

Research Article

Detecting the Early Stage of *Phaeosphaeria* Leaf Spot Infestations in Maize Crop Using In Situ Hyperspectral Data and Guided Regularized Random Forest Algorithm

Elhadi Adam,¹ Houtao Deng,² John Odindi,³ Elfatih M. Abdel-Rahman,^{3,4,5} and Onesimo Mutanga³

¹*School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Private Bag 3, Wits, Johannesburg 2050, South Africa*

²*Intuit Mountain View, Mountain View, CA, USA*

³*School of Agriculture, Earth and Environmental Sciences, Pietermaritzburg Campus, University of KwaZulu-Natal, Scottsville P/Bag X01, Pietermaritzburg 3209, South Africa*

⁴*African Insect Science for Food and Health (ICIPE), P.O. Box 30772, Nairobi 00100, Kenya*

⁵*Department of Agronomy, Faculty of Agriculture, University of Khartoum, 13314 Khartoum North, Sudan*

Correspondence should be addressed to Elhadi Adam; Elhadi.Adam@wits.ac.za

Received 20 August 2016; Accepted 19 December 2016; Published 28 March 2017

Academic Editor: K. S. V. Krishna Rao

Copyright © 2017 Elhadi Adam et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Phaeosphaeria leaf spot (PLS) is considered one of the major diseases that threaten the stability of maize production in tropical and subtropical African regions. The objective of the present study was to investigate the use of hyperspectral data in detecting the early stage of PLS in tropical maize. Field data were collected from healthy and the early stage of PLS over two years (2013 and 2014) using a handheld spectroradiometer. An integration of a newly developed guided regularized random forest (GRRF) and a traditional random forest (RF) was used for feature selection and classification, respectively. The 2013 dataset was used to train the model, while the 2014 dataset was used as independent test dataset. Results showed that there were statistically significant differences in biochemical concentration between the healthy leaves and leaves that were at an early stage of PLS infestation. The newly developed GRRF was able to reduce the high dimensionality of hyperspectral data by selecting key wavelengths with less autocorrelation. These wavelengths are located at 420 nm, 795 nm, 779 nm, 1543 nm, 1747 nm, and 1010 nm. Using these variables ($n = 6$), a random forest classifier was able to discriminate between the healthy maize and maize at an early stage of PLS infestation with an overall accuracy of 88% and a kappa value of 0.75. Overall, our study showed potential application of hyperspectral data, GRRF feature selection, and RF classifiers in detecting the early stage of PLS infestation in tropical maize.

1. Introduction

Maize (*Zea mays* L.) accounts for 15–50% of energy in human diets in sub-Saharan Africa and is a staple diet for majority of the African population (Kagoda et al. 2010). In some sub-Saharan African countries, maize is considered a major feed for livestock [1, 2]. In South Africa, maize is the second most popular crop after sugar cane and accounts for over 50% production within the Southern African Development Community (SADC) region [3]. The crop is grown in two main zones within the country: a marginal western belt

and a reliable and higher productivity eastern core. In these areas, maize is highly dependent on climatic variables such as temperature and precipitation [3]. Less than 10% of the crop is produced under irrigation [3]. Thus, climatic variability and change are key factors influencing interannual maize production in the country.

Commonly, changes in temperature and precipitation and increased frequency and intensity of extreme weather events influence agricultural productivity and food safety through a number of pathways [4]. For example, climate change characterized by increase in temperature and change

in rainfall patterns influences the onset, persistence, and patterns of crop bacteria, viruses, parasites, and fungi [4]. Such changes may also affect plants' physiology and host susceptibility, which may result in the emergence, redistribution, and changes in the incidence and intensity of plant diseases and pest infestations [4, 5].

One of the major diseases that threatens maize production in tropical and subtropical growing areas is *Phaeosphaeria* leaf spot [6, 7]. *Phaeosphaeria* leaf spot (PLS) is a maize foliar disease caused by the ascomycete fungus *Phaeosphaeria maydis* (Henn.). First noted in India, it has recently spread widely in other parts of the world such as Brazil [6, 7], USA [8], and Central, East, and Southern Africa [9, 10]. The disease is predominant in areas of high rainfall and moderate temperatures, a common characteristic of the higher tropical and subtropical elevations [9, 11, 12].

Early symptoms of the PLS are small dark green water-soaked leaf spots which may be circular, oval, elliptic, slightly elongated, and often 0.3 to 2.0 cm in diameter [6, 12]. Typically, lesions are scattered over the leaf surface and have a chlorotic appearance [9, 13]. These turn to pale green, straw-colored, bleached and necrotic, or dried with dark brown margins [9, 13]. Under favourable conditions, these lesions may coalesce to large irregular shapes and blight the entire leaf, in some cases infecting the stem [6].

PLS can result in a considerable reduction in photosynthetic leaf area as the spots coalesce [9]. This can cause premature leaf drying, thus reducing plant cycle, decrease in grain size and weight, and in extreme cases result in early plant death [6, 14]. Thus, substantial grain yield losses ranging from 11% to 60% in PLS susceptible cultivars have been reported in the United States of America and Brazil [8, 14]. Such losses may be attributed to nontranslocation of nitrogen in infected plants [14]. Whereas grain yield losses from PLS are yet to be quantified in South Africa, the maize genotypes in the country have shown to be favourable to PLS infestation [9, 10]. Due to the importance of maize production in the tropical and subtropical regions, losses by the disease not only affect yield, but also micro and macro socioeconomic systems [10, 14]. This, therefore, necessitates the adoption of effective disease management approaches to sustain production.

Using fungicides like mancozeb, applied before or at the early stages of the disease, control and management of PLS has been successful in Brazil [9]. Management approaches using fungicides therefore require that information on PLS infestation is available at appropriate scales and is summarized in a way that allows for suitable and timeous management practice at the right place [15–17].

Traditionally, field survey data, based on expert visual inspection, have been used to identify PLS infestation. However, this requires continuous monitoring, which might be prohibitively expensive, time-consuming, and in some cases impractical on large farms [18, 19]. Recently, remote sensing datasets and techniques have emerged as valuable means to detect and measure crop disease incidences in real time at both regional and farm scales [17, 20–23]. Detailed information and discussion on the use of different remotely sensed data in detecting crop diseases and

pests can be found in two recent comprehensive review papers [17, 24].

The maize physiological characteristics due to PLS infestation are caused by a change in some of the biochemical composition and internal as well as external leaf structure [6, 9, 14]. Consequently, the maize canopy spectral reflectance will also change accordingly across the relevant sections of the electromagnetic spectrum. These spectral characteristics form the basis for use of remote sensing in spatial modelling of crops stressed by the disease [25]. In this paper, we seek to detect the early stage of PLS in tropical maize using a combination of biochemical data and field spectral reflectance measurements.

2. Material and Methods

2.1. The Study Area and the Experimental Setup. This study was conducted in Cedara experimental farm, located in KwaZulu-Natal province, South Africa (30°16 E, 29°32 S, and 1076 metres above sea level). The plantings were done in two replications in November 2013 and 2014. The size of the plot was 3 m long with two interrow spacing of 0.75 m and 0.3 m, respectively. Two seeds were planted per station and later thinned to one. The plant population density was about 44,000 per hectare. Field data was collected during vegetative (VT) growth stage when the last branch of the tassel is visible [26]. The plots were classified based on the number of leaves showing the disease symptoms. The early stage of PLS disease was then assessed based on visual assessment of the plots [9]. *Phaeosphaeria* leaf spot (PLS) disease early stage was assessed fortnightly from the first appearance of symptoms based on visual assessment of the leaf using a 1–9 rating scale [9]: where 1 = 0%, 2 = <1%, 3 = 1–3%, 4 = 4–6%, 5 = 7–12%, 6 = 13–25%, 7 = 26–50%, 8 = 51–75%, and 9 = 75–100% leaf areas showing disease symptoms. The scores were further classified into the following disease reaction types: 1.0 = symptomless (healthy) and 2.0–3.0 = early stage. The other disease stages from 4 to 9 (moderate and severe) were not considered in this study.

2.2. Ground-Based Hyperspectral Measurements. Leaf reflectance spectra were obtained using an Analytical Spectral Devices (ASD Inc., Boulder, CO, USA) FieldSpec³spectrometer. The spectra were collected under sunny and clear-sky conditions between 10:00 am and 02:00 pm local time over two seasons, on January 5, 2013 and 25 January, 2014. The FieldSpec³spectrometer has a spectral range of 350 to 2500 nm and registers radiation at 1.4 nm intervals for the 350–1000 nm spectral region and 2 nm intervals for the 1000–2500 nm spectral regions. Measurements were then interpolated to 1 nm spectral resolution across the spectrum [27]. From each plot, three to five leaves from the top canopy of maize crop were sampled. For each sample unit, piles of maize leaves were arranged and placed randomly on top of a black thick cardboard [28]. The leaf reflectance was then taken immediately at a nadir-looking angle from about 25 cm above the leaves. About 15 to 20 measurements were made from each pile of leaves by moving randomly over each canopy, to derive the representative reflectance

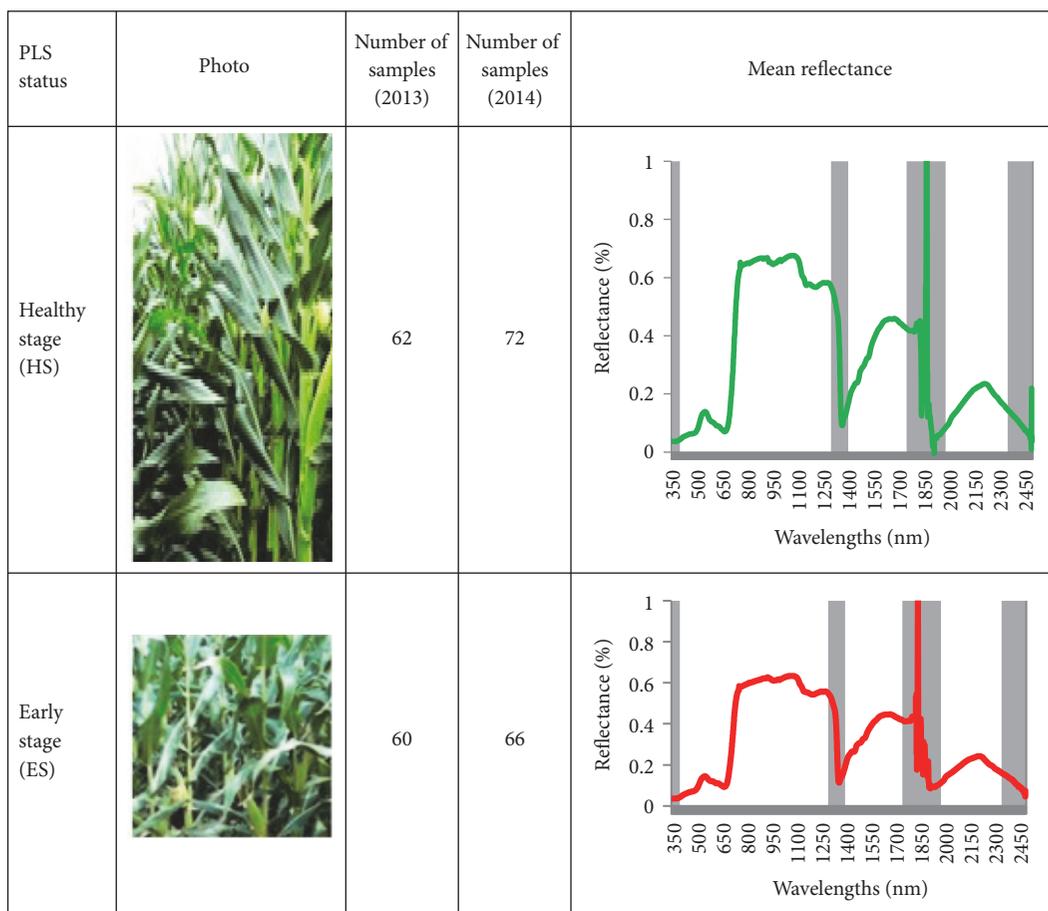


FIGURE 1: The sample plots of 2013 and 2014 with the mean reflectance of healthy and early stage of PLS using data collected in 2013. The grey areas indicate the spectral ranges excluded from the analysis due to the external effects.

spectra for the canopy (Figure 1). These spectral measurements were then averaged to represent the final spectral measurement for each leaf sample. A white reference spectral measure on the calibration panel was performed every 10–20 measurements to offset any change in the atmospheric condition and sun irradiance spectrum [27]. In total, 66 and 72 plots for the PLS early stage and healthy maize were sampled, respectively. Using similar procedure, field spectral measurements were replicated on 15 January, 2014 on the same experimental farm and under similar conditions. Spectral reflectance from 66 to 72 plots for the PLS early stage and healthy maize was collected, respectively.

2.3. Leaf Sampling and Biochemical Analysis. Disease infestation affects the amount and quality of chemical composition and physical structure of the leaves and spectral properties [29]. To test whether PLS has a significant impact on the chemical composition of maize leaves, the piles of leaf samples ($n = 2-5$) were packed immediately after the spectral measurements from both the healthy and the PLS early stage. The samples were then pooled, bagged, dried at 70°C for 48 h, and sent for full biochemical analysis at the Department of Agriculture and Environmental Affairs Feed laboratory in Cedara, South Africa. A *t*-test was used to determine the

effect of PLS disease on the biochemical composition of the leaf. The *t*-test was used to determine whether the PLS disease caused any significant difference in the chemical characteristic of the healthy leaves.

2.4. Hyperspectral Data Analysis. Reflectance values of 528 wavelengths distributed in four spectral regions (i.e., 350–399, 1300 nm–1400 nm, 1750 nm–1980 nm, and 2350 nm–2500 nm) were removed from the maize spectra due to noise and atmospheric water absorption [30]. Therefore, only 1623 out of 2151 wavelengths were used in the spectral analysis. One of the most notable difficulties in hyperspectral data processing is the hyperdimensionality of the data, which requires sufficient training samples to simplify the complexity of classification and prediction processes [31–34]. Practically, in most of the hyperspectral applications, the number of training samples (n) is limited with respect to the large number of hyperspectral bands (p) [34]. Therefore, variable selection methods have been widely used to select a compact variable without loss of predictive power of hyperspectral data [33]. In this study, a recently developed method, the guided regularized random forest (GRRF) [35, 36], was tested for hyperspectral band selection and classification.

2.5. Random Forest Classifier. Random forest is an ensemble learning technique developed by Breiman [37] to improve the classification and regression of trees (CART) by combining a large set of decision trees. The RF [37] grows multiple unpruned trees (*ntree*) on bootstrap samples of the original data. Each tree is grown on a bootstrap sample (2/3 of the original data known as “in-bag” data) taken with replacement from the original data. Trees are split to many nodes using random subsets of variables (*mtry*), and the default *mtry* value is the square root of the total number of variables. From the *mtry* selected variables, the variable that yields the highest decrease in impurity is chosen to split the samples at each node [37]. A tree is grown to its maximum size without pruning until the nodes are pure. That is, the nodes hold samples of the same class or contain certain number of samples. A prediction of the response variable (e.g., PLS early stage) is made by aggregating the prediction over all trees. In a classification application, a majority vote from all the trees in the ensemble determines the final prediction [37]. A more detailed description of RF can be found in Breiman [37] and Touw et al. (2012) among others. We used RF because RF naturally handles different scales, interactions, and nonlinearities among other numerical and categorical features [36].

2.6. Feature Selection via Guided Regularized Random Forest. Ordinary random forest has been widely used in hyperspectral data reduction. However, its preference to highly correlated predictor variable in identifying variables in high-dimensional spectral space has been identified as its major limitation [38, 39]. Moreover, while RF only provides insight into the importance of each variable in classification process, it does not automatically select the optimal number of variables that could yield the lowest error rate [40]. The new approach tested in this study was first developed and tested by Deng and Runger [36] in a small and simple dataset.

Random forest provides an internal measure of variable importance using the Gini index. The Gini index at node v is defined as

$$\text{Gini}(v) = \sum_{k=1}^k p_k^v (1 - p_k^v), \quad (1)$$

where p_k^v is the proportion of class k observations at node v . The Gini information gain (X_i, v) is the difference between the impurity at node v and the weighted average of impurities at each child node of v . The weights are proportional to the number of samples assigned to each child from the split at node v as defined in

$$\text{Gain}(X_i, v) = \text{Gain}(v) - w_L \text{Gini}(v_L) - w_R \text{Gini}(v_R), \quad (2)$$

where $\text{Gini}(v_L)$ and $\text{Gini}(v_R)$ are the Gini indices and w_L and w_R are the weights for the left and right child nodes.

To identify the key predictors, researchers have leveraged random forest for feature selection. For example, the recursive feature elimination (RFE) framework [41] and forward

variable selection [40, 42] build multiple random forests in order to obtain an optimal subset of features that best explains the phenomena of interest. However, these methods are computationally intensive. Consequently, Deng and Runger [35] proposed a regularization framework that can be applied to random forest (regularized random forest) and boosted trees (regularized boosted trees). The regularization framework avoids selecting a new feature for splitting the data in a tree node when that feature produces similar information to the feature already selected. The regularized framework builds one model that may considerably reduce the training time. Guided regularized random forest (GRRF) is an enhanced regularized algorithm that uses the importance scores from an ordinary random forest to guide the feature selection process [36].

The guided regularized random forest is built similarly to random forest, but uses a regularized version of information gain at each node v as in

$$\text{Gain}(X_i, v) = \begin{cases} \lambda_i \cdot \text{Gain}(X_i, v), & X_i \notin F \\ \text{Gain}(X_i, v), & X_i \in F, \end{cases} \quad (3)$$

where F is the feature set selected in the previous nodes and $\lambda_i \in (0,1]$ is called the coefficient of regularization for (X_i) and can be calculated as follows:

$$\lambda_i = (1 - \gamma) + \gamma \text{Imp}_i, \quad (4)$$

where $\text{Imp}_i \in [0,1]$ is the normalized importance score for X_i from an ordinary random forest built on the data set and $\gamma \in [0,1]$ is called the importance coefficient. For a feature X_i that does not have the maximum importance score 1, a larger γ leads to a smaller λ_i and, thus, a larger penalty on $\text{Gain}(X_i, v)$ when X_i has not been used in the nodes prior to node v .

Comparative studies have shown that GRRF is effective in selecting high-quality feature subsets while maintaining predictive accuracies [36].

2.7. Accuracy Assessment. The accuracy of RF classifier was assessed using the independent test dataset collected during the following growing season (2014) under the same PLS conditions. Out-of-bag error (OOB) [37], which provides an unbiased estimate of error of the RF, was used to estimate the misclassification. A confusion matrix was subsequently constructed to compute the overall accuracy (OA), user’s accuracy (UA), and producer’s accuracy (PA) as a criteria for evaluating the generalization ability (accuracy) of the RF classifiers [43]. OA is a ratio (%) between the number of correctly classified samples and the number of test samples, while UA represents the likelihood that a sample belongs to specific class and the classifier accurately assigns it such class. PA expresses the probability of a certain class being correctly recognized. Furthermore, kappa analysis that uses the k statistic was also calculated to determine if one error matrix is significantly different from another. The kappa coefficient provides a measure of the actual agreement between reference data and a random classifier. If the kappa coefficients are equal to one or close to one, then there is perfect agreement [44].

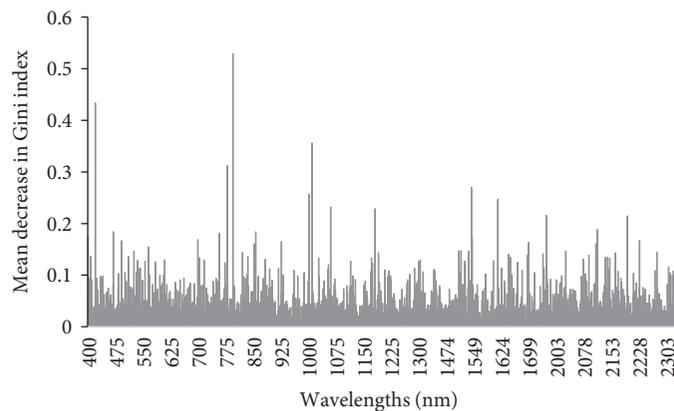


FIGURE 2: The importance scores of variables (wavelengths) as measured by ordinary RF using mean decrease in Gini index. The highest mean index is the most important variable.

3. Results and Discussion

3.1. Chemical Analysis of the Leaves. Since the effects of PLS on maize's biochemical characteristic have never been established, it was necessary to explore whether the leaves' chemical composition between the PLS early stage infestation and healthy leaves was significantly different. A *t*-test was used to determine if differences in the chemical composition (NPK) between the PLS early stage and healthy leaves were significant. Results that showed statistically significant difference in biochemical concentration was observed between healthy leaves and PLS-infested leaves for nitrogen (N), calcium (Ca), magnesium (Mg), copper (Cu), manganese (Mn), and phosphorus (P) concentration. PLS is known to affect translocation of nitrogen and reduce plant cycle and photosynthetic activity and accelerate leaf senescence, which reduces grain size and weight [7, 9]. However, this conclusion should be treated with caution as it is based on a general analysis to understand the effects of PLS on the spectral characteristic of the healthy maize leaves. More replicated biochemical analyses under controlled environment are required to better understand the effects of PLS on maize growth.

3.2. Variables Importance Measurement and Selection. The new variable selection procedure used in this study was able to reduce the high dimensionality of the hyperspectral data by eliminating irrelevant or redundant wavelengths. The importance of variables (wavelengths) in discriminating the PLS early stage and healthy maize leaves as determined by the ordinary RF classifier is shown in Figure 2. The most important wavebands are located in the red edge (670–780 nm) and near infrared (700–1200) portions of the electromagnetic spectrum. Very few bands are located in visible (400–500 nm) and the shortwave infrared (1900–2300 nm) sections of the electromagnetic spectrum. Some possible explanations for the selection of these wavelengths (within the visible and red edge regions) as the most important in discriminating PLS early stage are that these regions are more sensitive to vegetation biochemical properties such as canopy chlorophyll and nitrogen contents [45, 46]. Results

in this study have shown that PLS infestation leads to significant changes in the biochemical properties between early stages of PLS and healthy leaves. Changes in these leaf properties result in a shift in the red edge curve and increase the reflectance in the visible region [47], hence the selection of these regions. Other PLS symptoms that could be explained by the selection of the infrared wavelengths are that PLS accelerates leaf senescence and decreases grain size and weight [7]. Figure 2 indicates that many variables (wavelengths) share the same maximal Gini information gain at a node. Therefore, the importance scores from the ordinary RF were used to facilitate GRRF's selection of subset wavelengths that can better discriminate between the early symptoms of PLS and healthy maize leaves.

The GRRF was able to select 6 wavelengths using the ranking output of ordinary RF. The best subsets of wavelengths are located at 420 nm, 795 nm, 779 nm, 1543 nm, 1747 nm, and 1010 nm (Figure 3). These 6 spectral wavebands produced a minimal OOB error of 9.42% using the training dataset compared to 15.78% OOB error rate when the total number of wavelengths ($n = 1623$) was used. The subset selected by GRRF not only has fewer features compared with the entire variables, but also leads to lower OOB error on the training datasets. This could be explained by the fact that in the model-based analysis, the use of less important or redundant hyperspectral wavelengths leads to a decrease in the model accuracy, because the noise in the redundant data propagates through the classifier's performance [33]. The noise may not only decrease the performance of a weak classifier with a limited capability in handling the small variables, but may also affect the performance of more advanced classifiers such as random forest [35].

Results from the present experiment thus reaffirm previous findings [35, 36, 48] which show that the integrated approach between ordinary RF and GRRF is able to select small subsets of powerful variables in a high dimensionality data by an efficient computation procedure and achieve a competitive performance accuracy. Consequently, it is worth considering GRRF for variables selection in hyperspectral applications in the future. However, this assertion requires additional testing and comparisons with different variable

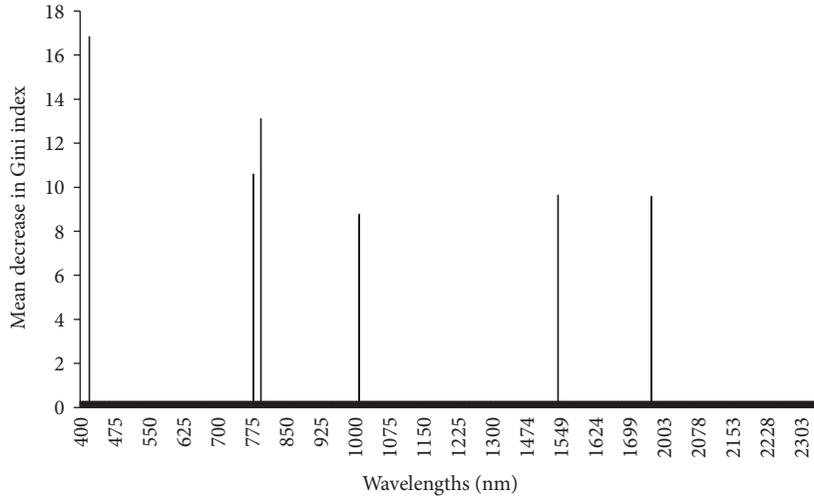


FIGURE 3: The importance scores of variables (wavelengths) as measured by ordinary RF using mean decrease in Gini index. The highest mean index is the most important variable.

TABLE 1: The confusion matrix showing overall accuracy (OA) and kappa for classifying the early stages of PLS (ES) and the healthy stage (HS) using the independent test dataset.

ES	Using 1623 wavelengths		ES	Using the selected 6 wavelengths	
	HS	Total		HS	Total
51	15	66	55	11	66
6	66	72	6	66	72
61	77	138	61	77	138
	AO = 81.88%			AO = 87.68%	
	Kappa = 0.64			Kappa = 0.75	

selection methods in different types of datasets before it is adopted as a substitute for data dimensionality reduction. If proven reliable, this integration could significantly save on time spent on complex computational procedure for hyperspectral data analysis.

3.3. Accuracy Assessment. The six wavelengths identified by GRRF were used as input variables into RF classifier to discriminate between the early stage symptoms (ES) of PLS and healthy stage (HS) of the maize leaves. Random forest parameters (*ntree* and *mtry*) were optimized using the training dataset (2013 dataset) and the model tested on an independent test dataset (2014 dataset). The results indicate that with the best setting of *ntree* (7500) and *mtry* (2) RF classifier yielded an overall accuracy of 81.88% using all the variables ($n = 1623$). The results were improved to 87.68% when the subset of selected ($n = 6$) variables was used (Table 1).

Based exclusively on the overall accuracy (AO), the use the wavelengths selected by GRRF ($n = 6$) proved to be more accurate (AO = 87.68) than the use of all wavelengths ($n = 1623$) for detecting the early stage of PLS.

Whereas GRRF can be used as a classification algorithm, we preferred to use the traditional RF as a classifier. We opted for this approach as GRRF is designed for feature selection,

and the trees are not constructed independently; therefore, the classification model may have a higher variance than the traditional random forest [36]. Moreover, the traditional random forest has been used successfully as a classifier in different types of dataset, particularly remotely sensed data [49–51]. Therefore, we considered the traditional RF as a classifier to provide an effective and efficient evaluation for the newly developed feature selection method in binary application. Results from this study indicate that GRRF was able to select high-quality feature subsets that significantly improved the classification performance of RF in detecting the early stage of PLS.

4. Conclusions

The objective of the present study therefore was to investigate the potential use of remotely sensed data in detecting the early stage of PLS in tropical maize. An extensive set of in situ hyperspectral measurements was collected over two different seasons, and an integrated new approach of GRRF and ordinary RF was investigated for variable selection and classification process. The relatively high overall accuracy obtained in this study indicates that the early stage of PLS in tropical maize can be detected using selected hyperspectral wavelengths.

From the results of the present study, we can conclude the following:

- (1) Phaeosphaeria leaf spot could be detected accurately at the early stage using hyperspectral data. This may provide insight on the choice of appropriate spatial and temporal management practices.
- (2) The new GRRF method produced high-quality feature subsets for the traditional RF classifier. Therefore, it could be considered as an effective and efficient feature selection tool for high data dimensionality reduction in hyperspectral applications.

Overall, our study presents a successful application of hyperspectral data, GRRF feature selection, and RF classifiers in detecting the early stage of PLS. This could be valuable in precision agriculture, specifically the management and control of the PLS. However, these results should be interpreted with caution as our study was based on analysing the spectral characteristics of PLS only. More studies are therefore needed in investigating the optimal spectral and spatial resolutions for PLS detection and upscaling these results to spaceborne or airborne sensor resolutions.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

All the authors were involved in the conceptualisation of the research idea. Elhadi Adam, John Odindi, Elfatih Abdel-Rahman, and John Derera designed the experiment. Elhadi Adam, John Odindi, and Elfatih Abdel-Rahman collected the data. Elhadi Adam and Houtao Deng analysed the data. Elhadi Adam drafted the manuscript. All authors advised on the content and edited the original and subsequent draft versions. Elhadi Adam prepared the manuscript for submission.

Acknowledgments

The authors are grateful to Mr Perushan Rajah and Miss Philisiwe Makhaye for assisting with the field data collection. Use of Cedara (KwaZulu Natal Department of Agriculture, South Africa) research facilities is sincerely acknowledged. Gratitude is extended to the R development core team for their very powerful open source packages for statistical analysis.

References

- [1] A. S. Langyintuo, W. Mwangi, A. O. Diallo, J. MacRobert, J. Dixon, and M. Bänziger, "Challenges of the maize seed industry in eastern and southern Africa: a compelling case for private-public intervention to promote growth," *Food Policy*, vol. 35, no. 4, pp. 323–331, 2010.
- [2] N. J. Walker and R. E. Schulze, "An assessment of sustainable maize production under different management and climate scenarios for smallholder agro-ecosystems in KwaZulu-Natal, South Africa," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 31, no. 15–16, pp. 995–1002, 2006.
- [3] J. K. A. Benhin, "South African crop farming and climate change: an economic assessment of impacts," *Global Environmental Change*, vol. 18, no. 4, pp. 666–678, 2008.
- [4] M. C. Tirado, R. Clarke, L. A. Jaykus, A. McQuatters-Gollop, and J. M. Frank, "Climate change and food safety: a review," *Food Research International*, vol. 43, no. 7, pp. 1745–1765, 2010.
- [5] S. Chakraborty, A. V. Tiedemann, and P. S. Teng, "Climate change: potential impact on plant diseases," *Environmental Pollution*, vol. 108, no. 3, pp. 317–326, 2000.
- [6] R. Gonçalves, J. Figueiredo, E. Pedro et al., "Etiology of Phaeosphaeria leaf spot disease of maize," *Journal of Plant Pathology*, vol. 95, no. 3, pp. 559–569, 2013.
- [7] J. Moreira, D. Bento, A. Souza, and C. Souza Jr., "QTL mapping for reaction to Phaeosphaeria leaf spot in a tropical maize population," *Theoretical and Applied Genetics*, vol. 119, no. 8, pp. 1361–1369, 2009.
- [8] M. Carson, "Yield loss potential of Phaeosphaeria leaf spot of maize caused by Phaeosphaeria maydis in the United States," *Plant Disease*, vol. 89, no. 9, pp. 986–988, 2005.
- [9] J. Sibiya, P. Tongoona, J. Derera, N. van Rij, and I. Makanda, "Combining ability analysis for Phaeosphaeria leaf spot resistance and grain yield in tropical advanced maize inbred lines," *Field Crops Research*, vol. 120, no. 1, pp. 86–93, 2011.
- [10] J. Derera, P. Tongoona, B. Vivek, N. van Rij, and M. Laing, "Gene action determining Phaeosphaeria leaf spot disease resistance in experimental maize hybrids," *South African Journal of Plant and Soil*, vol. 24, no. 3, p. 138, 2007.
- [11] M. Carson, C. Stuber, and M. Senior, "Quantitative trait loci conditioning resistance to phaeosphaeria leaf spot of maize caused by Phaeosphaeria maydis," *Plant Disease*, vol. 89, no. 6, pp. 571–574, 2005.
- [12] M. Carson, "Vulnerability of US maize germ plasm to Phaeosphaeria leaf spot," *Plant Disease*, vol. 83, no. 5, pp. 462–464, 1999.
- [13] L. D. Paccola-Meirelles, W. F. Meirelles, S. N. Parentoni, I. E. Marriel, A. S. Ferreira, and C. R. Casela, "Reaction of maize inbred lines to a bacterium, *Pantoea ananas*, isolated from Phaeosphaeria leaf spot lesions," *Crop Breeding and Applied Biotechnology, Londrina/PR*, vol. 2, no. 4, pp. 587–590, 2002.
- [14] L. Paccola-Meirelles, A. Ferreira, W. Meirelles, I. Marriel, and C. Casela, "Detection of a bacterium associated with a leaf spot disease of maize in Brazil," *Journal of Phytopathology*, vol. 149, no. 5, pp. 275–279, 2001.
- [15] S. Geerts, D. Raes, M. Garcia, C. D. Castillo, and W. Buytaert, "Agro-climatic suitability mapping for crop production in the Bolivian Altiplano: a case study for quinoa," *Agricultural and Forest Meteorology*, vol. 139, no. 3, pp. 399–412, 2006.
- [16] C. A. J. M. de Bie, "Comparative performance of agro-ecosystems," in *Natural Resources*, 232 pages, Wageningen, Wageningen, 2000.
- [17] D. J. Mulla, "Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps," *Biosystems Engineering*, vol. 114, no. 4, pp. 358–371, 2013.
- [18] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. ALRahamneh, "Fast and accurate detection and classification of plant diseases," *International Journal of Computer Applications*, vol. 17, no. 1, pp. 31–38, 2011.
- [19] Z.-y. Liu, J.-f. Huang, and R.-x. Tao, "Characterizing and estimating fungal disease severity of rice brown spot with hyperspectral reflectance data," *Rice Science*, vol. 15, no. 3, pp. 232–242, 2008.

- [20] M. S. Moran, Y. Inoue, and E. M. Barnes, "Opportunities and limitations for image-based remote sensing in precision crop management," *Remote Sensing of Environment*, vol. 61, no. 3, pp. 319–346, 1997.
- [21] P. J. Pinter, J. L. Hatfield, J. S. Schepers et al., "Remote sensing for crop management," *Photogrammetric Engineering and Remote Sensing*, vol. 69, no. 6, pp. 647–664, 2003.
- [22] R. Calderón, J. Navas-Cortés, and P. Zarco-Tejada, "Early detection and quantification of Verticillium wilt in Olive using hyperspectral and thermal imagery over large areas," *Remote Sensing*, vol. 7, no. 5, pp. 5584–5610, 2015.
- [23] K. Yu, G. Leufen, M. Hunsche, G. Noga, X. Chen, and G. Bareth, "Investigation of leaf diseases and estimation of chlorophyll concentration in seven Barley Varieties using fluorescence and hyperspectral indices," *Remote Sensing*, vol. 6, no. 1, pp. 64–86, 2013.
- [24] K. Usha and B. Singh, "Potential applications of remote sensing in horticulture—a review," *Scientia Horticulturae*, vol. 153, pp. 71–83, 2013.
- [25] M. Zhang, Z. Qin, X. Liu, and S. L. Ustin, "Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 4, no. 4, pp. 295–310, 2003.
- [26] Q. He, A. Berg, Y. Li, C. E. Vallejos, and R. Wu, "Mapping genes for plant structure, development and evolution: functional mapping meets ontology," *Trends in Genetics*, vol. 26, no. 1, pp. 39–46, 2010.
- [27] Analytical Spectral Device (ASD), *Handheld Spectroradiometer: User's Guide, Version 4.05*, Boulder, USA, 2005.
- [28] M. Bian, A. K. Skidmore, M. Schlerf et al., "Predicting foliar biochemistry of tea (*Camellia sinensis*) using reflectance spectra measured at powder, leaf and canopy levels," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 78, pp. 148–156, 2013.
- [29] M. P. Grisham, R. M. Johnson, and P. V. Zimba, "Detecting sugarcane yellow leaf virus infection in asymptomatic leaves with hyperspectral remote sensing and associated leaf pigment changes," *Journal of Virological Methods*, vol. 167, no. 2, pp. 140–145, 2010.
- [30] D. Zhao, K. R. Reddy, V. G. Kakani, J. J. Read, and S. Koti, "Canopy reflectance in cotton for growth assessment and lint yield prediction," *European Journal of Agronomy*, vol. 26, no. 3, pp. 335–344, 2007.
- [31] S. Li, H. Wu, D. Wan, and J. Zhu, "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine," *Knowledge-Based Systems*, vol. 24, no. 1, pp. 40–48, 2011.
- [32] M. Pal, "Margin-based feature selection for hyperspectral data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, no. 3, pp. 212–220, 2009.
- [33] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 7, pp. 793–802, 2004.
- [34] P. Hsu, "Feature extraction of hyperspectral images using wavelet and matching pursuit," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 2, pp. 78–92, 2007.
- [35] H. Deng and G. Runger, "Feature selection via regularized trees," *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pp. 1–8, IEEE, 2012.
- [36] H. Deng and G. Runger, "Gene selection with guided regularized random forest," *Pattern Recognition*, vol. 46, no. 12, pp. 3483–3489, 2013.
- [37] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [38] C. Adjorlolo, O. Mutanga, M. A. Cho, and R. Ismail, "Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 21, pp. 535–544, 2013.
- [39] C. Strobl, A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," *BMC Bioinformatics*, vol. 9, no. 1, 307 pages, 2008.
- [40] E. M. Adam, O. Mutanga, D. Rugege, and R. Ismail, "Discriminating the papyrus vegetation (*Cyperus papyrus* L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP," *International Journal of Remote Sensing*, vol. 33, no. 2, pp. 552–569, 2012.
- [41] R. Ismail and O. Mutanga, "Discriminating the early stages of Sirex noctilio infestation using random forest and short-wave infrared (SWIR) wavelengths," *International Journal of Remote Sensing*, vol. 32, pp. 4249–4266, 2011.
- [42] K. Mansour, O. Mutanga, T. Everson, and E. Adam, "Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 70, no. 0, pp. 56–65, 2012.
- [43] P. Mather and B. Tso, *Classification Methods for Remotely Sensed Data*, CRC press/Taylor & Francis Group, London, Second edition, 2003.
- [44] R. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, 2008.
- [45] M. A. Cho, P. Debba, O. Mutanga, N. Dudeni-Tlhone, T. Magadla, and S. A. Khuluse, "Potential utility of the spectral red-edge region of SumbandilaSat imagery for assessing indigenous forest structure and health," *International Journal of Applied Earth Observation and Geoinformation*, vol. 16, pp. 85–93, 2012.
- [46] R. C. Muchow, M. J. Robertson, A. W. Wood, and B. A. Keating, "Effect of nitrogen on the time-course of sucrose accumulation in sugarcane," *Field Crops Research*, vol. 47, no. 2, pp. 143–153, 1996.
- [47] M. A. Cho and A. K. Skidmore, "A new technique for extracting the red edge position from hyperspectral data: the linear interpolation method," *Remote Sensing of Environment*, vol. 101, no. 2, pp. 181–193, 2006.
- [48] H. Deng, *Guided Random Forest in the RRF Package*, arXiv preprint arXiv:1306.0237, 2013.
- [49] G. P. Petropoulos, K. Arvanitis, and N. Sigrimis, "Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3800–3809, 2012.
- [50] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.
- [51] A. Stumpf and N. Kerle, "Combining random forests and object-oriented analysis for landslide mapping from very high resolution imagery," *Procedia Environmental Sciences*, vol. 3, no. 1, pp. 123–129, 2011.

