Applied Geomatics (2021) 13:389–400
https://doi.org/10.1007/s12518-021-00357-4

ORIGINAL PAPER

Estimating maize lethal necrosis (MLN) severity in Kenya using multispectral high-resolution data

Kyalo Richard[1] · Elfatih M. Abdel-Rahman[1,2] · Sevgan Subramanian[1] · Johnson O. Nyasani[1,3] · Michael Thiel[4] · Hossein J. Jozani[5] · Christian Borgemeister[6] · Bester T. Mudereri[1,7] · Tobias Landmann[1,8]

Received: 5 June 2020 / Accepted: 7 January 2021 / Published online: 21 January 2021

Abstract

Maize lethal necrosis (MLN) is a severe disease in maize that significantly reduces yields by up to 90% in maize-growing regions such as Kenya and other countries in Africa. The disease causes chlorotic mottling of leaves and severe stunting which leads to plant death. The spread of MLN in the maize-growing regions of Kenya has intensified since the first outbreak was reported in September 2011. In this study, the RapidEye (5 m) imagery was combined with field-based data of MLN severity to map three MLN severity levels in Bomet County, Kenya. Two RapidEye images were acquired during maize stem elongation and inflorescence stages, respectively, and thirty spectral indices for each RapidEye time step were computed. A two-step random forest (RF) classification algorithm was used to firstly create a maize field mask and to predict the MLN severity levels (mild, moderate, and high). The RF algorithm yielded an overall accuracy of 91.0%, representing high model performance in predicting the MLN severity levels in a complex cropping system. The normalized difference red edge index (NDRE) was highly sensitive to MLN detection and demonstrated the ability to detect MLN-caused crop stress earlier than the normalized difference vegetation index (NDVI) and the green normalized difference vegetation index (GNDVI). These results confirm the potential of the RapidEye sensor and machine learning to detect crop disease infestation rates and for use in MLN monitoring in fragmented agro-ecological landscapes.

Keywords Africa · Crop disease · Food security · Severity level

✉ Elfatih M. Abdel-Rahman
eabdel-rahman@icipe.org

[1] International Centre of Insect Physiology and Ecology (icipe), P.O. Box 30772, Nairobi 00100, Kenya

[2] Department of Agronomy, Faculty of Agriculture, University of Khartoum, 13314 Khartoum North, Sudan

[3] Crop Health Unit, Kenya Agricultural and Livestock Research Organization, Embu Research Centre, P.O. Box 27, Embu 60100, Kenya

[4] Department of Remote Sensing, University of Würzburg, Oswald-Külpe-Weg 86, 97074 Würzburg, Germany

[5] Faculty of Agricultural Sciences in the Tropics & Subtropics, University Hohenheim, Garbenstraße 13, 70599 Stuttgart, Germany

[6] Center for Development Research (ZEF), Department of Ecology and Natural Resources Management, University of Bonn, Walter-Flex-Str. 3, 53113 Bonn, Germany

[7] Department of Animal and Wildlife Science, Midlands State University, P. Bag, 9055 Gweru, Zimbabwe

[8] Remote Sensing Solutions Gmbh, Dingolfinger Str. 9, 81673 Munich, Germany

Introduction

Agriculture is the primary source of livelihood for most developing countries in Africa (Worldbank 2018). In Kenya, agriculture contributes up to 65% of the total labor force and a third to the Kenyan gross domestic product (GDP) (Omiti et al. 2009). The population of Kenya is rapidly growing and is expected to increase from 46 million to 95 million people by 2050 (Worldbank 2018). This inevitably increases food demand which has profound implications on agricultural productivity to ensure food security for the country (FAO et al. 2019).

Maize (*Zea mays* L.) is the main staple food crop in most countries in sub-Saharan Africa, covering over 25 million ha of smallholder farmers in the region and is the main staple crop for over 85% of the population in Kenya (FAO et al. 2019). In the last decade, Kenya has realized an approximate 12% drop in maize production attributable to a wide range of abiotic and biotic risk factors (AGRA 2017). The risk factors include rainfall and temperature variability, pests, and diseases that are likely to intensify under the anticipated climate change (Nhamo et al. 2019). Therefore, effective disease detection capacity and empirical control and mitigation strategies are necessary and timely to curb the surge in the loss of crop productivity in the country (Myers et al. 2017).

Among the many pests and diseases affecting maize production in Kenya is the maize lethal necrosis (MLN) which has emerged as one of the most severe diseases threatening livelihoods and income across sub-Saharan Africa (Hilker et al. 2017; Osunga et al. 2017). MLN was firstly reported in Kenya in September 2011 in the Longisa Division within Bomet County of Kenya (Wangai et al. 2012; Adams et al. 2013). In the same year, MLN spread rapidly to other major maize-growing regions along the Rift Valley and the western part of Kenya towards Lake Victoria (Osunga et al. 2017). In the following year, Kenya reported a major drop in maize harvest instigated by MLN. Thus, since 2012, the disease has continued to spread rapidly into other countries in East Africa such as Tanzania, Uganda, Rwanda, and Ethiopia, leading to a serious reduction in maize production across the region prompting an urgent need to curb and control the disease (Adams et al. 2013; Mahuku et al. 2015).

Globally, numerous MLN control and management methods have been proposed. For instance, farmers are advised to uproot affected plants during the early growing stage to ensure that the disease does not spread. Another mitigation measure is crop rotation, but farmers are often reluctant to shift from the continuous planting of maize because of the demand and income derived from the crop. Thus, efforts to control the disease have been less fruitful (Deressa and Demissie 2017).

The detailed information on MLN disease severity, incidence, and the related effects on quality and quantity of maize production are important prerequisites for improved disease management (Mahlein 2016). MLN-infected maize plants show a range of symptoms such as yellowing and mottling of leaves leading to premature plant death, tasseling failure resulting in warped maize cobs with little to no seeds (Ochieng et al. 2016; Deressa and Demissie 2017). Thus, visual estimates of disease symptoms in the field can determine disease severity and incidence (Bock et al. 2010). However, such an approach is expensive, time-consuming, and often insufficiently accurate because of human bias and insufficient and timely coverage of the affected areas (Benson et al. 2015). Consequently, there is a growing demand for more precise and automated methods of MLN disease monitoring to mitigate disease outbreaks by enabling the timely adoption of relevant management practices (Geerts et al. 2006).

Remote sensing and geospatial techniques have demonstrated high potential in detecting the presence and monitoring of the spread of agricultural pests and diseases including MLN (Osunga et al. 2017; Jozani et al. 2020). This is because the induced crop physiological stress and biophysical changes on the infested plant leaves can alter the reflectance spectra of plants that are detected by remote sensing sensors (Fang and Ramasamy 2015). Using plant leaves and canopy spectral signature can complement field-based protocols in distinguishing between healthy and various levels of damage of infested plants (Albayrak 2008; Mudereri et al. 2019b, 2020a). Moreover, remote sensing technologies can reveal the spatial and temporal distribution of pests and diseases over large areas at a relatively low cost.

Early research has demonstrated the utility and capability of using remotely sensed data to yield accurate results in the wide-area mapping of crop diseases. For instance, Song et al. (2017) evaluated Sentinel-2 satellite imagery for mapping cotton root rot, demonstrating that the technique can be used for precise disease identification if the image set is taken during the optimum root rot discrimination period. Zhang et al. (2016) used two-date multispectral satellite imagery for accurately mapping damage caused by fall armyworm (*Spodoptera frugiperda*) in maize at a regional scale, while Franke and Menz (2007) evaluated high-resolution QuickBird satellite multispectral imagery for detecting powdery mildew (*Blumeria graminis*) and leaf rust (*Puccinia recondita*) in winter wheat. These studies demonstrated that multispectral images are generally suitable to detect intra-field heterogeneities in plant vigor, particularly at late stages of fungal infections, but are only moderately appropriate for distinguishing early infection levels (Dhau et al. 2018a, 2019; Sibanda et al. 2019). To the best of our knowledge, research on MLN mapping has emphasized on the use of inferred species distribution modeling using bioclimatic variables (Osunga et al. 2017) while the study of Jozani et al. (2020) has attempted to directly detect the disease or its infestation level within the complex smallholder farmers' maize fields in sub-Saharan Africa using multispectral data. Notwithstanding, Jozani et al. (2020) detected only the highly severe MLN-infected maize fields and did not look at the possibility of mapping other MLN severity scores (e.g., mild and moderate). Detecting crop diseases early enough in the growing season before highly severe infections are established is of profound importance for timely disease monitoring and management.

Therefore, in this present study, we evaluated the potential of space-borne RapidEye multi-temporal data and an advanced random forest (RF) classification technique for mapping three MLN severity levels, viz., mildly, moderately, and highly severe in a complex, dynamic, and heterogeneous landscape, typical for rural sub-Saharan Africa. This information is important for an improved understanding of the progression of the disease over a large area and for the formulation and implementation of site-specific strategies for effective control of MLN.

Study area

The study was conducted in Bomet and Nyamira Counties located 300 km northwest of Nairobi, Kenya.

The study sites lie between 34.97° E to 35.06° E and − 0.76°S to − 0.83°S (Fig. 1) with an elevation range between 1800 and 3000 m above sea level. Bomet falls in a semi-humid climatic zone with a mean monthly temperature of 18 °C and a bimodal annual rainfall ranging between 1100 and 1500 mm (Jaetzold and Schmidt 1982). The climate is suitable for growing a wide range of crops; however, maize and tea are the most dominant crops in the region, with the majority of farmers practicing a maize-based mono-cropping system, especially in the southern part of Bomet County (Abdel-rahman et al. 2017).

Methodology

Figure 2 summarizes the methodological approach for mapping severity levels of MLN. A two-step hierarchical RF classification using bi-temporal RapidEye was employed. First, a land use/land cover (LULC) classification map was generated to delineate cropland from other LULC classes. We used the extracted maize crop mask from the first step to classify different MLN severity levels in maize fields (viz., mild, moderately, and highly infected maize plants).

Field data collection

Field data collection was conducted to identify different LULC classes from the study area and to measure the MLN severity levels within maize fields. Stratified random sampling was followed to collect both the LULC and the MLN severity reference data. A handheld global positioning system (GPS) device with an error of ± 3 m was used to locate the reference control points. Once a field was identified, we delineated the field boundaries (polygon) within a minimum area of 10 × 10 m. To avoid the edge effect, we collected the polygon data 2 m away from the edge of each field. To mitigate the effect of soil background on the crop spectral features, we only sampled the field crops that were about 3 weeks old at the first image acquisition date. The reference data for both LULC classification and MLN severity mapping were randomly divided into 70% training and 30% validation sets. The training set was used to train the RF classifier, while the validation dataset was used to evaluate the accuracy.

Disease severity scores were determined using an expert knowledge approach based on Kusia (2014) and Mwatuni et al. (2020), whereby we conducted frequent field (4 × 1-week interval) visits to MLN-affected farms to assess the disease damage levels. Specifically, for each sampled farm, the maize plants were grouped in specific severity levels based on damage levels and visual inspection. The severity was rated

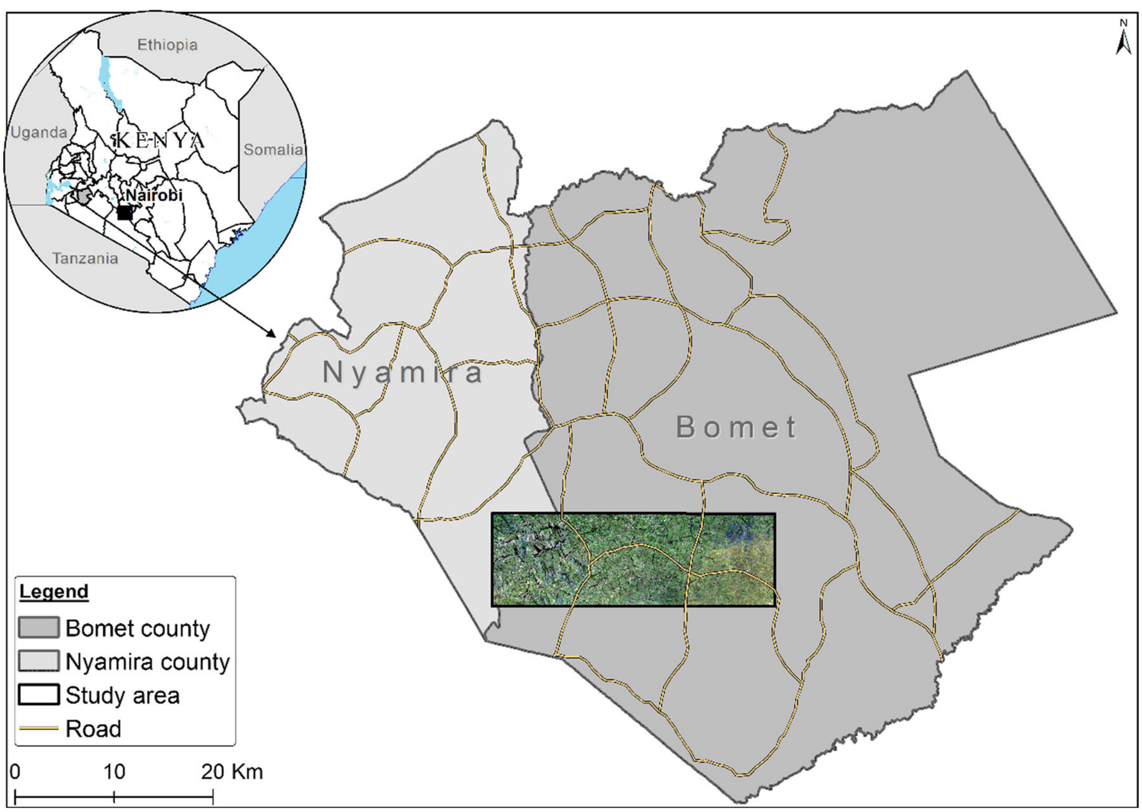

Fig. 1 Location of the study area in the Bomet and Nyamira Counties of Kenya

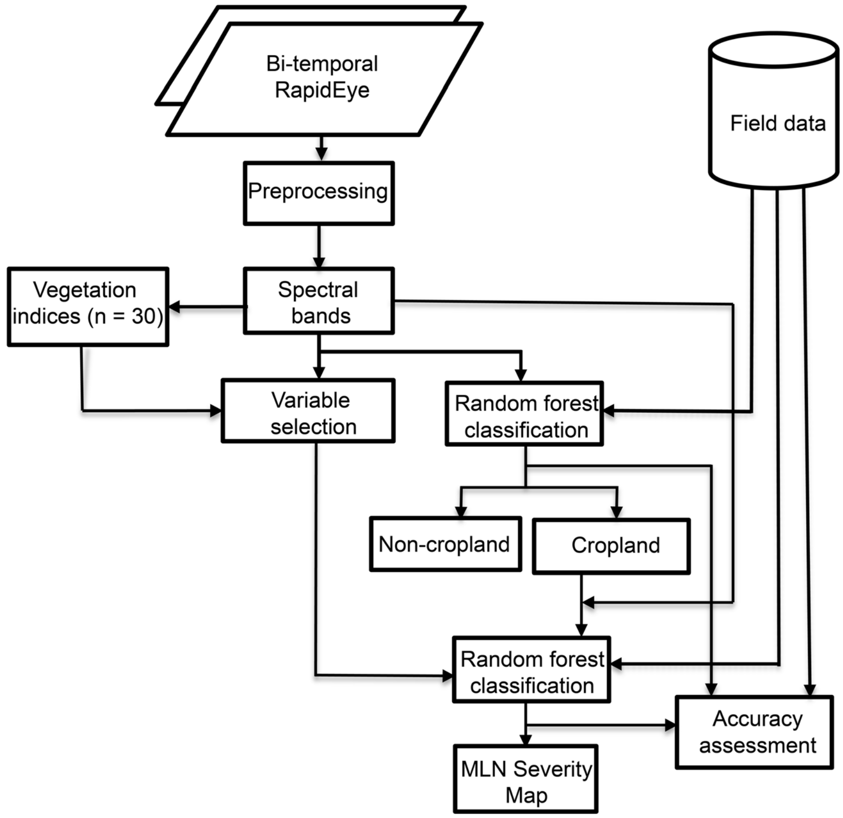

Fig. 2 Flowchart of the two-step random forest classification for mapping maize lethal necrosis (MLN) severity levels

using a scale of 0–5 as described by Paul and Munkvold (2004). The six scales were 0 (no disease), 1 (10–20% leaf area affected by the disease), 2 (21–40% leaf area affected by the disease), 3 (41–60% leaf area affected by the disease), 4 (61–80% leaf area affected by the disease), and 5 (81–100% leaf area affected by the disease). For consistency, MLN damage levels were classified as mild (< 20%), moderate (20–80%), and high (> 80%) based on a suggestion by Nutter and Schultz (1995).

RapidEye data preprocessing

Two RapidEye images of the 9th of December 2014 (RE1) and the 27th of January 2015 (RE2) were used. This period coincided with the maize stem elongation and the inflorescence stages, respectively, and the field data collection period. RapidEye is a commercial optical earth observation mission that consists of a constellation of five satellites with 5-m resolution and a swath width of 77 km with a revisit cycle of 5.5 days at nadir (RapidEye 2018). The RapidEye imagery is provided in five optical bands in the 400–850 nm range of the electromagnetic spectrum (Chabalala et al. 2020). The images used in this study were delivered as level 3A orthorectified products in the form of 25 × 25 km tiles georeferenced to the universal transverse mercator (UTM) projection.

Atmospheric correction was performed for each RapidEye tile independently using the atmospheric-topographic correction (ATCOR 3) software (Guanter et al. 2009). This application provides a sensor-specific atmospheric database of look-up-tables (LUT) which contain results of pre-calculated radiative transfer calculations based on the moderate resolution atmospheric transmission (MODTRAN-5) model (Berk et al. 2008). All images were co-registered (image-to-image) to ensure the alignment of the corresponding pixels. Subsequently, the RapidEye tiles were mosaiced into a single image file for each acquisition date. For each RapidEye image, 30 spectral vegetation indices (SVIs) were computed and combined with the individual bands (blue, green, red, red edge, and near infrared) as input predictor variables to improve the MLN severity level classification accuracy. Readers are referred to Kyalo et al. (2017) for the full list of the 30 SVIs.

Random forest algorithm

We used the RF machine learning classifier to predict the LULC classes, used for generating the crop mask, and in mapping the MLN severity levels (Breiman 2001). RF was chosen as the preferred classification method since it has been proven to be robust to outliers and noise and consistently demonstrated capability to handle high-dimensional datasets without suffering from overfitting (Chemura et al. 2017b; Hengl et al.

2018; Mudereri et al. 2019a). RF builds an ensemble of individual decision trees from which the final prediction is based using majority voting criteria. Each decision tree is trained using a bootstrap sample consisting of two-thirds of the training data drawn with replacement, and the remaining one-third of the data, which is not included in the bootstrapped training sample, is used to test the classification and estimate the out-of-bag (OOB) error (Breiman 2001; Chemura et al. 2017a).

RF uses two user-defined parameters, the number of trees (*ntree*), and the number of variables used to split the nodes (*mtry*). The default *ntree* is 500, while the default value for *mtry* is the square root of the total number of explanatory variables used in the study. To improve the classification accuracy, the two RF parameters were initially optimized based on the OOB error rate.

Variable selection and optimization

RF measures the importance of each predictive variable using the mean decrease in accuracy that is calculated using the OOB sample data (Georganos et al. 2018). However, the challenge was to select the least number of predictors that offer the best predictive power. In this regard, a backward feature elimination method (BFE) integrated with RF regression as part of the evaluation process was implemented (Mutanga et al. 2012). The BFE uses the ranking to identify the sequence in which to discard the least important predictors from the input datasets. The method starts with all the variables and then progressively eliminates the variable with the least contribution from the list. For each iteration, the model is optimized by selecting the best *mtry* and *ntree*, using a grid search and a 10-fold cross-validation method (Huang and Boutros 2016). The least contributing variable is eliminated, and the OOB error is calculated. The subset of the least number of variables with the smallest RMSE is then selected for the final classification model.

Accuracy assessment

The classification accuracy of the RF classifier was assessed using an independent set of field data (30%). The overall accuracy (OA) and the F1 score values were computed from the confusion matrices to evaluate the accuracy of generated classes. Also, the class-specific producer's accuracy (PA) and user's accuracy (UA) were calculated to evaluate the generalization ability of the RF classifier (Congalton 2001). A confusion matrix provides information on the correct predictions by comparing the classified map with ground information collected from the field. OA refers to the ratio of the correctly classified pixel to all pixels considered in the model evaluation. The F1 score is a per-category measure that corresponds to the harmonic mean of the UA and PA (Kyalo et al. 2017). PA refers to the error of omission which expresses the probability of a certain class to be correctly recognized, while UA is the error of commission which represents the likelihood that a sample belongs to a specific class and the classifier accurately assigns it to this class. Kappa statistics were also calculated to compare the significance between different error matrices generated from the generated classification results (McHugh 2012). The Kappa coefficient measures the actual agreement between the reference data and a random classifier with a value close to one, signifying perfect agreement. To reduce the common salt-and-pepper noise that is associated with high spatial resolution classification maps, a 3 × 3 cell majority filter was applied (Fierens and Rosin 1994; Su 2016). This approach replaces secluded cells with the class that matches a 3 × 3 cell matrix. Each filtered classified map was finally tested for accuracy.

Results

Random forest optimization

RF parameters (*ntree* and *mtry*) were optimized for the two-step classification for the different data sets using the grid search technique with tenfold cross-validation. The *ntree* value of 500 and *mtry* value of 3 settings yielded the least OOB error for the LULC classification. Also, the *ntree* value of 1000 and *mtry* of 5 yielded the best OOB error (4.8%) for the MLN severity mapping (Fig. 3).

Crop masking

Six major LULC classes were identified based on field observations made within the study area. Table 1 presents a summary of the results (OA and F1 score) when using 30% as evaluation data. The results revealed that the use of the RapidEye spectral bands gave an OA of 72.3% and 74.8% for a single classification of RE1 and RE2, respectively. The combination of the two RE1 and RE2 spectral bands improved the OA to 80.6%. Also, the F1 score for each class was generally above 0.80, except for soil class for the combination of RE1 and RE2.

Figure 4 shows the LULC map generated from the optimal combination of the two RE1 and RE2 images revealing that cropland and grassland are the major classes in the study site, with few tea plantations on the northern side of the study area. However, there was slight confusion between cropland with natural vegetation and forest resulting from the presence of big trees and pockets of bushes inside the cropland as observed from the field.

Table 2 represents the confusion matrix for the per-pixel evaluation for the LULC classification using RF. In general, all LULC classes achieved > 90% UA, except for natural vegetation which had 79.89% due to spectral confusion with

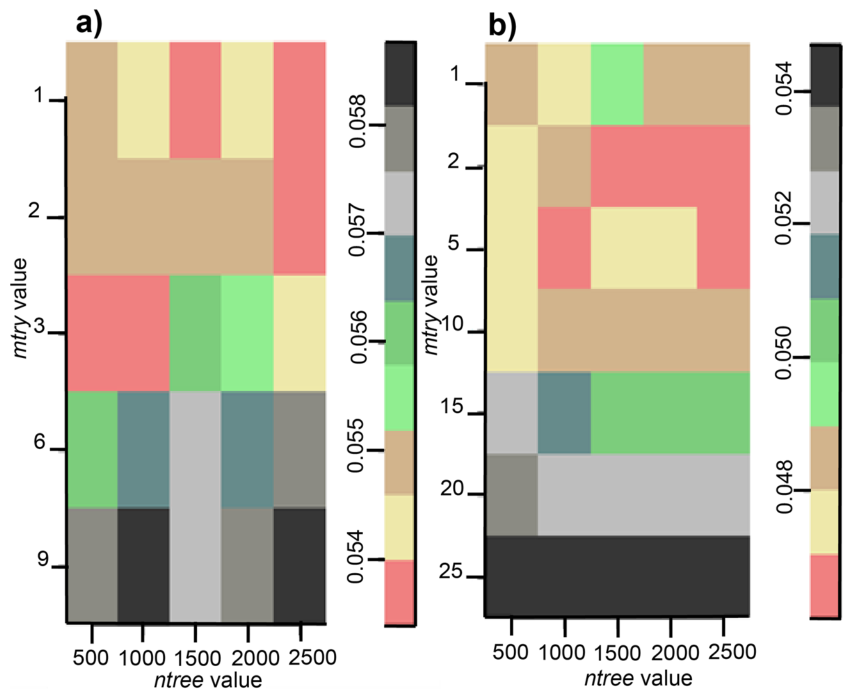

Fig. 3 Results of the random forest optimization grid for the land use/land cover classification result (a) and the maize lethal necrosis severity mapping result (b). The internal out-of-bag error rate calculated using the tenfold cross-validation and the training data. The color grids show the out-of-bag (OOB) error rate.

cropland. All LULC classes achieved > 90% PA except for cropland and natural vegetation classes which achieved 83.23% and 86.34%, respectively. Consequently, the UA was generally > 90% for all classes except the natural vegetation class which had 79.89%. The latter can be attributed to the observed confusion between the "cropland" and "natural vegetation" classes as shown in the confusion matrix.

Variable selection for maize lethal necrosis severity classification

The progressive removal of the least important predictor variables resulted in the selection of seven spectral variables (indices and/or bands) which gave the least OOB error as shown in Fig. 5. The model with a fewer number of predicted variables was compared with the model of all the predictor variable dataset.

Four and three spectral variables, respectively, were selected as important variables from the two RE1 and RE2 images, captured during the maize stem elongation and inflorescence development stages, respectively (Table 3). Only one spectral band (band 5) for RE2 was selected among the most important variables. Also, Chlorophyll Index red edge (ChlRed-edge) vegetation indices calculated from both acquisitions were selected among the most significant predictor variables too. Besides, the variable importance technique in the RF was used to determine the influence of each spectral variable selected on the mapping accuracy. ChlRed-edge vegetation index from RE1 was the most important variable with a mean decrease accuracy of 0.22% followed by ChlRed-edge vegetation index from RE2 with a mean decrease accuracy of 0.19%. Band 5 and the transformed soil-adjusted vegetation index red edge (TSAVI) from RE2 were the third and fourth most significant variables, respectively (Table 3).

RE1 and RE2 are RapidEye 1 and RapidEye 2 images, respectively

Maize lethal necrosis severity classification

Table 4 presents the accuracy assessment error matrix for the classification map generated by the RF's most important spectral variables to map three MLN severity classes (i.e., mild,

Table 1 Overall and class-wise accuracies for land use/land cover mapping using 30% of test data

		Class-wise accuracies—F1 score					
Image	Overall accuracy (%)	Cropland	Forest	Grassland	Natural vegetation	Soil	Tea
RE1	72.34	0.69	0.81	0.69	0.74	0.83	0.81
RE2	74.80	0.73	0.87	0.84	0.77	0.84	0.70
RE1 + RE2	80.63	0.82	0.92	0.89	0.85	0.78	0.82

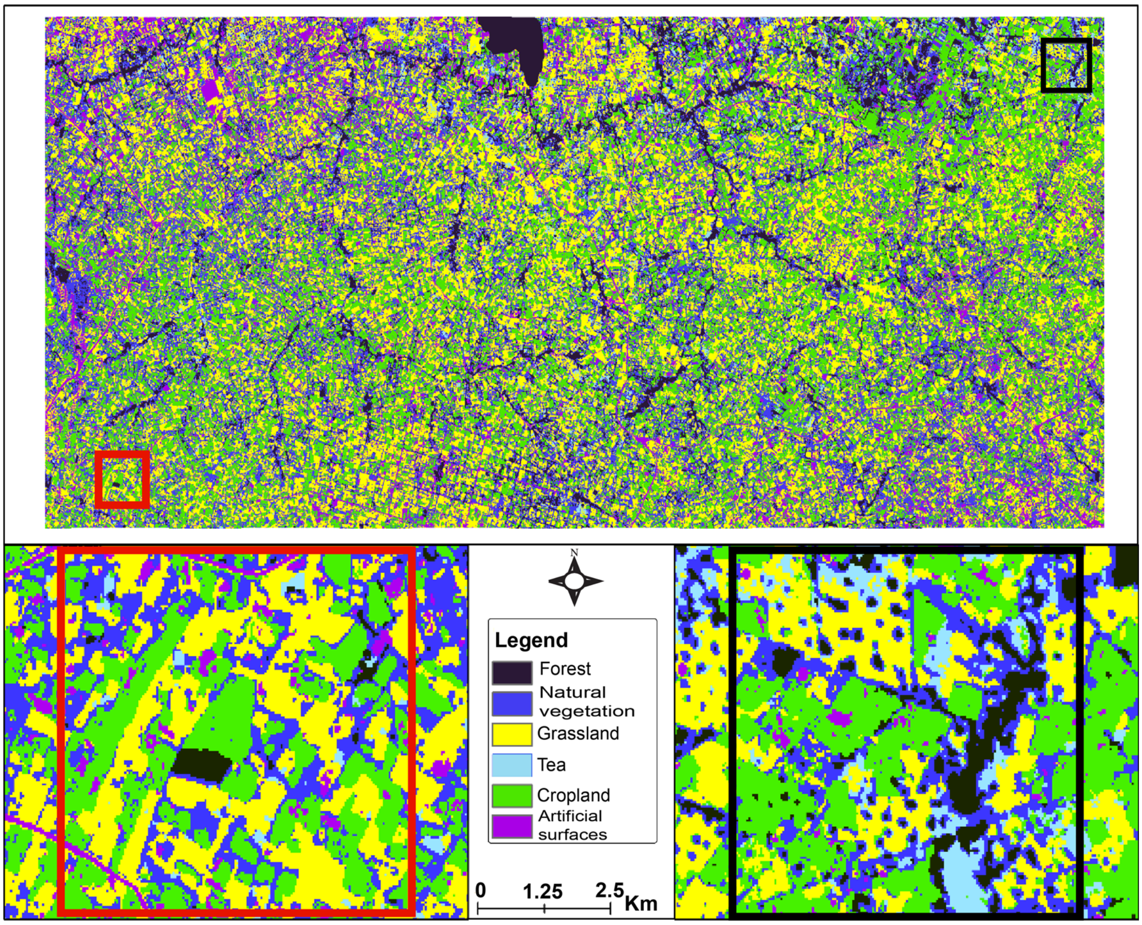

Fig. 4 Land use/land cover map obtained using a random forest classifier and the two RapidEye images (RE1 and RE2)

moderate, and high). The OA for mapping MLN disease was 73.33%. The PA, which indicates the probability of actual areas being correctly classified, was 60.48% for the mild MLN severity, 60.95% for the moderate, and 98.57% for the high severity classes. The UA attained was 66.84% for the mild, 61.24% for the moderate, and 89.61% for the high severity classes.

To improve the MLN disease mapping accuracy, we combined the mild and moderate severity classes which depicted high confusion because of similar spectral characteristics for most of the sampled farms. This improved the OA from 73.33% to 90.18% and Kappa from 0.60 to 0.92. The final thematic MLN severity map for the two severity classes (mild and high) produced via the RF algorithm is shown in Fig. 6. The red color represents maize farms with high severity, while the blue color depicts the mildly infected fields. As shown in the zoomed portion of the map, some of the

Table 2 Confusion matrix for land use/land cover classification using random forest classification with two early-season RapidEye images (RE1 and RE2) and one late-season Landsat image. UA is user's accuracy and PA is producer's accuracy

Class	Cropland	Forest	Grassland	Natural vegetation	Soil	Tea	Total	PA (%)
Cropland	268	0	1	48	0	5	322	83.2
Forest	0	303	0	7	0	12	322	94.0
Grassland	2	0	304	4	12	0	322	94.5
Natural vegetation	18	12	8	278	0	6	322	86.3
soil	9	0	5	3	305	0	322	94.7
Tea	0	13	0	8	0	301	322	93.5
Total	297	328	318	348	317	324	1932	
UA (%)	90.2	92.4	95.6	79.9	96.2	92.9		

*Overall accuracy = 91.0% and Kappa = 0.89

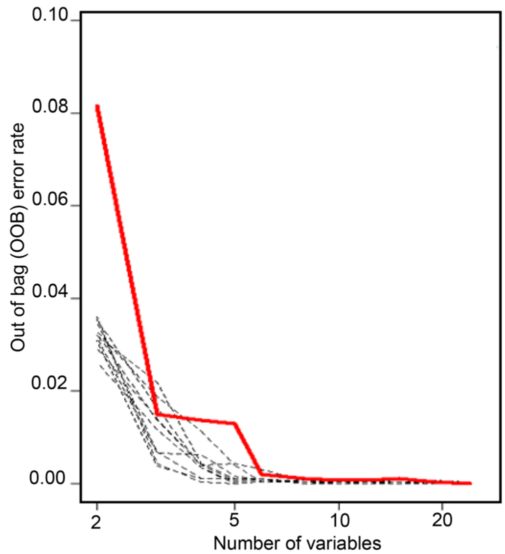

Fig. 5 The optimal number of predictor variables selected based on the random forest backward feature elimination search function using out-of-bag (OOB) error

maize fields harbored both mildly and highly severe MLN-affected plants which agree with our field observations (Fig. 6).

Discussion

This study explored the usefulness of bi-temporal RapidEye imagery and a RF classification tool for mapping the MLN severity levels in heterogeneous agro-ecological landscapes in Kenya. A two-step optimized RF classification was used to extract a crop mask from LULC classification and finally to generate an MLN severity map for the Bomet County and southern part of Nyamira County, a major maize-growing area in Kenya heavily affected by the disease.

Utilization of the two RE1 and RE2 images acquired for the study area during the maize early growing stages did not yield promising results in delineating cropland from other LULC classes. This could be attributed to the late plowing of some fields as observed during field visits. Thus, the use of early-season RE1 image alone captured insignificant portions of the phenological development of the maize plants, hence the failure to produce an accurate crop mask (Forkuor et al. 2014). Subsequently, due to the high costs of RapidEye imagery, the study managed two acquisitions during the maize stem elongation and inflorescence stages, respectively. Combining the two acquired RE1 and RE2 images improved the classification accuracy by providing additional information on late cultivated fields which significantly improved the accuracy for extracting the crop mask from LULC classification from 80.63% to 91.05%. Similarly, Crnojevic et al. (2014) used freely available Landsat-8 data with a single RapidEye image to improve the classification of small agricultural fields in northern Serbia.

The high OA of our LULC classification map supports the growing evidence that RF is a reliable classifier for heterogeneous landscapes (Nguyen et al. 2018). For instance, our results revealed a good separability for all the LULC classes apart from the slight confusion between cropland and natural vegetation classes. These results demonstrate the effectiveness of RF classifier to distinguish cropland from other LULC classes in a highly fragmented landscape. The observed overlaps between cropland and natural vegetation classes are well known and can be attributed to the spectral similarity among the vegetation and cropland caused by the presence of small pockets of shrubs within the agricultural land (Forkuor et al. 2015). Most farmers in our study area maintain fruit trees such

Table 4 Random forest classification confusion matrix for three maize lethal necrosis (MLN) severity classes (mild, moderate, and high) using the seven most important spectral variables with a 30% test dataset. UA is user's accuracy and PA is producer's accuracy

MLN class	Mild	Moderate	High	Total	PA (%)
Mild	127	78	5	210	60.48
Moderate	63	128	19	210	60.95
High	0	3	207	210	98.57
Total	190	209	231	630	
UA (%)	66.84	61.24	89.61		

*Overall accuracy = 73.33% and Kappa = 0.60

Table 3 Spectral variables selected as the most important predictor variables for mapping maize lethal necrosis severity levels using the random forest backward feature elimination procedure

Acquisition	Spectral variable	Abbreviation	Mean decrease accuracy (%)
RE1	Chlorophyll Index red edge	ChlRed-edge	0.22
RE2	Chlorophyll Index red edge	ChlRed-edge	0.19
RE2	RE band 5 (red edge)	Band 5	0.09
RE2	Transformed soil-adjusted vegetation index red edge	TSAVI	0.091
RE1	Green normalized difference vegetation index	GDVI	0.09
RE1	Normalized difference vegetation index	NDVI	0.08
RE1	Normalized difference red edge	NDRE	0.08

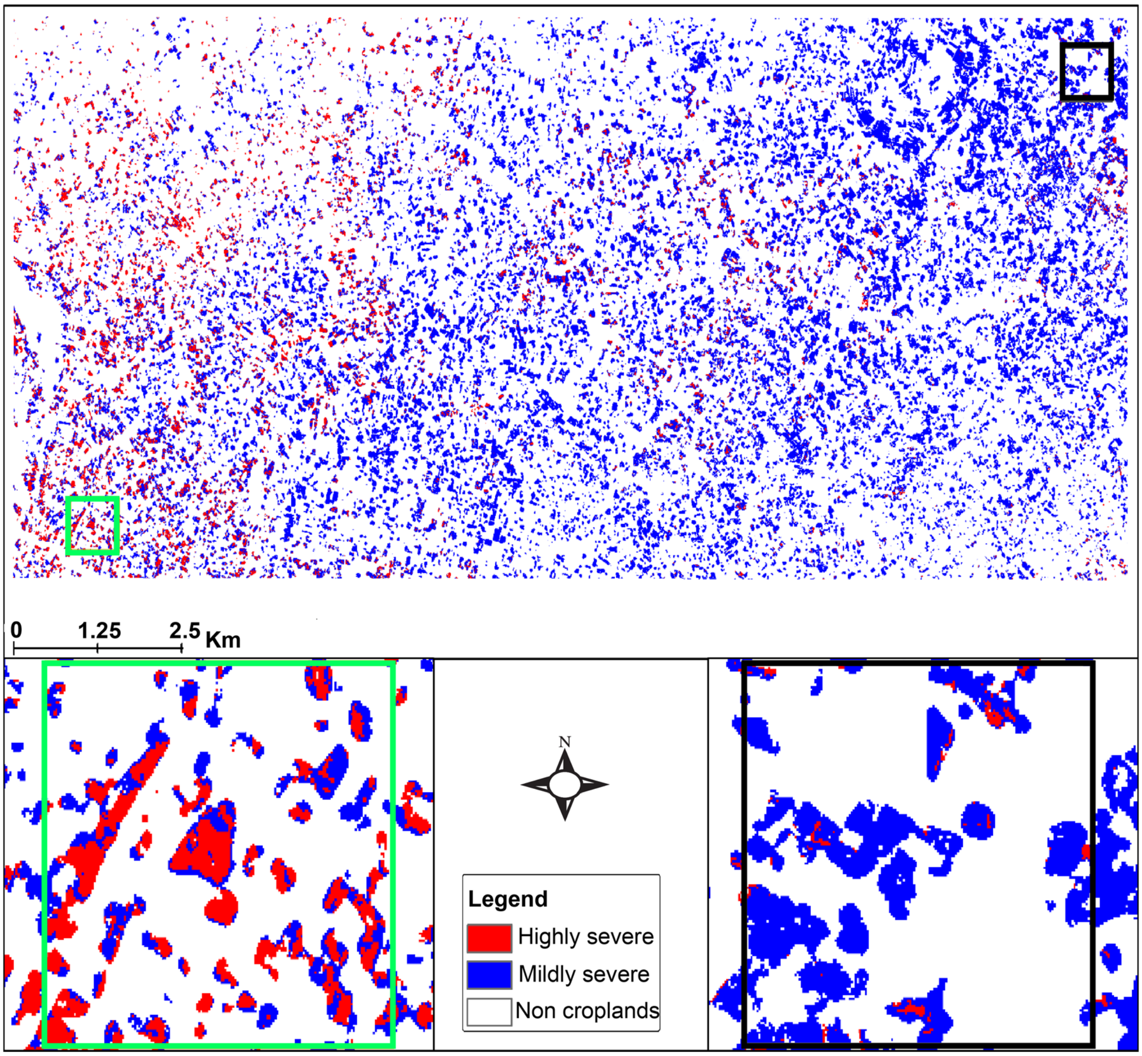

Fig. 6 The spatial distribution of maize lethal necrosis severity levels using the seven most important spectral variables selected by the random forest algorithm

as mangoes and banana trees within their fields, resulting in heterogeneity and spectral confusion between crops and other vegetation classes (Ayanu et al. 2015).

Essentially, MLN severity levels can accurately be distinguished if there are no other major stressors present that produce similar plant symptoms to those of the disease (Zhang et al. 2012). Field observations confirmed that MLN disease was the dominant stressor and that there was a minimal amount of interference from other biotic and abiotic factors in the sampled maize fields. To minimize such interference, we collected training polygons 5 m away from the farm edges to avoid edge effects and waterlogging which was observed to affect maize growing at the field edges in some of our sampled maize fields. Nevertheless, care was taken to ensure that infected fields were correctly identified by visually comparing each classification map with its original NDVI and true color images in the study.

The optimal predictor variables selected using optimized RF backward feature elimination technique were four SVIs (NDVI, GDVI, ChlRed-edge, and NDRE) extracted from the RE1 image acquired during the maize stem elongation stage combined with two SVIs (TSAVI and ChlRed-edge) and one spectral band (band 5) from the RE2 image acquired during maize inflorescence. These variables proved capable to discriminate two distinguishable MLN severity classes (mild and high) with the highest OA of 90.18% and a Kappa value of 0.92.

TSAVI was selected among the important variables for mapping MLN severity because of its ability to minimize soil brightness that influences spectral vegetation features involving red edge and NIR wavelengths (Mudereri et al. 2020b). Besides, TSAVI reduced soil background conditions which imposed extensive influence on partial canopy spectra and calculated SVIs. Similar results were reported by Dhau et al. (2018b) who found that soil-adjusted vegetation index (SAVI) was among the most important vegetation indices for detecting and mapping of maize streak virus using RE imagery.

Also, GDVI, NDRE, and NDVI were sensitive to MLN severity probably because severely infected maize plants are characterized by a low chlorophyll ratio followed by ultimate variations in leaf area. Previous studies showed the importance of NDVI in the monitoring of crop stress and disease detection (Eitel et al. 2011). Yet, GNDVI can better predict the leaf area index (LAI) than the conventional NDVI, while NDRE has demonstrated the ability to detect crop stress earlier than NDVI and GNDVI which are traditionally used for plant health monitoring (Wang et al. 2007). Inclusion of these vegetation indices by RF variable selection showed that changes in chlorophyll content are more sensitive to disease severity than changes in water content (Wang et al. 2016). Most notably, the presence of the red-edge band provided critical and subtle measurements of vegetation properties such as chlorophyll content necessary for distinguishing between healthy and disease-affected plants (Song et al. 2017). Therefore, our study supports the conclusion that strategically positioned bands such as the red edge found in new generation multispectral imagery contain more spectral information, useful for disease mapping in crop plants (Eitel et al. 2011; Chabalala et al. 2020).

Comparing the classification results generated using three MLN severity classes (mild, moderate, and high) with only two severity classes by merging mild and moderate-severe classes (mild and high) improved the overall accuracy by 16%. This implied that there was enormous spectral confusion between maize fields that were mildly and moderately affected by MLN. This confusion can be attributed to the fact that disease estimation is faced with much difficulty at the onset of early symptoms due to spectral similarity between slightly infected and non-infected fields (Ashourloo et al. 2014).

Based on the results from this study, a better understanding of the spatiotemporal characteristics of plant diseases is crucial in developing detection tools that are applicable for multi-temporal analyses and the temporal dimension of crop diseases. Therefore, sensor-based identification must be explored further to establish on what resolution and magnitudes disease infestation can be mapped with other sensors (Fang and Ramasamy 2015). Considering that the occurrence of plant diseases is dependent on explicit environmental factors and that diseases often exhibit a heterogeneous distribution, optical sensing techniques are useful in identifying primary disease foci and within field disease severity patterns (Melesse et al. 2007).

Conclusions

Monitoring of MLN severity levels is of immense practical importance, given that the disease tends to develop rapidly and that it is presently very difficult to precisely forecast its development. In this study, a method for mapping MLN severity using bi-temporal RapidEye satellite remote sensing data and optimized machine learning algorithm was developed and tested, ensuring systematic monitoring of MLN damage levels over a large area. Our results indicate the suitability of remote sensing data as a complementary tool for disease monitoring which could help in the development of effective disease control strategies. Although a low temporal resolution dataset with the high spatial resolution is a restrictive factor for practical implementation, the launch of future observation systems with improved repetition rates such as Sentinel-2 can broaden the field of applications. Therefore, explicit geospatial and timely synoptic tools are needed for the monitoring of pests and disease damage levels to facilitate better and more targeted mitigation measures in maize and other important crops. Besides, the effectiveness of remote sensing based on the spatiotemporal dynamics of MLN should be investigated in future studies to understand linkages between maize pests and disease hotspots with the underlying ecological factors for better and precise monitoring and management practices.

Acknowledgments This work received financial support from the German Federal Ministry for Economic Cooperation and Development (BMZ) commissioned and administered through the Deutsche Gesellschaft fur Internationale Zusammenarbeit (GIZ) Fund for International Agricultural Research (FIA) "Better implementation of crop season breaks for management of Maize Lethal Necrosis Virus in East Africa – Can remote sensing be an option?" grant number 11.7860.7-001.00; UK's Foreign, Commonwealth & Development Office (FCDO); the Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); the Federal Democratic Republic of Ethiopia; and the Government of the Republic of Kenya. We are grateful to Julius-Maximilians-University Würzburg, Department of Remote Sensing, for their cooperation and help with satellite data preprocessing. *The views expressed herein do not necessarily reflect the official opinion of the donors.*

References

Abdel-rahman EM, Landmann T, Kyalo R et al (2017) Predicting stem borer density in maize using RapidEye data and generalized linear models. Int J Appl Earth Obs Geoinf 57:61–74. https://doi.org/10.1016/j.jag.2016.12.008

Adams IP, Miano DW, Kinyua ZM, Wangai A, Kimani E, Phiri N, Reeder R, Harju V, Glover R, Hany U, Souza-Richards R, Deb Nath P, Nixon T, Fox A, Barnes A, Smith J, Skelton A, Thwaites R, Mumford R, Boonham N (2013) Use of next-generation sequencing for the identification and characterization of maize chlorotic mottle virus and sugarcane mosaic virus causing maize lethal necrosis in Kenya. Plant Pathol 62:741–749. https://doi.org/10.1111/j.1365-3059.2012.02690.x

AGRA (2017) Africa agriculture status report: the business of smallholder agriculture in sub-Saharan Africa (Issue 5). Nairobi, Kenya: Alliance for a Green Revolution in Africa (AGRA), Issue No. 5. Alliance for a Green Revolution in Africa

Albayrak S (2008) Use of reflectance measurements for the detection of N, P, K, ADF and NDF contents in sainfoin pasture. Sensors 8: 7275–7286. https://doi.org/10.3390/s8117275

Ashourloo D, Mobasheri MR, Huete A (2014) Developing two spectral disease indices for detection of wheat leaf rust (Puccinia triticina). Remote Sens 6:4723–4740. https://doi.org/10.3390/rs6064723

Ayanu Y, Conrad C, Jentsch A, Koellner T (2015) Unveiling undercover cropland inside forests using landscape variables: a supplement to remote sensing image classification. PLoS One 10:1–21. https://doi.org/10.1371/journal.pone.0130079

Benson JM, Poland JA, Benson BM, Stromberg EL, Nelson RJ (2015) Resistance to gray leaf spot of maize: genetic architecture and mechanisms elucidated through nested association mapping and near-isogenic line analysis. PLoS Genet 11:1–24. https://doi.org/10.1371/journal.pgen.1005045

Berk A, Anderson G, Acharya P, Shettle E (2008) MODTRAN®5.2.0.0 USER'S MANUAL A. Spectral Sciences, Inc

Bock CH, Poole GH, Parker PE, Gottwald TR (2010) Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. Crit Rev Plant Sci 29:59–107. https://doi.org/10.1080/07352681003617285

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Chabalala Y, Adam E, Oumar Z, Ramoelo A (2020) Exploiting the capabilities of Sentinel-2 and RapidEye for predicting grass nitrogen across different grass communities in a protected area. Applied Geomatics 12:379–395. https://doi.org/10.1007/s12518-020-00305-8

Chemura A, Mutanga O, Dube T (2017a) Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. Precis Agric 18:859–881. https://doi.org/10.1007/s11119-016-9495-0

Chemura A, Mutanga O, Odindi J (2017b) Empirical modeling of leaf chlorophyll content in coffee (Coffea Arabica) plantations with Sentinel-2 MSI data: effects of spectral settings, spatial resolution, and crop canopy cover. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10:5541–5550. https://doi.org/10.1109/JSTARS.2017.2750325

Congalton RG (2001) Accuracy assessment and validation of remotely sensed and other spatial information. Int J Wildland Fire 10:321–328. https://doi.org/10.1071/wf01031

Crnojevic V, Lugonja P, Brkljac B, Brunet B (2014) Classification of small agricultural fields using combined Landsat-8 and RapidEye imagery: case study of northern Serbia. J Appl Remote Sens 8:083512. https://doi.org/10.1117/1.jrs.8.083512

Deressa T, Demissie G (2017) Maize lethal necrosis disease (MLND) – a review. Journal of Natural Sciences Research 7:38–42

Dhau I, Adam E, Mutanga O, Ayisi K, Abdel-Rahman EM, Odindi J, Masocha M (2018a) Testing the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot disease in maize crop. Geocarto International 33:1223–1236. https://doi.org/10.1080/10106049.2017.1343391

Dhau I, Adam E, Mutanga O, Ayisi KK (2018b) Detecting the severity of maize streak virus infestations in maize crop using in situ hyperspectral data. Transactions of the Royal Society of South Africa 73:8–15. https://doi.org/10.1080/0035919X.2017.1370034

Dhau I, Adam E, Ayisi KK, Mutanga O (2019) Detection and mapping of maize streak virus using RapidEye satellite imagery. Geocarto International 34:856–866. https://doi.org/10.1080/10106049.2018.1450448

Eitel JUH, Vierling LA, Litvak ME, Long DS, Schulthess U, Ager AA, Krofcheck DJ, Stoscheck L (2011) Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. Remote Sens Environ 115:3640–3646. https://doi.org/10.1016/j.rse.2011.09.002

Fang Y, Ramasamy RP (2015) Current and prospective methods for plant disease detection. Biosensors 5:537–561. https://doi.org/10.3390/bios5030537

FAO, IFAD, UNICEF, et al (2019) The State of Food Security and Nutrition in the World 2019. Safeguarding against economic slowdowns and downturns. Rome, FAO Licence: CC BY-NC-SA 3.0 IGO

Fierens F, Rosin PL (1994) Filtering remote sensing data in the spatial and feature domains. In: Desachy J (ed) Image and Signal Processing for Remote Sensing. SPIE, pp 472–482

Forkuor G, Conrad C, Thiel M, Ullmann T, Zoungrana E (2014) Integration of optical and synthetic aperture radar imagery for improving crop mapping in Northwestern Benin. West Africa.:6472–6499. https://doi.org/10.3390/rs6076472

Forkuor G, Conrad C, Thiel M, Landmann T, Barry B (2015) Evaluating the sequential masking classification approach for improving crop discrimination in the Sudanian Savanna of West Africa. Comput Electron Agric 118:380–389. https://doi.org/10.1016/j.compag.2015.09.020

Franke J, Menz G (2007) Multi-temporal wheat disease detection by multi-spectral remote sensing. Precis Agric 8:161–172. https://doi.org/10.1007/s11119-007-9036-y

Geerts S, Raes D, Garcia M, del Castillo C, Buytaert W (2006) Agro-climatic suitability mapping for crop production in the Bolivian Altiplano: a case study for quinoa. Agric For Meteorol 139:399–412. https://doi.org/10.1016/j.agrformet.2006.08.018

Georganos S, Grippa T, Vanhuysse S, Lennert M, Shimoni M, Kalogirou S, Wolff E (2018) Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. GIScience and Remote Sensing 55:221–242. https://doi.org/10.1080/15481603.2017.1408892

Guanter L, Richter R, Kaufmann H (2009) On the application of the MODTRAN4 atmospheric radiative transfer code to optical remote sensing. Int J Remote Sens 30:1407–1424. https://doi.org/10.1080/01431160802438555

Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 2018:e5518. https://doi.org/10.7717/peerj.5518

Hilker FM, Allen LJS, Bokil VA, Briggs CJ, Feng Z, Garrett KA, Gross LJ, Hamelin FM, Jeger MJ, Manore CA, Power AG, Redinbaugh MG, Rúa MA, Cunniffe NJ (2017) Modeling virus coinfection to inform management of maize lethal necrosis in Kenya. Phytopathology 107:1095–1108. https://doi.org/10.1094/PHYTO-03-17-0080-FI

Huang BFF, Boutros PC (2016) The parameter sensitivity of random forests. BMC Bioinformatics 17:1–13. https://doi.org/10.1186/s12859-016-1228-x

Jaetzold R, Schmidt H (1982) Farm management handbook of Kenya. Ministry of Agriculture, Nairobi

Jozani HJ, Thiel M, Abdel-rahman EM et al (2020) Investigation of maize lethal necrosis (MLN) severity and cropping systems mapping in agro-ecological maize systems in Bomet, Kenya utilizing RapidEye and Landsat-8 imagery. Geology, Ecology, and Landscapes 00:1–16. https://doi.org/10.1080/24749508.2020.1761195

Kusia ES (2014) Characterization of maize chlorotic mottle virus and sugarcane mosaic virus causing maize lethal necrosis disease and spatial distribution of their alternative hosts in Kenya. Pan African University, Kenya

Kyalo R, Abdel-Rahman EM, Subramanian S et al (2017) Maize cropping systems mapping using RapidEye observations in agro-ecological landscapes in Kenya. Sensors 17:2537. https://doi.org/10.3390/s17112537

Mahlein A-K (2016) Present and future trends in plant disease detection. Plant Dis 100:1–11. https://doi.org/10.1007/s13398-014-0173-7.2

Mahuku G, Lockhart BE, Wanjala B, Jones MW, Kimunye JN, Stewart LR, Cassone BJ, Sevgan S, Nyasani JO, Kusia E, Kumar PL, Niblett

CL, Kiggundu A, Asea G, Pappu HR, Wangai A, Prasanna BM, Redinbaugh MG (2015) Maize lethal necrosis (MLN), an emerging threat to maize-based food security in sub-Saharan Africa. Phytopathology 105:956–965. https://doi.org/10.1094/PHYTO-12-14-0367-FI

McHugh ML (2012) Lessons in biostatistics interrater reliability: the kappa statistic. Biochemica Medica 22:276–282

Melesse AM, Weng Q, Thenkabail PS, Senay GB (2007) Remote sensing sensors and applications in environmental resources mapping and modelling. Sensors 7:3209–3241. https://doi.org/10.3390/s7123209

Mudereri BT, Chitata T, Mukanga C, Mupfiga ET, Gwatirisa C, Dube T (2019a) Can biophysical parameters derived from Sentinel-2 spaceborne sensor improve land cover characterization in semi-arid regions? Geocarto International:1–20. https://doi.org/10.1080/10106049.2019.1695956

Mudereri BT, Dube T, Adel-Rahman EM, Niassy S, Kimathi E, Khan Z, Landmann T (2019b) A comparative analysis of PlanetScope and Sentinel-2 space-borne sensors in mapping Striga weed using guided regularised random forest classification ensemble. ISPRS - International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences XLII-2(W13):701–708. https://doi.org/10.5194/isprs-archives-XLII-2-W13-701-2019

Mudereri BT, Abdel-Rahman E., Dube T, et al (2020a) Potential of resampled multispectral data for detecting Desmodium-Brachiaria intercropped with maize in a "Push-Pull" system. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B3-2:1017–1022. doi: https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1017-2020

Mudereri BT, Dube T, Niassy S, Kimathi E, Landmann T, Khan Z, Abdel-Rahman EM (2020b) Is it possible to discern Striga weed (Striga hermonthica) infestation levels in maize agro-ecological systems using in-situ spectroscopy? Int J Appl Earth Obs Geoinf 85: 102008. https://doi.org/10.1016/j.jag.2019.102008

Mutanga O, Adam E, Cho MA (2012) High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm. Int J Appl Earth Obs Geoinf 18:399–406. https://doi.org/10.1016/j.jag.2012.03.012

Mwatuni FM, Nyende AB, Njuguna J, Xiong Z, Machuka E, Stomeo F (2020) Occurrence, genetic diversity, and recombination of maize lethal necrosis disease-causing viruses in Kenya. Virus Res 286: 198081. https://doi.org/10.1016/j.virusres.2020.198081

Myers SS, Smith MR, Guth S, Golden CD, Vaitla B, Mueller ND, Dangour AD, Huybers P (2017) Climate change and global food systems: potential impacts on food security and undernutrition. Annu Rev Public Health 38:259–277. https://doi.org/10.1146/annurev-publhealth-031816-044356

Nguyen HTT, Doan TM, Radeloff V (2018) Applying random forest classification to map land use/land cover using Landsat 8 OLI. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 42:363–367. https://doi.org/10.5194/isprs-archives-XLII-3-W4-363-2018

Nhamo L, Matchaya G, Mabhaudhi T et al (2019) Cereal production trends under climate change: impacts and adaptation strategies in Southern Africa. Agriculture (Switzerland) 9:1–16. https://doi.org/10.3390/agriculture9020030

Nutter FW, Schultz PM (1995) Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. Can J Plant Pathol 17:174–184. https://doi.org/10.1080/07060669509500709

Ochieng J, Kirimi L, Mathenge M (2016) Effects of climate variability and change on agricultural production: the case of small scale farmers in Kenya. NJAS - Wageningen Journal of Life Sciences 77:71–78. https://doi.org/10.1016/j.njas.2016.03.005

Omiti JM, Otieno DJ, Nyanamba TO, McCullough E (2009) Factors influencing the intensity of market participation by smallholder farmers: a case study of rural and peri-urban areas of Kenya. Afjare 3:2009. https://doi.org/10.22004/ag.econ.56958

Osunga M, Mutua F, Mugo R (2017) Spatial modelling of maize lethal necrosis disease in Bomet County, Kenya. Journal of Geosciences and Geomatics, Vol 5, 2017, Pages 251-258 5:251–258. https://doi.org/10.12691/JGG-5-5-4

Paul PA, Munkvold GP (2004) A model-based approach to preplanting risk assessment for gray leaf spot of maize. Phytopathology 94: 1350–1357. https://doi.org/10.1094/PHYTO.2004.94.12.1350

RapidEye (2018) RapidEye Mosaic TM product specifications. http://blackbridge.com/rapideye/upload/RapidEye.

Sibanda M, Mutanga O, Dube T, et al (2019) The utility of the upcoming HysPIRI's simulated spectral settings in detecting maize gray leafy spot in relation to sentinel-2 MSI, VenµS, and Landsat 8 OLI sensors. Agronomy 9:. doi:https://doi.org/10.3390/agronomy9120846

Song X, Yang C, Wu M, Zhao C, Yang G, Hoffmann W, Huang W (2017) Evaluation of Sentinel-2A satellite imagery for mapping cotton root rot. Remote Sens 9:1–17. https://doi.org/10.3390/rs9090906

Su TC (2016) A filter-based post-processing technique for improving homogeneity of pixel-wise classification data. European Journal of Remote Sensing 49:531–552. https://doi.org/10.5721/EuJRS20164928

Wang F, Huang J, Tang Y, Wang X (2007) New vegetation index and its application in estimating leaf area index of rice. Rice Sci 14:195–203. https://doi.org/10.1016/s1672-6308(07)60027-4

Wang L, Zhou X, Zhu X et al (2016) Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. Crop Journal 4:212–219. https://doi.org/10.1016/j.cj.2016.01.008

Wangai AW, Redinbaugh MG, Kinyua ZM, Miano DW, Leley PK, Kasina M, Mahuku G, Scheets K, Jeffers D (2012) First report of maize chlorotic mottle virus and maize lethal necrosis in Kenya. Plant Dis 96:1582. https://doi.org/10.1094/PDIS-06-12-0576-PDN

Worldbank (2018) Rural population: sub-Saharan Africa

Zhang J, Pu R, Huang W, Yuan L, Luo J, Wang J (2012) Using in-situ hyperspectral data for detecting and discriminating yellow rust disease from nutrient stresses. Field Crop Res 134:165–174. https://doi.org/10.1016/j.fcr.2012.05.011

Zhang J, Huang Y, Yuan L, Yang G, Chen L, Zhao C (2016) Using satellite multispectral imagery for damage mapping of armyworm (Spodoptera frugiperda) in maize at a regional scale. Pest Manag Sci 72:335–348. https://doi.org/10.1002/ps.4003