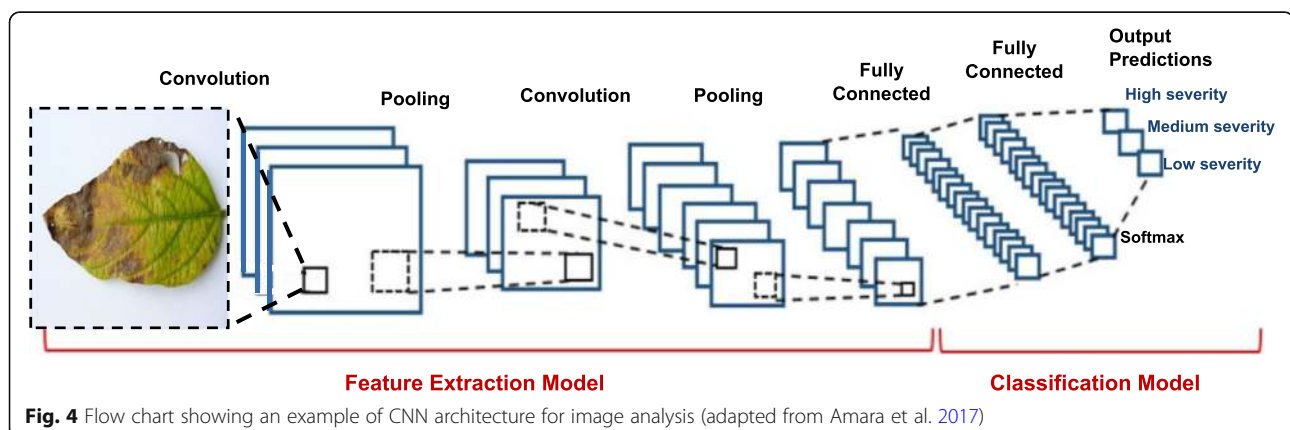

Fig. 4 Flow chart showing an example of CNN architecture for image analysis (adapted from Amara et al. 2017)

considerably to achieve the same power in a hypothesis test (Bock et al. 2010b; Chiang et al. 2014, 2016a, 2016b, 2019).

Much of the variation in image analysis may be attributed to two factors. Firstly, conditions under which the images were captured and the variety of symptoms in the images. Studies using VIS spectrum images captured in the field often report lower accuracies. Examples of images captured under variable conditions include the systems proposed by Macedo-Cruz et al. (2011), Barbedo (2017) and Hu et al. (2017) (resulting in 92, 91, and 84% accuracy, respectively); images captured under controlled conditions include methods proposed by Patil and Bodhe (2011), Kruse et al. (2014) and Stewart et al. (2016) (resulting in 98, 95, and 94% accuracy, respectively). Secondly, the actual reference values to which the estimates are compared will affect accuracy. Where the reference is a visual estimate, subjectivity will be directly related to the perceptions of the rater (Bock et al. 2008a).

Sources of error affecting accuracy

Operator Operators must accurately pair the diagnosis guidelines with the symptoms. Even manual measurements using image analysis have some subjectivity. Actual values based on image analysis used to validate automatic methods (or other methods of assessment) are variable (Barbedo 2013; Bock et al. 2008a). But the error should be small.

Variation in symptoms, host and background To work effectively, deep learning models must be trained using images covering a wide range of conditions. For most other techniques segmentation of leaf and disease is required (Barbedo 2016a). Threshold values and other parameters derived under one set of conditions generally fail under a different set of conditions due to variation in brightness, contrast, reflections, weather conditions and numerous other factors (Barbedo 2014). Symptoms may vary depending on stage of development (Patil and Bodhe 2011) and the interaction with environmental factors (Mutka et al. 2016). Separating image components automatically with field-acquired images is a challenging and complex task and solutions are only recently being developed (Zhang et al. 2018a). Automatic segmentation can be easier if a screen is placed behind the leaf prior to image capture (El Jarroudi et al. 2015; Pethybridge and Nelson 2015; Shrivastava et al. 2015), but this makes image capture more time-consuming and problematic. Thus, most methods using field-captured images rely on the user to manually segment the leaf (Barbedo 2014, 2016b, 2017).

Issues with image acquisition and differentiating diseased vs. healthy areas There is subjectivity in determining the edges of some symptoms (Barbedo 2014; Stewart et al. 2016). Leaves are not always flat causing perspective problems (Barbedo 2014), or require flattening (Clément et al. 2015). Small symptoms may be confused with debris (Barbedo 2014). Shadows, leaf veins, and other parts of the plant may mimic symptoms, causing error (Olmstead et al. 2001; Bade and Carmona 2011; Barbedo 2014; Clément et al. 2015; Barbedo 2016a). Groups of lesions may merge, impairing a counting process (Bock et al. 2008a; Bade and Carmona 2011). The presence of other disorders may exacerbate delineation of the symptoms of interest (Bock et al. 2008a, 2009a; El Jarroudi et al. 2015; Barbedo 2016b). Specular reflections may render parts of the leaf featureless (Steddom et al. 2005; Peressotti et al. 2011; Barbedo 2016a). Image compression may introduce distortions and artifacts (Steddom et al. 2005; Bock et al. 2010a). Symptom complexity affects the difficulty of the task (Bock et al. 2008a; Barbedo 2017), which has led some authors to argue that different algorithms are needed for each symptom (Contreras-Medina et al. 2012), or each host-pathogen pair (Mutka and Bart 2015). AI techniques can address some of these issues if trained with sufficiently comprehensive data. Factors that cause loss of information (specular reflections, shadows, etc.) can only be addressed by appropriate protocols during image capture.

Automatic image capture in the field can result in underlying leaves being obscured. Perspectives will be variable. This is an issue for plants with dense canopies if severity measurement on lower leaves is needed (Wiesner-Hanks et al. 2018).

Actual values Evaluation of measurements obtained using VIS image analysis is not straightforward. Generally, the "gold standard" reference is generated manually by image analysis (Peressotti et al. 2011; El Jarroudi et al. 2015), by expert visual estimation, or rarely other methods (Martin and Rybicki 1998). Due to subjectivity, even manually delineated image analysis may harbor operator error, and thus the systems developed are dependent on the references they are tasked to mimic; they could vary if other "gold standard" references were used.

System limitations As effective as various new techniques are, including deep learning, sometimes images in the visible range do not carry enough information for distinction of severity classes. In such cases, combining different imaging methods may be a viable solution (Berdugo et al. 2014), perhaps with the sacrifice of higher costs and reduced mobility.
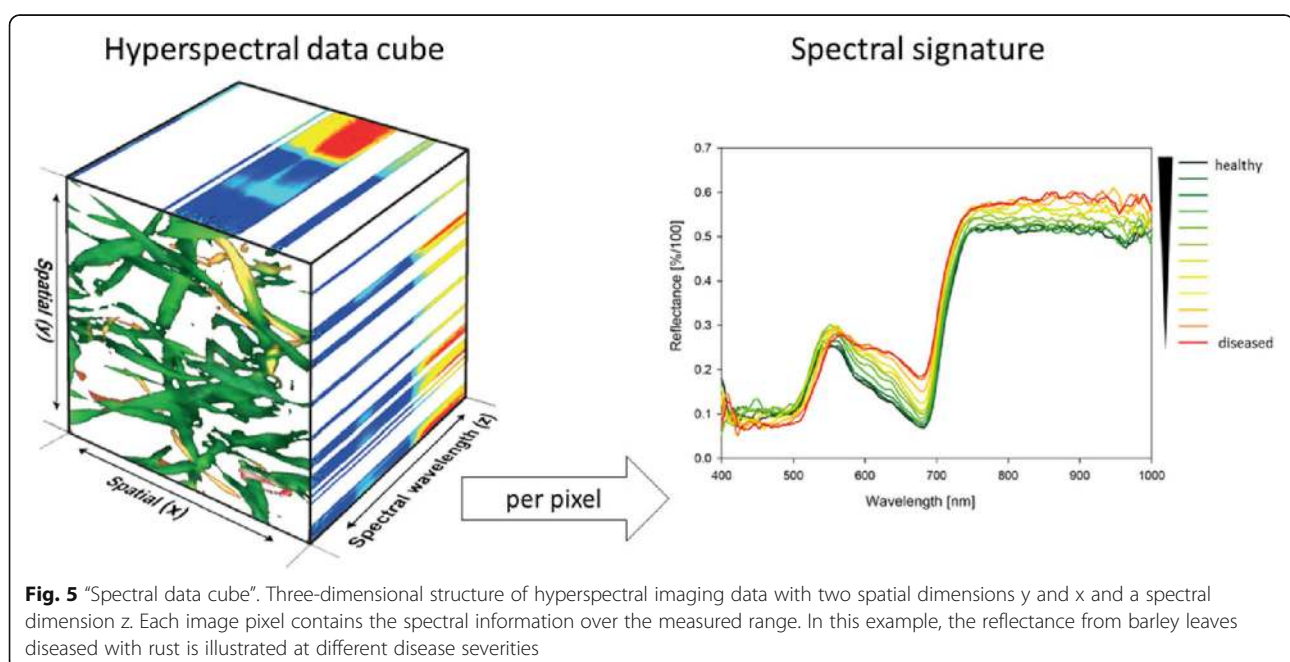
Application in research and practice

Scales of application Although VIS image analysis can be applied at different scales, the majority of the studies are at the scale of individual plant organs (Barbedo 2014; Kruse et al. 2014; Clément et al. 2015; El Jarroudi et al. 2015; Pethybridge and Nelson 2015; Barbedo 2016a, 2016b, 2017; Esgario et al. 2019; Ghosal et al. 2018; Ramcharan et al. 2019; Zhang et al. 2019) or the crop canopy (Macedo-Cruz et al. 2011; Laflamme et al. 2016; Naik et al. 2017). Image analysis of microscopic samples requires a sophisticated lab-based system (Ihlow et al. 2008).

Uses of image analysis Applications of RGB image-based severity measurement include: crop breeding and phenotyping, in which the objective is to rapidly measure severity on numerous specimens (Peressotti et al. 2011; De Coninck et al. 2012; Stewart and McDonald 2014; Laflamme et al. 2016; Naik et al. 2017; Ghosal et al. 2018; Karisto et al. 2018); the effect of disease on yield (Macedo-Cruz et al. 2011); to compare various treatments (Clément et al. 2015); in precision agriculture, in which the objective is to pinpoint areas where symptoms are more severe for a more focused control of the disease (Kruse et al. 2014), including aspects of biocontrol (Berner and Paxson 2003); and for general crop management, in which the objective is to provide information to aid decision making (Zhou et al. 2013; Barbedo 2014; Pethybridge and Nelson 2015; Barbedo 2016a, 2016b, 2017; Hu et al. 2017).

Image analysis software for disease severity measurement is available for mobile devices (Pethybridge and Nelson 2015; Manso et al. 2019). Mobile device-based applications generally require the user to set thresholds, which can lead to inconsistencies (Bock et al. 2008a, 2009c). Software was recently developed automating severity estimation using Fuzzy Logic rules and image segmentation for the mobile application 'Leaf Doctor' (Sibiya and Sumbwanyambe 2019).

Image capture using mobile platforms (UAVs, ground robots etc) is being studied in the field, although disease detection is the primary focus (Johnson et al. 2003; Garcia-Ruiz et al. 2013; de Castro et al. 2015). Measurement of severity with VIS spectrum image analysis using mobile platforms is less common (Lelong et al. 2008; Sugiura et al. 2016; Duarte-Carvajalino et al. 2018; Franceschini et al. 2019; Ganthaler et al. 2018; Liu et al. 2018), but is an area of research need. An automated VIS image analysis system on a UAV for measuring severity had moderate precision compared to visual rating ($R^2 = 0.73$), but was deemed acceptable for rating potato resistance to late blight (Sugiura et al. 2016). Zhang et al. (2018b) found RGB images taken using a UAV were less effective ($R^2 \leq 0.554$) in differentiating severity of sheath blight of rice compared to HSI sensors ($R^2 \leq 0.627$). VIS image analysis to measure disease severity is not yet routinely used outside the research realm. There are a few examples of controlled environment, high-throughput systems used routinely for research purposes. Karisto et al. (2018) described automated VIS image analysis to measure severity of Septoria leaf blotch on wheat. There was a good relationship between image

Fig. 5 "Spectral data cube". Three-dimensional structure of hyperspectral imaging data with two spatial dimensions y and x and a spectral dimension z. Each image pixel contains the spectral information over the measured range. In this example, the reflectance from barley leaves diseased with rust is illustrated at different disease severities

analyzed measurements and visual estimates (Lin's concordance correlation, ρ_c = 0.76 to 0.99, depending on rater (Stewart and McDonald 2014)). Microscopic imaging of powdery mildew on barley for genotype screening was considered ready for high-throughput processing (Ihlow et al. 2008). But both still require time-consuming sample preparation.

Spectral sensor technology to measure plant disease severity

MSI and HSI sensors measure the light reflected by an object. In plant disease detection and severity measurement this might be a single plant organ (leaf, fruit, and/or storage root), a plant, or a crop stand. Several studies have demonstrated that diseases can be detected accurately even before symptoms are visible to the human eye (Rumpf et al. 2010; Zhao et al. 2017). Indeed, detecting the quantity of disease at very early stages is valuable for disease management decisions, and neither raters nor VIS image analysis can detect latent disease. Furthermore, HSI is non-invasive and non-destructive, and is an objective method, and if automated can significantly reduce the workload compared to other methods of assessment (Walter et al. 2015; Mahlein 2016; Virlet et al. 2017).

Characteristics of light reflectance from plants

The optical properties of plants are determined mainly by their reflectance, transmission and absorbance of light. Diseases affect these signature characteristics.

Reflectance of light from plants Reflectance depends on leaf properties. Transmission and absorbance are influenced by pigments and water (Gates et al. 1965; Curran 1989). Reflectance is caused by biochemical properties that result in a mixed signal (Gates et al. 1965; Carter and Knapp 2001; Gay et al. 2008). The visible range (400–700 nm) is characterized by absorption by chlorophyll, carotenoids and anthocyanins (Gay et al. 2008). According to Hindle (2008), NIR and SWIR stimulate molecular motion that induces absorption or reflection by compounds having characteristic spectral patterns. The NIR reflectance of leaves is determined mainly by the leaf and cell structures and the canopy architecture (Gates et al. 1965; Elvidge 1990). The NIR and SWIR regions have bands that are absorbed by water (particularly the SWIR region) (Seelig et al. 2008).

How do plant diseases influence the optical properties of plants? The pathogen causes changes in physiological and biochemical processes in the host (Mahlein et al. 2010), resulting in disease, often accompanied by symptoms. The pathogen and symptom types have consequences for the detectability and measurement of disease severity. Each host-parasite interaction has a specific spatial and temporal dynamic, impacting different wavebands during pathogenesis (Wahabzada et al. 2015; Wahabzada et al. 2016). Sensors offer the potential to extract new features of disease severity and dynamics, and a new way to visualize and analyze severity. Progress in disease symptoms can be directly related to HSI measurements (as "metro maps" or "disease traces", Kuska et al. 2015; Wahabzada et al. 2015, 2016). Metro maps of plant disease dynamics explicitly track the host-pathogen interaction, providing an abstract yet interpretable view of disease progress.

Methods of hyperspectral image acquisition

In contrast to RGB cameras having a spatial resolution of several megapixels, spectral sensors include high-resolution techniques with greater spectral resolution (Fig. 5; Mahlein et al. 2018). HSI and MSI sensors assess narrow wavebands in specific ranges of the electromagnetic spectrum in combination with a high spatial resolution. The VIS and NIR region (400–1000 nm) have the highest information content for monitoring plant stress. The ultraviolet-range (UV, 250–400 nm) (Brugger et al. 2019) and SWIR-range (1000–2500 nm) (Wahabzada et al. 2015) provide information as well. Spectral sensors can be characterized by resolution (number of wavebands per nm) and the type of the detector. Often, MSI sensors cover the RGB range in addition to NIR but provide less data due to lower spectral resolution, although they are lightweight and cost less (Mahlein et al. 2018). In contrast, HSI sensors are more complex, heavier, expensive and the measurement takes longer, demanding strict protocols. Systems consist of the sensor, a light source and a control unit for measuring, storing and processing the data (Thomas et al. 2018b).

Choice of HSI sensor in combination with the measuring design and platform is the basis of a data set. Accuracy and resolution are influenced by the distance between the sensor and the object. Thus, airborne or space borne systems have lower spatial resolution compared to near-range systems. Data preprocessing and analysis is closely linked and individually designed depending on the sensor, setup and purpose of measuring (Behmann et al. 2015a; Mishra et al. 2018).

Non-imaging sensors Non-imaging HSI sensors do not provide spatial information. The focal length of the viewing angle and the distance to the target determine the size of the measured area. The signal comprises mixed information from healthy and diseased areas, affecting the sensitivity and specificity, so early detection and measurement of symptoms by non-imaging sensors is limited, especially at low disease severities. Measurement of severity of mixed infections is challenging using

non-imaging sensors. Mahlein et al. (2010, 2012b) found the detection limit using non-imaging HSI for Cercospora leaf spot (CLS) and powdery mildew of sugar beet was 10 and 20% diseased leaf area, respectively.

Imaging sensors Imaging HSI sensors collect extra information on shape, gradient or color of the spatial dimension (Behmann et al. 2015a). There are push-broom and whisk-broom scanners that capture the spectral information of a pixel point or a pixel line at the same time, respectively. The image emerges due to movement of the sensor and has high spatial and spectral resolution. Depending on image size, image acquisition time may take minutes, limiting imaging sensors to motionless objects (Thomas et al. 2017).

Other HSI sensors Filter-based HSI sensors do not require the sensor to move and are generally faster than push- and whisk-broom sensors, but the subject must be motionless. HSI snapshot cameras capture images akin to RGB cameras, but have lower resolution compared to push- or whisk-broom sensors, although they have a fast image acquisition time (Thomas et al. 2017).

Choice of sensor platform
It is critical to consider purpose and subject. HSI sensor setups can be handheld or mounted on a platform (vehicles, robots, UAVs, airplanes or satellites). Choosing the right sensor in combination with the right measurement scale is the key requirement for successful field measurement. Possible targets could be early disease detection/identification, or quantifying disease incidence or severity. Drone measurements at a height of 50 m above the crop in combination with a low spatial resolution hyperspectral camera will not detect single leaf lesions compared to a measuring device close to the leaf canopy that has high spatial resolution. Pixel-wise attribution of diseased and healthy tissue is conducive to observe spectral reflectance patterns of diseases in detail. It should be noted that some disease symptoms can only be distinguished from other diseases and stresses when using HSI imaging with high spatial resolution.

Data handling, training and analysis
There are several approaches for analyzing HSI and MSI data – but no standard one. Data preprocessing typically consists of normalization to a white reference standard and dark current images (Behmann et al. 2015a). A smoothing of the data can be performed. Often the background and parts of the image which are not required for further analysis are masked to reduce the data complexity.

Vegetation indices A common and straightforward way to analyze hyperspectral images are vegetation indices (VI) (Devadas et al. 2009; Ashourloo et al. 2014; Behmann et al. 2015a). VIs are algorithms based on band ratios. Often 2–6 bands are involved. VIs are used to highlight a specific factor while reducing data complexity and the impact of other factors (Jackson and Huete 1991; Gitelson et al. 2014; Blackburn 2007). Several well-described VIs have been used for the detection or quantification of diseases, but weren't specifically developed for that purpose. Moreover, VIs are related to pigment content, vitality, biomass, water content and so on. For the analysis of MSI data, VIs are often the method of choice.

Some disease specific VIs have been developed (Mahlein et al. 2013; Ashourloo et al. 2014; Oerke et al. 2016). The correlation between disease severity and reflectance wavebands are calculated and those wavebands with the highest correlations are integrated into disease specific indices. Comparative studies have demonstrated that disease specific VIs are superior to standard VIs (Mahlein et al. 2013; Ashourloo et al. 2014). An overview of VIs for the detection and/or quantification of diseases is presented, including disease specific VIs (Table 5).

Symptom recognition and analysis As for VIS image analysis, hyperspectral image analysis is challenging. The aim is to extract a small proportion of relevant information from the hyperspectral signal (Behmann et al. 2015b). Algorithms are developed to learn and make predictions about the data (Kersting et al. 2016) and can cope with hundreds of wave bands used for detection, quantification and characterization of plant diseases in the laboratory, greenhouse and field (Behmann et al. 2015b; Singh et al. 2016). Either the entire spectral data set can be analyzed, and patterns identified, or feature selection methods can be applied to reduce the data complexity. As with VIS image analysis methods, there are supervised and unsupervised learning approaches.

Supervised approaches like regression and classification demand annotated training data. Provision of training data is a limiting factor in severity measurement as sufficiently large image sets of annotated data for specific diseases under a full range of conditions are not available.

Compared to supervised approaches, unsupervised approaches are less well explored, but do not rely on annotation and training data. Unsupervised methods can be assigned to pattern recognition in hyperspectral image data. A 'crossover' is a data driven learning model that relies on the actual data set, and not on predefined models; the algorithm utilizes extreme data points to define archetypal signatures, including latent aspects of the data (Wahabzada et al. 2015, 2016).

Approaches using AI for measuring severity are based on deep learning. In contrast to the predefined features of machine learning approaches, deep

Table 5 Examples of different general spectral vegetation indices and disease-specific vegetation indices used to detect and measure severity of various plant disease

Type	Name	Parameter / Disease	Calculation[a]	Reference
General spectral vegetation indices	Normalized Difference Vegetation Index (NDVI)	Chlorophyll	$(R_{800}$-$R_{670})/(R_{800} + R_{670})$	Rouse et al. (1974) Tucker et al. (1991)
	Plant Senescence Reflectance Index (PSRI)	Pigment content, Senescence	$(R_{680}$-$R_{500})/R_{750}$	Merzlyak et al. (1999)
	Carotenoid Reflectance Index (CARI)	Carotenoids	$1/R_{510}$-$1/R_{550}$	Gitelson et al. (2002)
	Anthocyanin Reflectance Index (ARI)	Anthocyanins/Yellow rust (wheat)	$1/R_{510}$-$1/R_{700}$	Gitelson et al. (2001) Zheng et al. (2018)
	Photochemical Reflectance Index (PRI)	Photosynthesis, Carotenoids/Yellow rust (wheat)	$(R_{570}$-$R_{531})/(R_{570 + 531})$	Gamon et al. (1992) Zheng et al. (2018)
	Normalized Difference Spectral Index (NDSI)	Leaf spot diseases (peanut)	$(R_{800} - R_{670})/(R_{800} + R_{670})$	Rouse et al. (1974) Chen et al. (2019)
Disease-specific vegetation indices	Cercospora Leaf Spot Index (CLSI)	Cercospora leaf spot (sugar beet)	$(R_{698}$-$R_{570}/(R_{698} + R_{570})$-$R_{734}$	Mahlein et al. (2013)
	Sugar Beet Rust Index (SBRI)	Sugar beet rust (sugar beet)	$(R_{570} - R_{513})/(R_{570} + R_{513}) + (1/2 * R_{704})$	Mahlein et al. (2013)
	Powdery Mildew Index (PMI)	Powdery mildew (sugar beet)	$(R_{520} - R_{584})/(R_{520} + R_{584}) + R_{724}$	Mahlein et al. (2013)
	Leaf Rust Disease Severity Index 1 (LRDSI_1)	Brown rust (wheat)	$6.9*(R_{605}/R_{455})$-1.2	Ashourloo et al. (2014)
	Leaf Rust Disease Severity Index 2 (LRDSI_2)	Brown rust (wheat)	$4.2*(R_{695}/R_{455})$-3.8	Ashourloo et al. (2014)
	Red Edge Disease Stress Index (REDSI)	Yellow rust (wheat)	$(705$-$665)*(R_{783} - R_{665}) - (783$-$665)*(R_{705} - R_{665})2 \times R_{665}$	Zheng et al. (2018)
	Lemon Myrtle-Myrtle Rust Index (LMMRI)	Myrtle rust (myrtle)	$(R_{545}/R_{555})^{5/3}*R_{1505}/R_{2195}$	Heim et al. (2019)

[a]R_i; indicates reflectance at waveband i

learning models determine more abstract and more informative data representation within the process of optimization to a particular task. Deep learning offers potential to identify optimal features for the detection and measurement of a specific disease. As with RGB images, CNNs show great potential as a component of deep learning. Nagasubramanian et al. (2017, 2019) applied a 3D CNN for detection of charcoal rot on soybean using close-range VIS-NIR hyperspectral images and achieved a detection accuracy of 97% and was able to predict lesion length on most stems. However, these technologies demand substantial training data. Establishing a library of ground-truthed data for different diseases is crucial to the successful implementation of deep learning for disease quantification.

Related to general disease severity measurement, the importance of early detection (a "pre-visible symptom severity measurement") cannot be overstated and is critical in many circumstances; HSI can excel when severity is nascent.

Ground truthing, accuracy and measuring disease severity with spectral sensors

Various actual values or "ground truthing" have been used in HSI disease severity measurement including

visual estimates based on nominal or ordinal scales (Huang et al. 2007; Wang et al. 2016; Leucker et al. 2017), described stages of symptom progression (Kuska et al. 2015; Wahabzada et al. 2015, 2016; Zhu et al. 2017), and molecular quantification of the pathogen (Thomas et al. 2017; Zhao et al. 2017). An increasing number of studies have demonstrated that HSI and MSI data can be used to accurately detect, differentiate and quantify symptoms of plant diseases (Mahlein et al. 2012a). However, as noted, accuracy is not necessarily measured using the 0 to 100% scale as it has historically been for visual estimates or even for VIS image analysis. It may be related directly to the physiological, biochemical, structural and development changes in the host and pathogen. Comparing estimated or measured symptoms using the 0 to 100% scale to HSI, measurements can easily be done as HSI sensors provide pixel-based results on disease status (Fig. 6). The relation among visual rating and sensor measurement can be evaluated by post-classification routines and confusion matrixes.

Accuracy of detection can be robust. Apan et al. (2004) detected sugarcane orange rust with 96.9% accuracy compared to visually ground-truthed data; Bravo et al. (2003) used in-field spectral images for early detection of yellow rust infected wheat with 96% when
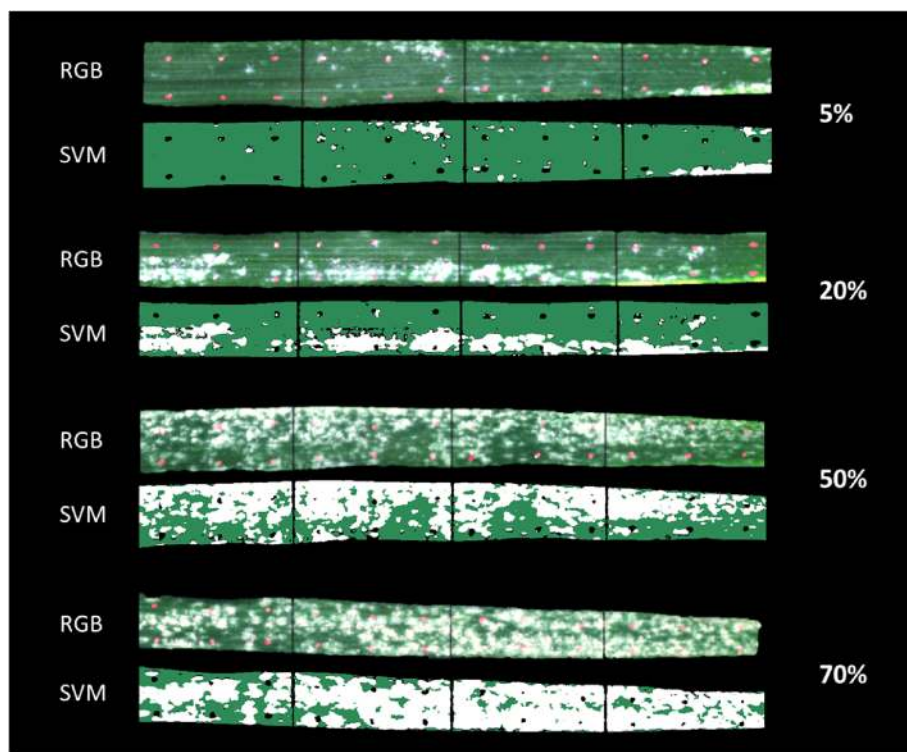
Fig. 6 RGB images and false-color classification of diseased pixels of wheat leaves with symptoms of powdery mildew caused by *Blumera graminis* f.sp. *tritici*. Hyperspectral images were acquired using a Specim V10 camera system, and classification was performed using Support Vector Machines (SVM). Percentage of diseased leaf area assessed by SVM classification is indicated on the right; classification accuracy ranged from 90% to 95%

compared to a visually-assessed disease map; Hillnhütter et al. (2011, 2012) discriminated symptoms caused by the nematode *Heterodera schachtii* and the soil borne fungus *Rhizoctonia solani* in sugar beet under both field and controlled conditions (spectral reflectance data and manual symptom assessment were correlated, $P < 0.01$); Delalieux et al. (2007, Delalieux et al. 2009a, 2009b) identified narrow waveband ratios with c-values (the c-index is derived from Received Operator Curves maximizing sensitivity for low values of the false-positive fraction) ranging from 0.80 to 0.88 for detecting scab (caused by *Venturia inaequalis*) on apple.

For measuring severity, Wahabzada et al. (2015, 2016) used advanced data mining techniques to define cardinal points during pathogenesis and differentiate spatial and temporal development of symptom dynamics of foliar diseases (caused by *Pyrenophora teres, Puccinia hordei* and *Blumeria graminis hordei*) of barley. Disease was quantified by counting the number of diseased pixels to equate to the stage of infection which has a relationship with severity (leaf area diseased), although severity (as a percent area diseased) was not explicitly performed. Some of these ideas are ushering in novel paradigms in the progress of disease severity for HSI. Huang et al. (2007) demonstrated reliable measurement of severity using a 9-class ordinal scale for severity of yellow rust in wheat ($R^2 = 0.91$). Other studies have explored classification accuracy using ordinal groupings in classes of visually assessed specimens as the assumed gold standard (Bravo et al. 2003; Alisaac et al. 2018; Thomas et al. 2018a; Alisaac et al. 2019), including the use of confusion matrices. Regression analysis of visual estimates of diseased wheat spikes on a percentage scale and hyperspectral measurements also had demonstrable reliability (R^2 up to 0.828, Kobayashi et al. 2016). Thomas et al. (2017), using pathogen DNA to ground-truth achieved a coefficient of determination (R^2) of 0.72 from 3 to 9 days after infection of barley with *Blumeria graminis* f.sp. *hordei*.

Sources of error affecting accuracy

Illumination Measurements in the field can be performed using shading and artificial light. If sunlight is used, robust checks against variation in sunlight intensity are critical (Wendel and Underwood 2017). Interpolation approaches may fail through lack of continuous illumination (Suomalainen et al. 2014). Solar altitude, clouds, dew or dust can be problematic. The application of suitable radiation transfer models may help reduce environmental effects (Jay et al. 2016) but is complex and time consuming. Appropriate calibration to reflectance standards or the continuous assessment of radiation intensity is necessary. Varying

illumination issues are more acute in direct sunlight and less severe under cloudy conditions, where the light is more diffuse. So far there are no standard calibration methods, the method of choice has to be designed depending on the senor-platform and illumination situation (Banerjee et al. 2020). For HSI under laboratory conditions, calibration routines are well established (Behmann et al. 2015a).

Motion Crop motion due to wind can be an issue. Most HSI sensors record information with a small temporal offset. With line scanning HSI cameras, the single lines are measured consecutively, and movement distorts the spatial image, whereas the spectral information remains valid (Thomas et al. 2017). Filter based systems often demand several seconds to record an image. If the object moves, the spectrum will consist of the reflectance information from different leaf areas and possibly even the ground, which cannot be corrected as the movement geometry is unknown. However, averaging the entire hyperspectral image mostly eliminates the effect, but spatial resolution is lost and the resulting data is comparable to that obtained using a simple spectrometer.

Mixed infection and mixed stress Quantification of a disease can be hindered by simultaneous stress (biotic or abiotic) or mixed infection. This aspect has only begun to be addressed. Studies are needed to demonstrate the potential of HSI to simultaneously identify and quantify multiple stressors or diseases.

Technical setup Leaves at different levels in a complex canopy require different exposure times. Shadows complicate saturation and since the choice of the exposure time is based on the brightest object, the exposure time is often much lower than required for shaded leaves low in the canopy, resulting in a noisier image.

Characteristics of the disease distribution Disease distributions may affect the ease with which the sensor can access specimens to sample. Some diseases spread from the lower leaves to the upper leaves through wind or the kinetic energy of rain droplets (e.g Septoria leaf blotch). Also, Septoria leaf blotch has a prolonged biotrophic phase. Thus, the upper leaves may not reflect the true disease severity in the crop stand when measurements are captured from above the canopy. Wind borne pathogens may be more likely to infect upper parts of a plant. In cereals, this favors the detection of foliar rust diseases or powdery mildews.

These challenges notwithstanding, HSI has great potential to provide a sophisticated, accurate and rapid method to measure disease severity at multiple spatial scales. The challenges are technically surmountable, and

the advances over the last several years demonstrate the utility of this technology.

Application in research and practice

Controlled conditions Many studies have measured disease severity using HSI under controlled conditions in the laboratory (Delalieux et al. 2009a, 2009b; Arens et al. 2016; Leucker et al. 2017). High spatial resolution can be obtained by hyperspectral microscopes (Kuska et al. 2015; Leucker et al. 2016), detecting plant-pathogen interactions at the submillimeter scale, before they are visible, or detectable using field-based HSI systems. Scale independent transfer of characteristic spectral signatures may be possible (Bohnenkamp et al. 2019), whereby spectral signatures of different diseases over time is used for detection and quantification models at different spatial scales. The approach will help process large numbers of complex host-pathogen interactions and the impact of mixed infections or abiotic stressors.

Field conditions HSI measurement of disease severity under field conditions is particularly challenging (Bravo et al. 2003; West et al. 2003). As with systems under controlled conditions, these are at an early experimental phase. Applied systems do not yet exist. Variable environmental conditions and biological heterogeneity impair the quality of field data. Additionally, the infection biology and epidemiology of a disease may impact detectability and measurability (West et al. 2003; Mahlein et al. 2019).

Contrasting the methods

An overview of the methods is presented in Fig. 7, and some of the advantages and disadvantages of the methods are contrasted (Table 6). Clearly, they have different levels of subjectivity, speed, scalability and cost. Accuracy also varies. Inexperienced, untrained/uninstructed and unaided raters can be wildly inaccurate in severity estimation. But trained, well-instructed and aided raters can provide very accurate estimates. Raters are slow, may be more expensive, and have low throughput. Scalability for visual rating is limited to plot or at most, field levels of assessment. However, both VIS and HSI/MSI image analysis offer less variable measurements of severity under tightly controlled conditions. Both can offer high throughput. Early detection and measurement of severity, particularly by HSI or MSI (and other remote sensors) is a major advantage and is being realized in the research arena. However, both HSI and MSI are limited in field situations as they are currently less capable of dealing with the wide variability in host, pathogen and disease characteristics experienced in the field. Raters, when well-trained and instructed can differentiate symptoms of diseases and suitable samples for assessment. Visual estimation of disease severity will be widely used
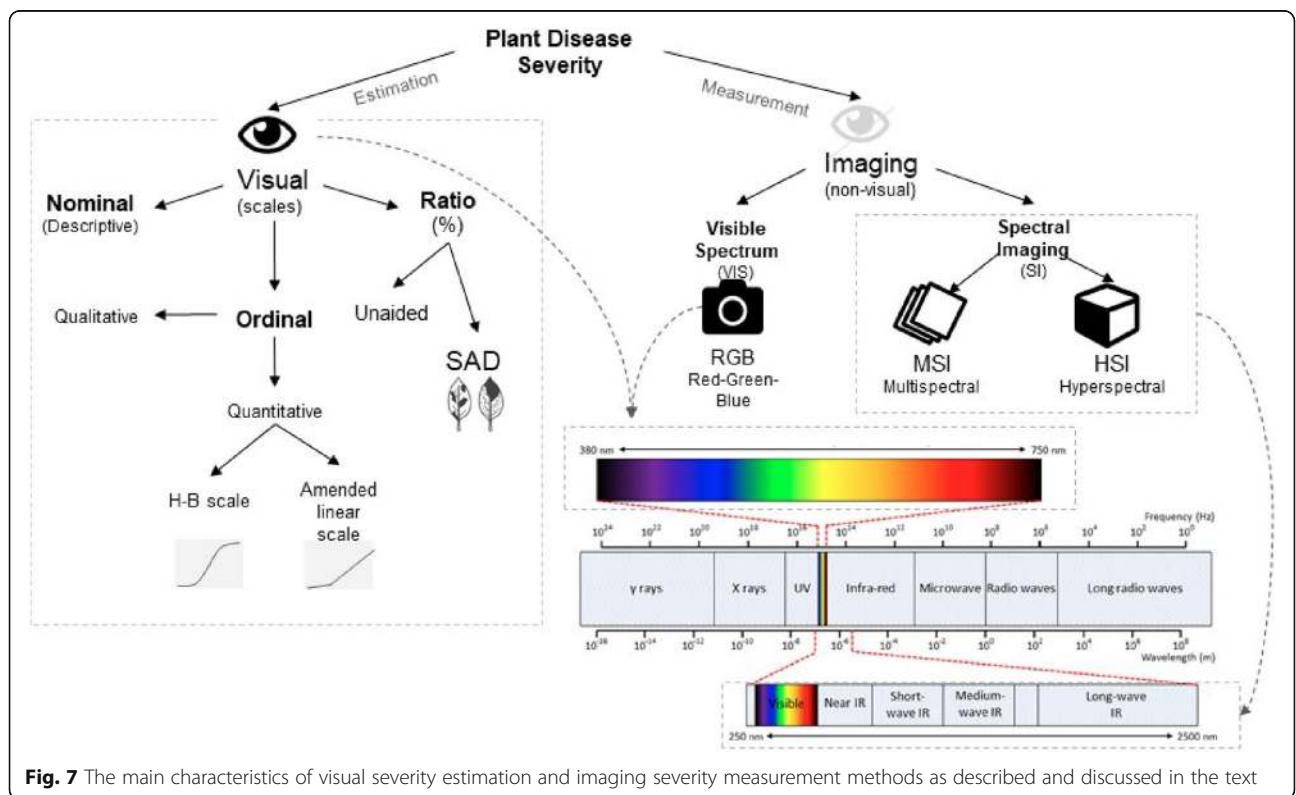

Fig. 7 The main characteristics of visual severity estimation and imaging severity measurement methods as described and discussed in the text

Table 6 A comparison of different criteria for visual assessment, visible spectrum image analysis (RGB) and hyperspectral image analysis as methods for obtaining plant disease severity data

Aspect	Severity assessment method		
	Visual	RGB	HSI
Training, re-training or adaptation	Yes	Yes	Yes
Disease-specificity	No	Yes	Yes
Subjectiveness	Yes	No	No
High throughput	No	Yes	Yes
Technical expertise	Medium	High	High
Cost of equipment (capital)	Low	Low	High
Cost of personnel (operational)	High	Low	Medium
Spatial coverage (scale)	Low	High	High
Canopy penetration	High	Low	Low
Operational speed	Low	Medium-High	Medium

for many years yet and may be needed alongside automated systems for validation and ground-truthing of new or improved fully automated AI-based methods for the foreseeable future.

Visual rating, when performed by trained, well-instructed and aided raters has probably reached its zenith of accuracy. But much is left to be understood regarding visual severity estimation, and the level of improvement will vary according to disease symptoms and how consistency within and among raters can be improved. In contrast, both VIS and HSI/MSI image analysis are rapidly evolving fields with ever more sophisticated approaches being developed and used for image acquisition and processing to measure severity. This is clear in the recent development of high-throughput systems for measuring disease under controlled conditions. Although measuring disease severity under field conditions remains challenging, the technical hurdles are being addressed and various systems have been demonstrated to have some utility, if not yet of practical value. It is possible that a combination of manual operations with automated measures will be required to overcome some limitations.

Visual rating of plant disease severity remains the most widely performed method for all purposes of field research where severity is a required variable. Very few mobile, or field operated VIS and HSI/MSI image analysis systems are routinely used in plant breeding, plant disease management, or for other purposes requiring severity measurement. This will doubtless change as research makes more advances facilitating the field application of VIS and HSI/MSI image analysis. As described, new tools based on AI have demonstrated capability and the potential to overcome many of the

barriers. Already some small companies and start-ups provide HSI services for crop monitoring. These may be a model for the future where plant disease assessment is a standard service using HSI and may be provided using various platforms. Furthermore, new digital technologies must be linked to existing prognosis and expert systems with integration into disease thresholding models for real-time management of disease. VIS and HSI/MSI image analysis will continue to play a more prominent role for quantifying disease in research and practice.

Most visual estimates are assessed for accuracy based on the percentage scale, which offers high resolution for differentiating severity of disease. VIS image analysis under tightly controlled conditions can accurately measure disease either when manually operated or automated based on the percentage scale. But under field conditions accuracy is less certain, and the measurements are most often compared to a limited number of classes on an ordinal scale (up to 9 classes), which results in lower resolution to differentiate severity compared to the percentage scale. However, sample sizes can be rapidly and easily increased with VIS image analysis, which can improve the power of a hypothesis test. Severity data collected by HSI/MSI sensors is sometimes related to the percentage scale, but often the data are related to an ordinal or nominal scale rating of the ground-truthed samples, or to characteristic stages during the pathogenesis process. This may provide a new paradigm for rating severity other than using a ratio, ordinal or nominal scales.

A major challenge for both VIS and HSI/MSI is training image sample sizes covering the range in variability of symptoms and conditions expected to be experienced. This will require considerable effort. A possible solution is citizen science (Barbedo 2019), in which non-professional volunteers collect and/or process data (Silvertown 2009). Practitioners and stakeholders could capture images in the field and an expert could annotate these. This idea has been implemented by Plantix™ (https://plantix.net/en/, PEAT, Berlin). This, and other studies referenced provide a sound basis for being optimistic for the technology in the future.

Furthermore, accuracies of different methods cannot be directly compared unless they are tested against identical gold standards or actual values. Thus, inferring the state of art quantitatively is challenging. It is worth noting that sharing the datasets used in published studies is being encouraged by many journals, so it might be possible to test new methodologies with the data used in prior experiments (Barbedo 2019), thus enabling more direct comparisons. Examples of accuracies attained by each of these methods are summarized by examples (Table 7). These and other studies have demonstrated that all three methods can provide accurate estimates or measurements of disease severity. However, VIS and HSI/MSI image analysis are still primarily at a research and developmental stage. Remote sensor-based methods are becoming less expensive, readily available and portable, and have the advantage of high throughput and scalability. However, the capability of raters in providing accurate estimates should not be overlooked as more sophisticated methods become available. Indeed, it behooves us to assure that the accuracy and reliability being attained by remote sensing methods is providing information at least sufficient for the purpose. Methods of validation should be in place to determine this – use of actual values or ground-truthing in all studies is critical to the ongoing process of ensuring accuracy.

Some needs for future research in visual disease assessment, RGB and HSI image analysis

This section is structured to pose specific questions and issues that need to be addressed through research. It does not intend to be exhaustive, but suggestive of some important avenues for future study.

Visual severity estimation When dealing with multiple raters, some individual or environment-related sources of errors that may affect accuracy remain unknown:

- Do raters' characteristics such as information processing speed (reflective or impulsive) affect accuracy?
- Does the environment (heat, cold etc.) affect accuracy of estimates?

We need to continue to optimize quantitative ordinal scales and SAD design to ensure that accuracy is maximized:

- Are there ordinal scales applicable for different pathosystems, regardless of severity range?
- How do we design SADs for diseases with different characteristics (lesion size, shape, colors, etc)?
- Do the number of diagrams in a SADs affect severity estimates?
- Is it possible to develop a few generic SADs to cover the range of leaf types and diseases that have to be assessed?
- Is one SAD representative of a percent sufficient as a reference diagram?
- How can instruction be performed to maximize accuracy?

The role of training in plant disease severity estimation is only partially explored:

Table 7 Comparison of different disease severity assessment methods in relation to accuracy-related measures, statistical methods and scale type and resolution used for method validation

Method	Actual value	Scale type and number of classes used for comparison	Resolution to differentiate severity	Statistic used to assess accuracy	Range	Reference
Visual assessment	Images traced from slide projections for contrast and image analyzed	Ratio scale 0 to 100% (100)	100	Regression, intercept and slope coefficient (a and b)	$R^2 = 0.913$ to 0.960; $a = 1.064$ to 12.958; $b = 1.029$ to 1.245	Martin and Rybicki (1998
	Images on paper cut by symptom border and weighed	Ratio scale 0 to 100% (100)	100	LCC (ρ_c)	$\rho_c = 0.51$–0.99	Nita et al. (2003)
	Manual image analysis	Ratio scale 0 to 100% (100)	100	Regression, intercept and slope coefficient (a and b)	$R^2 = 0.51$ to 0.93; $a = -1.90$ to 41.38; $b = 0.65$ to 1.24	Godoy et al. (2006)
	Manual image analysis	Ratio scale 0 to 100% (100)	100	Regression, intercept and slope coefficient (a and b) LCC (ρ_c)	Depending on symptom types: $R^2 = 0.59$ to 0.88; $a = -1.52$ to 2.83; $b = 0.10$ to 1.21; $\rho_c = 0.85$ to 0.94	Bock et al. (2008a)
	Manual image analysis	Ratio scale 0 to 100% (100)	100	Regression, intercept and slope coefficient (a and b)	$R^2 = 0.88$ to 0.98; $a = -6.68$ to 5.09; $b = 0.75$ to 0.94	Michereff et al. (2009)
	Manual image analysis	Ratio scale 0 to 100% (100)	100	LCC (ρ_c)	Means of $\rho_c = 0.76$ to 0.98 depending on fruit perspective and use of SADs	Spolti et al. (2011)
	Manual image analysis	Ratio scale 0 to 100% (100)	100	LCC (ρ_c)	Mean of $\rho_c = 0.79$ and 0.89 (with and without SADs)	Yadav et al. (2013)
	Manual image analysis	Ratio scale 0 to 100% (100)	100	LCC (ρ_c)	$\rho_c = 0.83$ to 1.00 (mean = 0.95)	Bardsley and Ngugi (2013)
	Manual image analysis	Ratio scale 0 to 100% (100)	100	LCC (ρ_c)	Means of $\rho_c = 0.53$, 0.87, 0.86 and 0.87 (without SADs, with SADs, and with color or black and white SADs)	Schwanck and Del Ponte (2014)
VIS (RGB) image analysis	Images on paper cut by symptom border and weighed	Ratio scale 0 to 100% (100)	100	Regression, intercept and slope coefficient (a and b)	$R^2 = 0.996$; $a = -0.91$; $b = 0.99$	Lindow (1983)
	Planimeter measurement, various pathosystems	Ratio scale 0 to 100% (100)	100	Regression, intercept and slope	$R^2 = 0.976$ to 0.992; $a = 0.914$ to 1.06; $b = -0.17$ to -4.35	Lindow and Webb (1983)
	Images traced from slide projections for contrast and image analyzed	Ratio scale 0 to 100% (100)	100	Regression, intercept and slope	$R^2 = 0.971$ to 0.985; $a = -0.877$ to 0.610; $b = 0.999$ to 1.045	Martin and Rybicki (1998)
	Manual image analysis	Area in pixels	–	Regression, intercept and slope	$R^2 = 0.980$;	Peressotti et al(

Bock *et al. Phytopathology Research*   (2020) 2:9

Page 23 of 30

Table 7 Comparison of different disease severity assessment methods in relation to accuracy-related measures, statistical methods and scale type and resolution used for method validation *(Continued)*

Method	Actual value	Scale type and number of classes used for comparison	Resolution to differentiate severity	Statistic used to assess accuracy	Range	Reference
					$a = 0.901$; $b = 16,097$	2011)
	Severity measured with multiplex real-time PCR	Ratio scale 0 to 100% (100)	100	Regression	$R^2 = 0.9945$	De Coninck et al(2012)
	Visual (pixels)	Ratio scale 0 to 100% (100)	100	Accuracy (%)	Overall accuracy = 96%	Barbedo (2014)
	Visual ratings by 16 raters (inspection deemed image analysis was accurate)	Ratio scale 0 to 100% (100)	100	LCC (ρ_c)	$\rho_c = 0.76$–0.99 (mean = 0.92)	Stewart and McDonald (2014)
	Visual (Pixels)	Ratio scale 0 to 100% (100)	100	PCC (r)	$r = 0.60$–0.90	Clément et al.(2015)
	Manual segmentation using Photoshop	Ratio scale 0 to 100% (100)	100	Quality of segmentation (Q_s)	$Q_s = 84.17\%$	Hu et al. (2017)
	Visual	Ordinal (4 classes)	Healthy stage, early stage, middle stage, end stage	Classification accuracy (%)	Classification accuracy: Healthy stage = 100%; Early stage = 93.1%; Middle stage = 83.3%; End stage = 97.0%	Wang et al(2017)
	Visual	Ordinal (5 classes)	Healthy, very low, low, high, very high	Classification accuracy (%) compared to other diseases and severities	Accuracy of severity measurement = 78.57–86.51% (depending on architecture of CNN)	Esgario et al. (2019)
	Visual	Ordinal (2 classes)	Mild, severe	Classification accuracy (%) compared to other diseases and severities	Severe symptoms = 70.4%; Mild symptoms = 29.4%	Ramcharan et al. (2019)
	Visual	Ordinal (3 classes)	Healthy, general, serious	Proportion accurately classified	0.91	Liang et al. (2019)
Multspectral (MSI) and Hyperspectral (HSI)	Visual	Ordinal (3 classes)	Low, medium high	Classification accuracy (%)	71 to 91%, depending on class (a 5-class scale had accuracy = 11 to 40%)	Coops et al. (2003)
	Visual	Ratio scale 0 to 100% (100)	100	Percentage results with error ≥ 5%	24.1%	Larsolle and Muhammed (2007)
	Visual	Ordinal (9 classes)	0, 1, 10, 20, 30, 45, 60, 80% or 100%	Regression, intercept and slope	$R^2 = 0.91$; $a = 2.40$; $b = -721.22$	Huang et al. (2007)
	Visual	Ordinal (4 classes)	Severe, medium, light, non-visible	None given	None given	Cui et al. (2009, 2010)
	Visual	Ordinal (3 classes)	Healthy tissue, light mycelium, dense mycelium	Classification accuracy (%) including 3 diseases and their severities	Healthy tissue = 100%; Overall accuracy with disease = 61.70 to 98.90%	Mahlein et al. (2012b)

Table 7 Comparison of different disease severity assessment methods in relation to accuracy-related measures, statistical methods and scale type and resolution used for method validation (*Continued*)

Method	Actual value	Scale type and number of classes used for comparison	Resolution to differentiate severity	Statistic used to assess accuracy	Range	Reference
	Symptom progression	Changing symptoms related to spectral changes	–	Metro maps	–	Wahabzada et al. (2015)
	Symptom progression	Changing symptoms related to spectral changes	–	Leaf traces (similar to above)	–	Kuska et al. (2015)
	Visual	Ordinal (9 classes)	0, 1, 5, 10, 20, 40, 60, 80 or 100	Regression (R^2) RMSE	$R^2 > 0.90$; RMSE< 0.15	Wang et al. (2016)
	Visual (in-field disease incidence of infected wheat spikes)	Ratio scale 0 to 100% (100)	100	Regression, intercept and slope (depending on VI)	$R^2 = 0.801, 0.828$; a = 0.2902, 0.4572; b = 0.0013, 0.0020	Kobayashi et al. (2016)
	DNA quantification (presymptomatic)	Continuous	DNA content	Regression (R^2)	$R^2 = 0.868$	Zhao et al. (2017)
	DNA quantification (presymptomatic to symptomatic)	Continuous	DNA content	Regression (R^2)	$R^2 = 0.72$	Thomas et al. (2017)
	Length of lesion	mm	mm	–	Predicted lesion length was proportional to the interior lesion length.	Nagasubramanian et al. (2017)
	Visual	Ordinal (3 classes)	Low (≤5%), moderate (5 to 20%), severe (> 20%) severity.	Classification accuracy (%)	94.83%	Thomas et al. (2018a, b)

– What kind of training is most appropriate?
– Must it be in the specific pathosystem?
– Should training use actual photographs of the target disease, computer-generated images, or a combination of both?

RGB image analysis Research is needed to determine if classification of severity using VIS image analysis and AI techniques provides the resolution and accuracy needed under field conditions.

– Can this be achieved using the 100% ratio scale?
– If ordinal type scales are used, how many classes are needed? How will that vary with pathosystem?
– How can RGB sensor-based systems penetrate the crop canopy where severity estimates of lower leaves might be required?

Databases of annotated images are needed for developing reliable and accurate automated systems based on AI:

– Is development of sufficient image databases for the numbers of diseases and crop combinations practical (true for both VIS and HSI/MSI image analysis)?
– If so, how best to coordinate the logistics of image acquisition?

Particularly for training using AI, systems need to be developed that do not need connectivity to a database:

– Can we develop more efficiently packaged mobile applications?

Explore further combining RGB with HSI/MSI or other techniques:

– Will this help maximize (and possibly synergize) information for accurate measurement of severity?

HSI/MSI and image processing Several of the issues that affect RGB image analysis are common to HSI/MSI too (for example, databases of appropriately ground-truthed images for accurately measuring severity).
Ideally it would be best if hyperspectral signatures were transferrable across scales:

– Can we transfer discriminating hyperspectral signatures to different scales (leaf – plant – field scale) for different diseases?
– If so, are they effective for measuring severity in the variable field situation?
– If scalability is indeed practical for most diseases, how to resolve the issue of proximal and distal sensing and resolution and still maintain accuracy of severity measurements (may not be an issue for detection)?

A major issue that remains is related to data quality:

– How does ground resolution, shadowing, crop motion and image capture influences accuracy of measurements?
– What standard is required for disease measurement?
– Are HSI/MSI measures based on disease development equally or more effective than traditional measures of severity using the percentage scale (metro maps, etc).
– Can more sophisticated mobile platforms or combinations of 3D sensors provide a method to resolve issues of architecture or hidden sampling units?

Intensive knowledge transfer is needed:

– What can we learn from other disciplines such as informatics, medicine, electrical engineering, etc.?

Authors' contributions

CHB led and coordinated the writing of the review, with emphasis on the section on visual disease assessment. EMD provided input on various sections including on SADs and visual disease assessment. JGAB led the section on VIS image analysis, and AKM led the section on HSI/MSI with input from DB. All authors coordinated writing of the introduction and conclusion sections. The author(s) read and approved the final manuscript.

Availability of data and materials
Not applicable.

Ethics approval and consent to participate
Not applicable (no human/animal subjects).

Consent for publication
Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]USDA-ARS Southeastern Fruit and Tree Nut Research Laboratory, Byron, GA 31008, USA. [2]Embrapa Agricultural Informatics, Campinas, SP 13083-886, Brazil. [3]Departamento de Fitopatologia, Universidade Federal de Viçosa, Viçosa, MG 36570-000, Brazil. [4]Institute for Crop Science and Resource Conservation-Phytomedicine, University of Bonn, Nussallee 9, 53115 Bonn, Germany. [5]Institute of Sugar Beet Research, Holtenser Landstrasse 77, 37079 Göttingen, Germany.

References

Abramoff MD, Magalhães PJ, Ram SJ. Image processing with ImageJ. Biophoton Int. 2004;11:36–42.

Alisaac E, Behmann J, Kuska MT, Dehne H, Mahlein A. Hyperspectral quantification of wheat resistance to Fusarium head blight: comparison of two *Fusarium* species. Eur J Plant Pathol. 2018;152:869–84.

Alisaac E, Behmann J, Rathgeb A, Karlovsky P, Dehne HW, Mahlein AK. Assessment of *Fusarium* infection and mycotoxin contamination of wheat kernels and flour using hyperspectral imaging. Toxins. 2019;11(10):556.

Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991.

Amara J, Bouaziz B, Algergawy A. A deep learning-based approach for banana leaf diseases classification. Stuttgart: BTW workshop; 2017. p. 79–88.

Anon. The measurement of potato blight. Trans Br Mycol Soc. 1947;31:140–1.

Anon. Instruction to authors. St Paul: American Phytopathology Society; 2020. https://apsjournals.apsnet.org/page/authorinformation#preparing.

Apan A, Held A, Phinn S, Markley J. Detecting sugarcane 'orange rust' disease using EO-1 Hyperion hyperspectral imagery. Int J Remote Sens. 2004;25:489–98.

Arens N, Backhaus A, Döll S, Fischer S, Seiffert U, Mock H-P. Non-invasive presymptomatic detection of *Cercospora beticola* infection and identification of early metabolic responses in sugar beet. Front Plant Sci. 2016;7:1377.

Ashourloo D, Mobasheri MR, Huete A. Developing two spectral disease indices for detection of wheat leaf rust (*Puccinia triticina*). Remote Sens. 2014;6:4723–40.

Bade CIA, Carmona MA. Comparison of methods to assess severity of common rust caused by *Puccinia sorghi* in maize. Trop Plant Pathol. 2011;36:264–6.

Baird JC, Norma E. Fundamentals of scaling and psychophysics. New York: Wiley; 1978.

Bakr EM. A new software for measuring leaf area, and area damaged by *Tetranychus urticae* Koch. J Appl Entomol. 2005;129:173–5.

Banerjee BP, Raval S, Cullen PJ. UAV-hyperspectral imaging of spectrally complex environments. Int J Remote Sens. 2020;41:4136–59.

Barbedo JGA. Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus. 2013;2:660.

Barbedo JGA. An automatic method to detect and measure leaf disease symptoms using digital image processing. Plant Dis. 2014;98:1709–16.

Barbedo JGA. A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing. Trop Plant Pathol. 2016a; 41:210–24.

Barbedo JGA. A review on the main challenges in automatic plant disease identification based on visible range images. Biosyst Eng. 2016b;144:52–60.

Barbedo JGA. A new automatic method for disease symptom segmentation in digital photographs of plant leaves. Eur J Plant Pathol. 2017;147:349–64.

Barbedo JGA. Plant disease identification from individual lesions and spots using deep learning. Biosyst Eng. 2019;180:96–107.

Barbedo JGA, Koenigkan LV, Halfeld-Vieira BA, Costa RV, Nechet KL, Godoy CV, et al. Annotated plant pathology databases for image-based detection and recognition of diseases. IEEE Lat Am Trans. 2018;16:1749–57.

Bardsley SJ, Ngugi HK. Reliability and accuracy of visual methods to quantify severity of foliar bacterial spot symptoms on peach and nectarine. Plant Pathol. 2013;62:460–74.

Behmann J, Mahlein A-K, Paulus S, Kuhlmann H, Oerke E-C, Plümer L. Calibration of hyperspectral close-range pushbroom cameras for plant phenotyping. ISPRS J Photogramm Remote Sens. 2015a;106:172–82.

Behmann J, Mahlein A-K, Rumpf T, Römer C, Plümer L. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precis Agric. 2015b;16:239–60.

Berdugo CA, Zito R, Paulus S, Mahlein AK. Fusion of sensor data for the detection and differentiation of plant diseases in cucumber. Plant Pathol. 2014;63:1344–56.

Berner DK, Paxson LX. Use of digital images to differentiate reactions of collections of yellow star thistle (*Centaurea solstitialis*) to infection by *Puccinia jaceae*. Biol Control. 2003;28:171–9.

Blackburn GA. Hyperspectral remote sensing of plant pigments. J Exp Bot. 2007; 58:855–67.

Bock CH, Chiang K-S. Disease incidence–severity relationships on leaflets, leaves, and fruit in the pecan–*Venturia effusa* pathosystem. Plant Dis. 2019;103:2865–76.

Bock CH, Chiang KS, del Ponte EM. Accuracy of plant specimen disease severity estimates: concepts, history, methods, ramifications and challenges for the future. CAB Rev. 2016a;11:1–21 https://doi.org/10.1079/PAVSNNR201611032.

Bock CH, Cook AZ, Parker PE, Gottwald TR. Automated image analysis of the severity of foliar citrus canker symptoms. Plant Dis. 2009c;93:660–5.

Bock CH, Gottwald TR, Parker PE, Cook AZ, Ferrandino F, Parnell S, et al. The Horsfall-Barratt scale and severity estimates of citrus canker. Eur J Plant Pathol. 2009b;125:23–38.

Bock CH, Gottwald TR, Parker PE, Ferrandino F, Welham S, van den Bosch F, et al. Some consequences of using the Horsfall-Barratt scale for hypothesis testing. Phytopathology. 2010b;100:1030–41.

Bock CH, Hotchkiss MW, Wood BW. Assessing disease severity: accuracy and reliability of rater estimates in relation to number of diagrams in a standard area diagram set. Plant Pathol. 2016b;65:261–72.

Bock CH, Nutter FW Jr. Detection and measurement of plant disease symptoms using visible-wavelength photography and image analysis. CAB Rev. 2011;6: 1–15 https://doi.org/10.1079/PAVSNNR20116027.

Bock CH, Parker PE, Cook AZ, Gottwald TR. Characteristics of the perception of different severity measures of citrus canker and the relationships between the various symptom types. Plant Dis. 2008a;92:927–39.

Bock CH, Parker PE, Cook AZ, Gottwald TR. Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. Plant Dis. 2008b;92:530–41.

Bock CH, Parker PE, Cook AZ, Gottwald TR. Comparison of assessment of citrus canker foliar symptoms by experienced and inexperienced raters. Plant Dis. 2009a;93:412–24.

Bock CH, Poole GH, Parker PE, Gottwald TR. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. Crit Rev Plant Sci. 2010a;29:59–107.

Bohnenkamp D, Behmann J, Mahlein A-K. In-field detection of yellow rust in wheat on the ground canopy and UAV scale. Remote Sens. 2019;11:2495.

Braido R, Goncalves-Zuliani AMO, Janeiro V, Carvalho SA, Junior JB, Bock CH, et al. Development and validation of standard area diagrams as assessment aids for estimating the severity of citrus canker on unripe oranges. Plant Dis. 2014;98:1543–50.

Bravo C, Moshou D, West J, McCartney A, Ramon H. Early disease detection in wheat fields using spectral reflectance. Biosyst Eng. 2003;84:137–45.

Brugger A, Behmann J, Paulus S, Luigs H-G, Kuska MT, Schramowski P, et al. Extending hyperspectral imaging for plant phenotyping to the UV-range. Remote Sens. 2019;11:1401.

Camargo A, Smith JS. An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosyst Eng. 2009;102:9–21.

Campbell CL, Madden LV. Introduction to plant disease epidemiology. New York: Wiley; 1990.

Carter GA, Knapp AK. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. Amer J Bot. 2001; 88(4):677–84.

Chaube HS, Singh US. Plant disease management: principles and practices. Boca Raton: CRC Press; 1991.

Chen F, Lou S, Fan Q, Wang C, Claverie M, Wang C, et al. Normalized difference vegetation index continuity of the Landsat 4-5 MSS and TM: investigations based on simulation. Remote Sens. 2019;11:1681.

Chester KS. Plant disease losses: their appraisal and interpretation. Plant Dis Rep. 1950;193(Suppl):190–362.

Chiang K-S, Bock CH, El Jarroudi M, Delfosse P, Lee IH, Liu HI. Effects of rater bias and assessment method on disease severity estimation with regard to hypothesis testing. Plant Pathol. 2016a;65:523–35.

Chiang K-S, Bock CH, Lee IH, El Jarroudi M, Delfosse P. Plant disease severity assessment - how rater bias, assessment method and experimental design affect hypothesis testing and resource use efficiency. Phytopathology. 2016b; 106:1451–64.

Chiang K-S, Liu HI, Bock CH. A discussion on disease severity index values. Part I: warning on inherent errors and suggestions to maximize accuracy. Ann Appl Biol. 2017a;171:139–54.

Chiang K-S, Liu HI, Chen YL, El Jarroudi M, Bock CH. Quantitative ordinal scale estimates of plant disease severity: comparing treatments using a proportional odds model. Phytopathology. 2019; https://doi.org/10.1094/PHYTO-10-18-0372-R.

Chiang K-S, Liu HI, Tsai JW, Tsai JR, Bock CH. A discussion on disease severity index values. Part II: using the disease severity index for null hypothesis testing. Ann Appl Biol. 2017b;171:490–505.

Chiang K-S, Liu SC, Bock CH, Gottwald TR. What interval characteristics make a good categorical disease assessment scale? Phytopathology. 2014;104:575–85.

Christ BJ. Effect of disease assessment method on ranking potato cultivars for resistance to early blight. Plant Dis. 1991;75:353–6.

Clément A, Verfaille T, Lormel C, Jaloux B. A new colour vision system to quantify automatically foliar discoloration caused by insect pests feeding on leaf cells. Biosyst Eng. 2015;133:128–40.

Cobb NA. Contribution to an economic knowledge of the Australian rusts (Uredinae). Agric Gaz NSW. 1892;3:60.

Contreras-Medina LM, Osornio-Rios RA, Torres-Pacheco I, Romero-Troncoso RJ, Guevara-González RG, Millan-Almaraz JR. Smart sensor for real-time quantification of common symptoms present in unhealthy plants. Sensors. 2012;12:784–805.

Cooke BM. Disease assessment and yield loss. In: Cooke BM, Jones DG, Kaye B, editors. The epidemiology of plant diseases. 2nd ed. The Netherlands: Springer; 2006.

Coops N, Stanford M, Old K, Dudzinski M, Culvenor D, Stone C. Assessment of *Dothistroma* needle blight of *Pinus radiata* using airborne hyperspectral imagery. Phytopathology. 2003;93:1524–32.

Cui D, Zhang Q, Li M, Hartman GL, Zhao Y. Image processing methods for quantitatively detecting soybean rust from multispectral images. Biosyst Eng. 2010;107:186–93.

Curran PJ. Remote sensing of foliar chemistry. Remote Sens Environ. 1989;30:271–8.

De Castro AI, Ehsani R, Ploetz RC, Crane JH, Buchanon S. Detection of laurel wilt disease in avocado using low altitude aerial imaging. PLoS One. 2015;10: e0124642.

De Coninck BMA, Amand O, Delauré SL, Lucas S, Hias N, Weyens G, et al. The use of digital image analysis and real-time PCR fine-tunes bioassays for quantification of Cercospora leaf spot disease in sugar beet breeding. Plant Pathol. 2012;61:76–84.

Debona D, Nascimento KJT, Rezende D, Rios JA, Bernardeli AMA, Silva LC, et al. A set of standard area diagrams to assess severity of frogeye leaf spot on soybean. Eur J Plant Pathol. 2015;142:603–14.

Del Ponte EM, Nelson SC, Pethybridge SJ. Evaluation of app-embedded disease scales for aiding visual severity estimation of Cercospora leaf spot of table beet. Plant Dis. 2019;103:1347–56.

Del Ponte EM, Pethybridge SJ, Bock CH, Michereff SJ, Machado FJ, Spolti P. Standard area diagrams for aiding severity estimation: scientometrics, pathosystems, and methodological trends in the last 25 years. Phytopathology. 2017;107:1161–74.

Delalieux S, Auwerkerken A, Verstraeten W, Somers B, Valcke R, Lhermitte S, et al. Hyperspectral reflectance and fluorescence imaging to detect scab induced stress in apple leaves. Remote Sens. 2009a;1:858–74.

Delalieux S, Somers B, Verstraeten WW, van Aardt JAN, Keulemans W, Coppin P. Hyperspectral indices to diagnose leaf biotic stress of apple plants, considering leaf phenology. Int J Remote Sens. 2009b;30:1887–912.

Delalieux S, van Aardt J, Keulemans W, Schrevens E, Coppin P. Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications. Eur J Agronomy. 2007;27:130–43.

Devadas R, Lamb DW, Simpfendorfer S, Backhouse D. Evaluating ten spectral vegetation indices for identifying rust infection in individual wheat leaves. Precis Agric. 2009;10:459–70.

Domiciano GP, Duarte HSS, Moreira EN, Rodrigues FA. Development and validation of a set of standard area diagrams to aid in estimation of spot blotch severity on wheat leaves. Plant Pathol. 2014;63:922–8.

Duan J, Zhao B, Wang Y, Yang W. Development and validation of a standard area diagram set to aid estimation of bacterial spot severity on tomato leaves. Eur J Plant Pathol. 2015;142:665–75.

Duarte HSS, Zambolim L, Capucho AS, Nogueira Júnior AF, Rosado AWC, Cardoso CR, et al. Development and validation of a set of standard area diagrams to estimate severity of potato early blight. Eur J Plant Pathol. 2013;137:249–57.

Duarte-Carvajalino JM, Alzate DF, Ramirez AA, Santa-Sepulveda JD, Fajardo-Rojas AE, Soto-Suárez M. Evaluating late blight severity in potato crops using unmanned aerial vehicles and machine learning algorithms. Remote Sens. 2018;10:1513.

El Jarroudi M, Kouadio AL, Mackels C, Tychon B, Delfosse P, Bock CH. A comparison between visual estimates and image analysis measurements to determine Septoria leaf blotch severity in winter wheat. Plant Pathol. 2015;64:355–64.

Elvidge CD. Visible and near infrared reflectance characteristics of dry plant materials. Int J Remote Sens. 1990;11:1775–95.

Esgario JGM, Krohling RA, Ventura JA. Deep learning for classification and severity estimation of coffee leaf biotic stress. arXiv. 2019; https://arxiv.org/pdf/1907.11561.pdf. (11 pages).

Fiorani F, Schurr U. Future scenarios for plant phenotyping. Annu Rev Plant Biol. 2013;64:267–91.

Forbes GA, Jeger MJ. Factors affecting the estimation of disease intensity in simulated plant structures. J Plant Dis Prot. 1987;94:113–20.

Forbes GA, Korva JT. The effect of using a Horsfall-Barratt scale on precision and accuracy of visual estimation of potato late blight severity in the field. Plant Pathol. 1994;43:675–82.

Franceschini MHD, Bartholomeus H, van Apeldoorn DF, Suomalainen J, Kooistra L. Feasibility of unmanned aerial vehicle optical imagery for early detection and severity assessment of late blight in potato. Remote Sens. 2019;11:224.

Fu LY, Wang Y-G, Liu CJ. Rank regression for analyzing ordinal qualitative data for treatment comparison. Phytopathology. 2012;102:1064–70.

Gamon JA, Peñuelas J, Field CB. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sens Environ. 1992;41: 35–44.

Ganthaler A, Losso A, Mayr S. Using image analysis for quantitative assessment of needle bladder rust disease of Norway spruce. Plant Pathol. 2018;67:1122–30.

Garcia-Ruiz F, Sankaran S, Maja JM, Lee WS, Rasmussen J, Ehsani R. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. Comput Electron Agric. 2013;91:106–15.

Gates DM, Keegan HJ, Schleter JC, Weidner VR. Spectral properties of plants. Appl Opt. 1965;4:11–20.

Gay A, Thomas H, Roca M, James C, Taylor J, Rowland J, et al. Nondestructive analysis of senescence in mesophyll cells by spectral resolution of protein synthesis-dependent pigment metabolism. New Phytol. 2008;179:663–74.

Gent DH, Claasen BJ, Tworney MC, Wolfenbarger SN, Woods JL. Susceptibility of hop crown buds to powdery mildew and its relation to perennation of *Podosphaera macularis*. Plant Dis. 2018;102:1316–25.

Ghosal S, Blystone D, Singh AK, Ganapathysubramanian B, Singh A, Sarkar S. An explainable deep machine vision framework for plant stress phenotyping. Proc Natl Acad Sci U S A. 2018;115:4613–8.

Gitelson AA, Merzlyak MN, Chivkunova OB. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. Photochem Photobiol. 2001;74:38–45.

Gitelson AA, Peng Y, Huemmrich KF. Relationship between fraction of radiation absorbed by photosynthesizing maize and soybean canopies and NDVI from remotely sensed data taken at close range and from MODIS 250 m resolution data. Remote Sens Environ. 2014;147:108–20.

Gitelson AA, Zur Y, Chivkunova OB, Merzlyak MN. Assessing carotenoid content in plant leaves with reflectance spectroscopy. Photochem Photobiol. 2002;75: 272–81.

Goclawski J, Sekulska-Nalewajko J, Kuzniak E. Neural network segmentation of images from stained cucurbits leaves with colour symptoms of biotic and abiotic stresses. Int J Appl Math Comput Sci. 2012;22:669–84.

Godoy CV, Koga LJ, Canteri MG. Diagrammatic scale for assessment of soybean rust severity. Fitopatol Bras. 2006;31:63–8 https://doi.org/10.1590/S0100-41582006000100011.

González-Domínguez E, Martins RB, Del Ponte EM, Michereff SM, García-Jiménez J, Armengol J. Development and validation of a standard area diagram set to aid assessment of severity of loquat scab on fruit. Eur J Plant Pathol. 2014; 139:419–28.

Goodwin PH, Hsiang T. Quantification of fungal infection of leaves with digital images and Scion image software. Methods Mol Biol. 2010;638:125–35.

Gottwald TR, da Graça JV, Bassanezi RB. Citrus Huanglongbing: the pathogen and its impact. Plant Health Prog. 2007;8(1) https://doi.org/10.1094/PHP-2007-0906-01-RV.

Hahn SK, Howland AK, Terry ER. Correlated resistance of cassava to mosaic and bacterial blight diseases. Euphytica. 1980;29:305–11.

Hamada NA, Moreira RR, Nesi CN, De Mio LLM. Pathogen dispersal and *Glomerella* leaf spot progress within apple canopy in Brazil. Plant Dis. 2019; 103:3209–17.

Hartung K, Piepho H-P. Are ordinal rating scales better than percent ratings? - a statistical and "psychological" view. Euphytica. 2007;155:15–26.

Hau B, Kranz J, König R. Errors in the assessment of plant disease severities. J Plant Dis Prot. 1989;96:649–74.

Haynes KG, Christ BJ, Weingartner DP, Douches DS, Thill CA, Secor G, et al. Foliar resistance to late blight in potato clones evaluated in national trials in 1997. Am J Potato Res. 2002;79:451–7.

Heim RHJ, Wright IJ, Allen AP, Geedicke I, Oldeland J. Developing a spectral disease index for myrtle rust (*Austropuccinia psidii*). Plant Pathol. 2019;68:738–45.

Hernández-Rabadán DL, Ramos-Quintana F, Guerrero JJ. Integrating SOMs and a Bayesian classifier for segmenting diseased plants in uncontrolled environments. Sci World J. 2014;2014:214674 https://doi.org/10.1155/2014/214674.

Hetzroni A, Miles GE, Engel BA, Hammer PA, Latin RX. Machine vision monitoring of plant health. Adv Space Res. 1994;14(11):203–12.

Hillnhütter C, Mahlein A-K, Sikora RA, Oerke EC. Use of imaging spectroscopy to discriminate symptoms caused by *Heterodera schachtii* and *Rhizoctonia solani* on sugar beet. Precis Agric. 2012;13:17–32.

Hillnhütter C, Mahlein A-K, Sikora RA, Oerke E-C. Remote sensing to detect plant stress induced by *Heterodera schachtii* and *Rhizoctonia solani* in sugar beet fields. Field Crop Res. 2011;122:70–7.

Hindle PH. Historical development. In: Burns DA, Ciurczak EW, editors. Handbook of near-infrared analysis. 3rd ed. Boca Raton: CRC Press; 2008. p. 3–6.

Horsfall JG, Barratt RW. An improved grading system for measuring plant disease. Phytopathology. 1945;35:655.

Horsfall JG, Cowling EB. Pathometry: the measurement of plant disease. In: Horsfall JG, Cowling EB, editors. Plant disease: an advanced treatise, vol. II. New York: Academic Press; 1978. p. 120–36.

Horsfall JG, Heuberger JW. Measuring magnitude of a defoliation disease of tomatoes. Phytopathology. 1942;32:226–32.

Horvath B, Vargas JJ. Analysis of dollar spot disease severity using digital image analysis. Int Turfgrass Soc Res J. 2005;10:196–201.

Hu Q-X, Tian J, He D-J. Wheat leaf lesion color image segmentation with improved multichannel selection based on the Chan–Vese model. Comput Electron Agric. 2017;135:260–8.

Huang K-Y. Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. Comput Electron Agric. 2007;57:3–11.

Huang W, Lamb DW, Niu Z, Zhang Y, Liu L, Wang J. Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. Precis Agric. 2007;8:187–97.

Hughes DP, Salathé M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv. 2015; https://arxiv.org/ftp/arxiv/papers/1511/1511.08060.pdf. (13 pages).

Hunter RE, Roberts DD. A disease grading system for pecan scab [*Fusicladium effusum*]. Pecan Quarterly. 1978;12:3–6.

Ihlow A, Schweizer P, Seiffert U. A high-throughput screening system for barley/powdery mildew interactions based on automated analysis of light micrographs. BMC Plant Biol. 2008;8:6.

Jackson RD, Huete AR. Interpreting vegetation indices. Prev Vet Med. 1991;11: 185–200.

James WC. An illustrated series of assessment keys for plant diseases, their preparation and usage. Can Plant Dis Surv. 1971;51:39–65.

James WC. Assessment of plant disease losses. Annu Rev Phytopathol. 1974;12: 27–48.

Jay S, Bendoula R, Hadoux X, Féret J-B, Gorretta N. A physically-based model for retrieving foliar biochemistry and leaf orientation using close-range imaging spectroscopy. Remote Sens Environ. 2016;177:220–36.

Johnson DA, Alldredge JR, Hamm PB, Frazier BE. Aerial photography used for spatial pattern analysis of late blight infection in irrigated potato circles. Phytopathology. 2003;93:805–12.

Jones MM, Stansly PA. Frequent low volume sprays of horticultural mineral oil (HMO) for psyllid and leafminer control. J Citrus Pathol. 2014;1(1):178 https://escholarship.org/uc/item/1z03d071.

Karisto P, Hund A, Yu K, Anderegg J, Walter A, Mascher F, et al. Ranking quantitative resistance to *Septoria tritici* blotch in elite wheat cultivars using automated image analysis. Phytopathology. 2018;108:568–81.

Kerguelen V, Hoddle MS. Measuring mite feeding damage on avocado leaves with automated image analysis software. The Florida Entomol. 1999;82:119–22.

Kersting K, Bauckhage C, Wahabzada M, Mahlein A-K, Steiner U, et al. Feeding the world with big data: uncovering spectral characteristics and dynamics of stressed plants. In: Lässig J, Kersting K, Morik K, editors. Computational sustainability. Cham: Springer; 2016. p. 99–120.

Kobayashi T, Sasahara M, Kanda E, Ishiguro K, Hase S, Torigoe Y. Assessment of rice panicle blast disease using airborne hyperspectral imagery. Open Agric J. 2016;10:28–34.

Koch H, Hau B. A psychological aspect of plant disease assessment. J Plant Dis Prot. 1980;87:587–93.

Kokko EG, Conner RL, Lee B, Kuzyk AD, Kozu GC. Quantification of common root rot symptoms in resistant and susceptible barley by image analysis. Can J Plant Pathol. 2000;22:38–43.

Kora C, McDonald MR, Boland GJ. Epidemiology of Sclerotinia rot of carrot caused by *Sclerotinia sclerotiorum*. Can J Plant Pathol. 2005;27:245–58.

Kranz J. Measuring plant disease. In: Kranz J, Rotem J, editors. Experimental techniques in plant disease epidemiology. New York: Springer Verlag; 1988. p. 35–50.

Kruse OMO, Prats-Montalbán JM, Indahl UG, Kvaal K, Ferrer A, Futsaether CM. Pixel classification methods for identifying and quantifying leaf surface injury from digital images. Comput Electron Agric. 2014;108:155–65.

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

Kuska MT, Mahlein A-K. Aiming at decision making in plant disease protection and phenotyping by the use of optical sensors. Eur J Plant Pathol. 2018;152: 987–92.

Kutcher HR, Turkington TK, McLaren DL, Irvine RB, Brar GS. Fungicide and cultivar management of leaf spot diseases of winter wheat in western Canada. Plant Dis. 2018;102:1828–33.

Kuźniak E, Świercz U, Chojak J, Sekulska-Nalewajko J, Gocławski J. Automated image analysis for quantification of histochemical detection of reactive oxygen species and necrotic infection symptoms in plant leaves. J Plant Interact. 2014;9:167–74.

Kwack MS, Kim EN, Lee H, Kim J-W, Chun S-C, Kim KD. Digital image analysis to measure lesion area of cucumber anthracnose by *Colletotrichum orbiculare*. J Gen Plant Pathol. 2005;71:418–21.

Laflamme B, Middleton M, Lo T, Desveaux D, Guttman DS. Image-based quantification of plant immunity and disease. Mol Plant-Microbe Interact. 2016;29:919–24.

Lamari L. ASSESS 2.0: image analysis software for plant disease quantification. St Paul: APS Press; 2002.

Large EC. Measuring plant disease. Annu Rev Phytopathol. 1966;4:9–26.

Larsolle A, Muhammed HH. Measuring crop status using multivariate analysis of hyperspectral field reflectance with application to disease severity and plant density. Precis Agric. 2007;8:37–47.

Lelong CCD, Burger P, Jubelin G, Roux BL, Kabbe S, Baret F. Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots. Sensors. 2008;8:3557–85.

Leucker M, Mahlein AK, Steiner U, Oerke EC. Improvement of lesion phenotyping in Cercospora beticola – sugar beet interaction by hyperspectral imaging. Phytopathology. 2016;106:177–84.

Leucker M, Wahabzada M, Kersting K, Peter M, Beyer W, Steiner U, et al. Hyperspectral imaging reveals the effect of sugar beet quantitative trait loci on Cercospora leaf spot resistance. Funct Plant Biol. 2017;44:1–9.

Liang W, Zhang H, Zhang G, Cao H-X. Rice blast disease recognition using a deep convolutional neural network. Sci Rep. 2019;9:2869.

Likert R. A technique for the measurement of attitudes. Arch Psychol. 1932;140:1–55.

Lindow SE. Estimating disease severity of single plants. Phytopathology. 1983;73: 1576–81.

Lindow SE, Webb RR. Quantification of foliar plant disease symptoms by microcomputer-digitized video image analysis. Phytopathology. 1983;73:520–4.

Liu W, Cao X, Fan J, Wang Z, Yan Z, Luo Y, et al. Detecting wheat powdery mildew and predicting grain yield using unmanned aerial photography. Plant Dis. 2018;102:1981–8.

Lloret J, Bosch I, Sendra S, Serrano A. A wireless sensor network for vineyard monitoring that uses image processing. Sensors. 2011;11:6165–96.

Macedo-Cruz A, Pajares G, Santos M, Vilegas-Romero I. Digital image sensor-based assessment of the status of oat (*Avena sativa* L.) crops after frost damage. Sensors. 2011;11:6015–36.

Madden LV, Hughes G, van den Bosch F. The study of plant disease epidemics. St Paul: APS Press; 2007.

Mahlein A-K. Plant disease detection by imaging sensors—parallels and specific demands for precision agriculture and plant phenotyping. Plant Dis. 2016;100: 241–51.

Mahlein A-K, Kuska MT, Behmann J, Polder G, Walter A. Hyperspectral sensors and imaging technologies in phytopathology: state of the art. Annu Rev Phytopathol. 2018;56:535–58.

Mahlein A-K, Kuska MT, Thomas S, Wahabzada M, Behmann J, Rascher U, et al. Quantitative and qualitative phenotyping of disease resistance of crops by hyperspectral sensors: seamless interlocking of phytopathology, sensors, and machine learning is needed! Curr Opin Plant Biol. 2019;50:156–62.

Mahlein A-K, Oerke EC, Steiner U, Dehne HW. Recent advances in sensing plant diseases for precision crop protection. Eur J Plant Pathol. 2012a;133:197–209.

Mahlein A-K, Rumpf T, Welke P, Dehne H-W, Plümer L, Steiner U, et al. Development of spectral indices for detecting and identifying plant diseases. Remote Sens Environ. 2013;128:21–30.

Mahlein A-K, Steiner U, Dehne H-W, Oerke E-C. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. Precis Agric. 2010;11:413–31.

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

Manso GL, Knidel H, Krohling RA, Ventura JA. A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust. arXiv. 2019; https://arxiv.org/pdf/1904.00742v1.pdf. (36 pages).

Martin DP, Rybicki EP. Microcomputer-based quantification of maize streak virus symptoms in *Zea mays*. Phytopathology. 1998;88:422–7.

McBride GB. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. NIWA Client Report. 2005:HAM2005–62 https://www.medcalc.org/download/pdf/McBride2005.pdf.

Merzlyak MN, Gitelson AA, Chivkunova OB, Rakitin VY. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiol Plant. 1999;106:135–41 https://doi.org/10.1034/j.1399-3054.1999. 106119.x.

Michereff SJ, Noronha MA, Lima GSA, Albert ÍCL, Melo EA, Gusmão LO. Diagrammatic scale to assess downy mildew severity in melon. Hortic Bras. 2009;27:76–9.

Mirik M, Michels GJ, Kassymzhanova-Mirik S, Elliott NC, Catana V, Jones DB, et al. Using digital image analysis and spectral reflectance data to quantify damage by greenbug (Hemitera: Aphididae) in winter wheat. Comput Electron Agric. 2006;51:86–98.

Mishra P, Nordon A, Tschannerl J, Lian G, Redfern S, Marshall S. Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products. J Food Eng. 2018;238:70–7.

Miyasaka SC, McCulloch CE, Nelson SC. Taro germplasm evaluated for resistance to taro leaf blight. Hort Technol. 2012;22:838–49.

Moore WC. The measurement of plant disease in the field: preliminary report of a sub-committee of the Society's plant pathology committee. Trans Br Mycol Soc. 1943;26:28–35.

Mutka AM, Bart RS. Image-based phenotyping of plant disease symptoms. Front Plant Sci. 2015;5:734.

Mutka AM, Fentress SJ, Sher JW, Berry JC, Pretz C, Nusinow DA, et al. Quantitative, image-based phenotyping methods provide insight into spatial and temporal dimensions of plant disease. Plant Physiol. 2016;172:650–60.

Mwebaze E, Owomugisha G. Machine learning for plant disease incidence and severity measurements from leaf images. In: Proceedings of the 15th IEEE international conference on machine learning and applications (ICMLA), Anaheim, USA; 2016. p. 158–63.

Nagasubramanian K, Jones S, Sarkar S, Singh AK, Singh A. Ganapathysubramanian B. Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean. arXiv. 2017; https://arxiv.org/ftp/arxiv/papers/1710/1710.04681.pdf. (20 pages).

Nagasubramanian K, Jones S, Singh AK, Sarkar S, Singh A, Ganapathysubramanian B. Plant disease identification using explainable 3D deep learning on hyperspectral images. Plant Methods. 2019;15:98.

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

Newell LC, Tysdal HM. Numbering and note-taking systems for use in improvement of forage crops. J Amer Soc Agron. 1945;37:736–49.

Newton AC, Hackett CA. Subjective components of mildew assessment on spring barley. Eur J Plant Pathol. 1994;100:395–412.

Nilsson H-E. Remote sensing and image analysis in plant pathology. Annu Rev Phytopathol. 1995;15:489–527.

Nita M, Ellis MA, Madden LV. Reliability and accuracy of visual estimation of Phomopsis leaf blight of strawberry. Phytopathology. 2003;93:995–1005.

Nutter FW Jr, Esker PD. The role of psychophysics in phytopathology. Eur J Plant Pathol. 2006;114:199–213.

Nutter FW Jr, Gleason ML, Jenco JH, Christians NC. Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. Phytopathology. 1993;83:806–12.

Nutter FW Jr, Litwiller D. A computer program to generate standard area diagrams to aid raters in assessing disease severity. Phytopathology. 1998;88:S117.

Nutter FW Jr, Schultz PM. Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. Can J Plant Pathol. 1995;17:174–84.

Nutter FW Jr, Teng PS, Shokes FM. Disease assessment terms and concepts. Plant Dis. 1991;75:1187–8.

O'Neal ME, Landis DA, Isaacs R. An inexpensive, accurate method for measuring leaf area and defoliation through digital image analysis. J Econ Entomol. 2002;95:1190–4.

Oerke E-C. Crop losses to pests. J Agric Sci. 2006;144:31–43.

Oerke E-C, Herzog K, Toepfer R. Hyperspectral phenotyping of the reaction of grapevine genotypes to *Plasmopara viticola*. J Exp Bot. 2016;67:5529–43.

Oerke E-C, Steiner U. Potential of digital thermography for disease control. In: Oerke E-C, Gerhards R, Menz G, Sikora R, editors. Precision crop protection-the challenge and use of heterogeneity. Dordrecht: Springer; 2010.

Olmstead JW, Lang GA, Grove GG. Assessment of severity of powdery mildew infection of sweet cherry leaves by digital image analysis. Hortic Sci. 2001;36:107–11.

Parker SR, Shaw MW, Royle DJ. The reliability of visual estimates of disease severity on cereal leaves. Plant Pathol. 1995a;44:856–64.

Parker SR, Shaw MW, Royle DJ. Reliable measurement of disease severity. Asp Appl Biol. 1995b;43:205–14.

Patil SB, Bodhe SK. Leaf disease severity measurement using image processing. Int J Engin Tech. 2011;3:297–301.

Paul PA, El-Allaf SM, Lipps PE, Madden LV. Relationships between incidence and severity of Fusarium head blight on winter wheat in Ohio. Phytopathology. 2005;95:1049–60.

Pedroso C, Lage DAC, Henz GP, Café-Filho AC. Development and validation of a diagrammatic scale for estimation of anthracnose on sweet pepper fruits for epidemiological studies. J Plant Pathol. 2011;93:219–25.

Peressotti E, Duchêne E, Merdinoglu D, Mestre P. A semiautomatic non-destructive method to quantify grapevine downy mildew sporulation. J Microbiol Methods. 2011;84:265–71.

Pethybridge SJ, Nelson SC. Leaf doctor: a new portable application for quantifying plant disease severity. Plant Dis. 2015;99:1310–6.

Price TV, Gross R, Wey JH, Osborne CF. A comparison of visual and digital image-processing methods in quantifying the severity of coffee leaf rust (*Hemileia vastatrix*). Aust J Exp Agric. 1993;33:97–101.

Ramcharan A, McCloskey P, Baronowski K, Mbiliyni N, Mrisho L, Ndalawha M, et al. A mobile-based deep learning model for cassava disease diagnosis. Front Plant Sci. 2019;10:272.

Rioux RA, Van Ryzin BJ, Kerns JP. Brachypodium: a potential model host for fungal pathogens of turfgrasses. Phytopathology. 2017;107:749–57.

Rouse JW, Haas RH, Schell JA, Deering DW. Monitoring vegetation systems in the Great Plains with ERTS. In: Freden SC, Mercanti EP, Becker M, editors. Third earth resources technology satellite–1 syposium NASA, NASA SP-351, Washington DC; 1974. p. 309–17.

Rumpf T, Mahlein A-K, Steiner U, Oerke E-C, Dehne H-W, Plümer L. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Comput Electron Agric. 2010;74:91–9.

Sankaran S, Mishra A, Ehsani R, Davis C. A review of advanced techniques for detecting plant diseases. Comput Electron Agric. 2010;72:1–13.

Savary S, Bregaglio S, Willocquet L, Gustafson D, Mason D'Croz D, Sparks A, et al. Crop health and its global impacts on the components of food security. Food Secur. 2017;9:311–27.

Savary S, Ficke A, Aubertot J-N, Hollier C. Crop losses due to diseases and their implications for global food production losses and food security. Food Secur. 2012;4:519–37.

Schwanck AA, Del Ponte EM. Accuracy and reliability of severity estimates using linear or logarithmic disease diagram sets in true colour or black and white: a study case for rice brown spot. J Phytopathol. 2014;162:670–82.

Seelig H-D, Hoehn A, Stodieck LS, Klaus DM, Adams WW III, Emery WJ. The assessment of leaf water content using leaf reflectance ratios in the visible, near-, and short-wave-infrared. Int J Remote Sens. 2008;29:3701–13.

Shah DA, Madden LV. Nonparametric analysis of ordinal data in designed factorial experiments. Phytopathology. 2004;94:33–43.

Sherwood RT, Berg CC, Hoover MR, Zeiders KE. Illusions in visual assessment of Stagonospora leaf spot of orchard grass. Phytopathology. 1983;73:173–7.

Shrivastava S, Singh SK, Hooda DS. Color sensing and image processing-based automatic soybean plant foliar disease severity detection and estimation. Multimed Tools Appl. 2015;74:11467–84 https://doi.org/10.1007/s11042-014-2239-0.

Sibiya M, Sumbwanyambe M. An algorithm for severity estimation of plant leaf diseases by the use of colour threshold image segmentation and fuzzy logic inference: a proposed algorithm to update a "leaf doctor" application. AgriEngineering. 2019;1:205–19.

Silvertown J. A new dawn for citizen science. Trends Ecol Evol. 2009;24:467–71.

Simko I, Jimenez-Berni JA, Sirault XRR. Phenomic approaches and tools for phytopathologists. Phytopathology. 2017;107:6–17.

Singh A, Ganapathysubramanian B, Singh AK, Sarkar S. Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci. 2016;21(2):110–24.

Škaloudová B, Křivan V, Zemek R. Computer-assisted estimation of leaf damage caused by spider mites. Comput Electron Agric. 2006;53:81–91.

Spolti P, Schneider L, Sanhueza RMV, Batzer JC, Gleason ML, Del Ponte EM. Improving sooty blotch and flyspeck severity estimation on apple fruit with the aid of standard area diagrams. Eur J Plant Pathol. 2011;129:21–9.

Steddom K, McMullen M, Schatz B, Rush CM. Comparing image format and resolution for assessment of foliar diseases of wheat. Plant Health Prog. 2005; https://doi.org/10.1094/PHP-2005-0516-01-RS.

Stevens SS. On the theory of scales of measurement. Science. 1946;103:677–80.

Stewart EL, Hagerty CH, Mikaberidze A, Mundt CC, Zhong Z, McDonald BA. An improved method for measuring quantitative resistance to the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. Phytopathology. 2016;106:782–8.

Stewart EL, McDonald BA. Measuring quantitative virulence in the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. Phytopathology. 2014;104:985–92.

Strange RN, Scott PR. Plant disease: a threat to global food security. Annu Rev Phytopathol. 2005;43:83–116.

Strayer-Scherer A, Liao YY, Young M, Ritchie L, Vallad GE, Santra S, et al. Advanced copper composites against copper-tolerant *Xanthomonas perforans* and tomato bacterial spot. Phytopathology. 2018;108:196–205.

Sugiura R, Tsuda S, Tamiya S, Itoh A, Nishiwaki K, Murakami N, et al. Field phenotyping system for the assessment of potato late blight resistance using RGB imagery from an unmanned aerial vehicle. Biosyst Eng. 2016;148:1–10.

Sun H, Wei J, Zhang J, Yang W. A comparison of disease severity measurements using image analysis and visual estimates using a category scale for genetic analysis of resistance to bacterial spot in tomato. Eur J Plant Pathol. 2014;139:125–36.

Suomalainen J, Anders N, Iqbal S, Roerink G, Franke J, Wenting P, et al. A lightweight hyperspectral mapping system and photogrammetric processing chain for unmanned aerial vehicles. Remote Sens. 2014;6:11013–30.

Thomas S, Behmann J, Steier A, Kraska T, Muller O, Rascher U, et al. Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. Plant Methods. 2018a;14:45.

Thomas S, Kuska MT, Bohnenkamp D, Brugger A, Alisaac E, Wahabzada M, et al. Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. J Plant Dis Protect. 2018b;125:5–20.

Thomas S, Wahabzada M, Kuska MT, Rascher U, Mahlein A-K. Observation of plant-pathogen interaction by simultaneous hyperspectral imaging reflection and transmission measurements. Funct Plant Biol. 2017;44:23–34.

Tomerlin JR, Howell TA. DISTRAIN: a computer program for training people to estimate disease severity on cereal leaves. Plant Dis. 1988;72:455–9.

Tucker CC, Chakraborty S. Quantitative assessment of lesion characteristics and disease severity using digital image processing. J Phytopathol. 1997; 145:273–8.

Tucker CJ, Dregne HE, Newcomb WW. Expansion and contraction of the Sahara Desert from 1980 to 1990. Science. 1991;253:299–301.

Vale FXR, Fernandes-Filho EI, Liberato JR. QUANT: a software for plant disease severity assessment. P 105. In: Proceedings of the 8th International Congress of Plant Pathology, 2-7 February 2003, Christchurch, New Zealand. Sydney: Published by Horticulture Australia; 2003.

Vieira RF, Paula Júnior TJ, Carneiro JES, Teixeira H, Queiroz TFN. Management of white mold in type III common bean with plant spacing and fungicide. Trop Plant Pathol. 2012;37:95–101.

Virlet N, Sabermanesh K, Sadeghi-Tehran P, Hawkesford MJ. Field Scanalyzer: an automated robotic field phenotyping platform for detailed crop monitoring. Funct Plant Biol. 2017;44:143–53.

Wahabzada M, Mahlein A-K, Bauckhage C, Steiner U, Oerke E-C, Kersting K. Metro maps of plant disease dynamics--automated mining of differences using hyperspectral images. PLoS One. 2015;10:e0116902.

Wahabzada M, Mahlein A-K, Bauckhage C, Steiner U, Oerke E-C, Kersting K. Plant phenotyping using probabilistic topic models: uncovering the hyperspectral language of plants. Sci Rep. 2016;6:22482.

Walter A, Liebisch F, Hund A. Plant phenotyping: from bean weighing to image analysis. Plant Methods. 2015;11:14.

Wang G, Sun Y, Wang J. Automatic image-based plant disease severity estimation using deep learning. Comput intell Neurosci. 2017;2017:2917536 https://doi.org/10.1155/2017/2917536.

Wang H, Qin F, Ruan L, Wang R, Liu Q, et al. Identification and severity determination of wheat stripe rust and wheat leaf rust based on hyperspectral data acquired using a black-paper-based measuring method. PLoS One. 2016;11:e0154648.

Wendel A, Underwood J. Illumination compensation in ground based hyperspectral imaging. ISPRS J Photogramm Remote Sens. 2017;129:162–78.

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

Wiesner-Hanks T, Stewart EL, Kaczmar N, DeChant C, Wu H, Nelson RJ, et al. Image set for deep learning: field images of maize annotated with disease symptoms. BMC Res Notes. 2018;11:440.

Wijekoon CP, Goodwin PH, Hsiang T. Quantifying fungal infection of plant leaves by digital image analysis using Scion image software. J Microbiol Methods. 2008;74:94–101.

Xie W, Yu K, Pauls KP, Navabi A. Application of image analysis in studies of quantitative disease resistance, exemplified using common bacterial blight–common bean pathosystem. Phytopathology. 2012;102:434–42.

Xu W, Haynes KG, Qu X. Characterization of early blight resistance in potato cultivars. Plant Dis. 2019;103:629–37.

Yadav NVS, de Vos SM, Bock CH, Wood BW. Development and validation of standard area diagrams to aid assessment of pecan scab symptoms on fruit. Plant Pathol. 2013;62:325–35.

Zhang D, Zhou X, Zhang J, Lan Y, Xu C, Liang D. Detection of rice sheath blight using an unmanned aerial system with high-resolution color and multispectral imaging. PLoS One. 2018a;13:e0187470.

Zhang J-H, Kong F-T, Wu J-Z, Han S-Q, Zhai Z-F. Automatic image segmentation method for cotton leaves with disease under natural environment. J Integr Agric. 2018b;17:1800–14.

Zhang S, You Z, Wu X. Plant disease leaf image segmentation based on superpixel clustering and EM algorithm. Neural Comput & Applic. 2019;31:S1225–32.

Zhao Y, Gu Y, Qin F, Li X, Ma Z, Zhao L, et al. Application of near-infrared spectroscopy to quantitatively determine relative content of *Puccinia striiformis* f. sp. *tritici* DNA in wheat leaves in incubation period. J Spectrosc. 2017;2017:9740295 https://doi.org/10.1155/2017/9740295.

Zheng Q, Huang W, Cui X, Shi Y, Liu L. New spectral index for detecting wheat yellow rust using Sentinel-2 multispectral imagery. Sensors. 2018;18:868.

Zhou Z, Zang Y, Li Y, Zhang Y, Wang P, Luo X. Rice plant-hopper infestation detection and classification algorithms based on fractal dimension values and fuzzy C-means. Math Comput Model. 2013;58:701–9.

Zhu H, Chu B, Zhang C, Liu F, Jiang L, He Y. Hyperspectral imaging for presymptomatic detection of tobacco disease with successive projections algorithm and machine-learning classifiers. Sci Rep. 2017;7:4125.