Assessing the suitability of data from Sentinel-1A and 2A for crop classification

Assessing the Suitability of Data from Sentinel-1A and 2A for Crop Classification

Sentinel-1A C-SAR and Sentinel-2A MultiSpectral Instrument (MSI) provide data applicable to the remote identification of crop type. In this study, six crop types (beans, beetroot, grass, maize, potato, and winter wheat) were identified using five C-SAR images and one MSI image acquired during the 2016 growing season. To assess the potential for accurate crop classification with existing supervised learning models, the four different approaches of kernel-based extreme learning machine (KELM), multilayer feedforward neural networks, random forests, and support vector machine were compared. Algorithm hyperparameters were tuned using Bayesian optimization. Overall, KELM yielded the highest performance, achieving an overall classification accuracy of 96.8%. Evaluation of the sensitivity of classification models and relative importance of data types using data-based sensitivity analysis showed that the set of VV polarisation data acquired on 24 July (Sentinel-1A) and band 4 data (Sentinel-2A) had the greatest potential for use in crop classification.

Keywords: Agricultural fields; classification; Hokkaido; machine learning; Sentinel-1A; Sentinel-2A

1. Introduction

The identification and mapping of crops is important for estimating potential harvest as well as for agricultural field management, and provides information for national and multinational agricultural agencies, insurance agencies, and regional agricultural boards. However, as of 2016 some local governments in Japan are still using manual effort to document field properties such as crop type and location (Ministry of Agriculture, Forestry and Fisheries, 2016). The high expense of these manual methods suggests a necessity to develop more efficient techniques. Remote sensing technology is a very useful tool for gathering a large amount of information simultaneously (Ryu *et al.*, 2011). While some *in situ* data is still required for generating and validating classification models, remote sensing is generally also effective at reducing labour costs.

In the present study, the applicability of data acquired from Sentinel-1A C-SAR and Sentinel-2A MultiSpectral Instrument (MSI) for generating crop maps was evaluated. Previous studies have investigated the use of C-band SAR data for monitoring vegetation state (Fieuzal and Baup, 2016; Haldar *et al.*, 2016) and for discriminating between crop types (Larranaga and Alvarez-Mozos, 2016). Multi-temporal SAR data following annual plant growing cycles are useful for clarifying temporal pattern changes (Costa, 2004). While the use of exclusively backscattering coefficients yielded an overall accuracy of less than 50% (Roychowdhury, 2016), more accurate classifications have been possible using a combination of Haralik textures, the polarization ratio and the local mean together with the VV backscattering coefficients (Inglada *et al.*, 2016). However, in some areas (including our study area) there were few opportunities to obtain polarimetric Sentinel-1A data.

Some studies have shown that phenology features derived from optical sensors are useful to estimate crop acreage (Nigam *et al.*, 2015; Zhang *et al.*, 2017). Biophysical parameters including fresh and dry weight and leaf area index (LAI) can also be retrieved from vegetation indices derived from the Landsat 8 OLI and the Landsat 7 ETM+ (Ahmadian *et al.*, 2016); these data have proven effective in identifying crop types with high accuracy (Goodin *et al.*, 2015). Moreover, red-edge or short wave infra-red reflectance data have been provided by various satellites such as RapidEye (Eitel *et al.*, 2007) and Landsat 8 OLI (Roy *et al.*, 2014), and have contributed to improvements in crop monitoring over large areas (Kim and Yeom, 2015; Sonobe *et al.*, 2017b). These data as provided by Sentinel-2A may prove useful for the same purpose. Sentinel-2A data have been shown to be suited for mapping urban green species and may help in reducing the amount of manual digitizing while sustaining a high level of accuracy (Rosina and Kopecka, 2016). Huang *et al*. (2017) further demonstrated that the near-infrared, short wave infrared and red-edge bands are useful for separating unburned and burned areas, due to these bands' sensitivity to vegetation state and soil moisture changes. As observations derived from optical sensors are sometimes influenced by cloud interference, multi-sensor approaches (combining optical and microwave data) may be used to improve classification accuracy (Sheoran and Haack, 2013; Eberhardt *et al.*, 2016). A significant improvement in classification accuracy was confirmed when Sentinel-1A SAR and Landsat8 satellite image time series were integrated (Inglada *et al.*, 2016). This indicates that integrating data from Sentinel-1A and 2A may also have great potential for high-accuracy crop classification.

In addition to good quality remote sensing data, classification algorithms are essential for generating accurate maps. Different machine learning approaches have been used for image classification over the past two decades (Pal *et al.*, 2013). The Support Vector Machine (SVM) has been one of the most effective classification approaches, and has been widely used with a Gaussian kernel function (Burges, 1998).

For example, a SVM classifier achieved an overall accuracy of 92.0 % both for identification of soil types and of five crop types (Foody and Mathur, 2004). The random forests (RF) approach has also been very successful for classification and regression using remote sensing data (Biau and Scornet, 2016), and was shown to perform as well as SVM in terms of classification accuracy and training time (Pal, 2005). A recently developed extension of machine learning, deep learning, has enabled the use of multilayer feedforward neural networks (FNN) which have also been applied to optical remote sensing data (Cooner *et al.*, 2016) and several classification approaches based on this technology have received scrutiny (Foody, 2000; Brown *et al.*, 2009). A more efficient fast learning neural algorithm for single hidden layer feedforward neural networks, called the extreme learning machine (ELM; Huang *et al.*, 2012) has been applied in a similar manner (Sonobe *et al.*, 2017a).

While these algorithms have been widely used for land cover classification, parameter tuning is always required and may result in the deterioration of accuracies (Xue *et al.*, 2017). For optimising the hyperparameters of machine learning algorithms, grid search strategies have been applied (Puertas *et al.*, 2013). However, as these may constitute a poor choice for configuring algorithms for new data sets, the use of Bayesian optimisation has been suggested. This is a framework for sequential optimisation of the hyperparameters of noisy, expansive black-box functions (Bergstra and Bengio, 2012), and represents one possible method to unify hyperparameter tuning for performance comparison among machine learning algorithms.

Evaluating the importance of each variable is useful in such comparisons. Although RF generates importance measures for variables, a bias in variable selection during the tree-building process may lead to biased variable importance measures (VIMs) when variables are correlated to some degree (Nicodemus *et al.*, 2010). Other algorithms are generally more difficult to implement, and few studies have engaged in cross-algorithm comparisons. One tool that allows robust assessment of multiple supervised learning black box data mining models is data-based sensitivity analysis (DSA; Cortez and Embrechts, 2013), and this approach to variable evaluation was used in the present study.

The main objectives of this paper are (1) to evaluate the potential of Sentinel-1 and -2 data for crop type classification and crop map generation, and (2) to identify whether reflectance values or gamma nought values are more suitable for classification.

2. Materials and Methods

2.1. Study area

The study area is located on Hokkaido, Japan, and encompasses the area 142°55′12″ to 143°05′51″ E, 42°52′48″ to 43°02′42″ N (Figure 1). The continental humid climate of the region features warm summers and cold winters, with an average annual temperature of 6°C and an annual precipitation of 920 mm.

<Figure 1>

The crops used in the study were several types of beans (soy, azuki, and kidney), maize, beetroot and potato, and various grasses. Figure 2 shows the stages of each crop. Beans and maize were sown in mid-May, while beetroot and potato were transplanted from late April to early May (Tokachi Subprefecture, 2016). Grasses, including timothy orchard grass and winter wheat, were sown in the previous year. Beans were harvested from late September to early November, beetroots in November, potatoes from late August to September, and winter wheat from late July to early August. Grasses were harvested twice a year, from late June to early July and in late August.

<Figure 2>

2.2. Reference data

Field location and attribute data, such as crop types, were based on manual surveys and provided by Tokachi Nosai (Obihiro, Hokkaido) as a polygon shape file. No more precise data exist for this area. Based on these data, a total of 4719 fields (981 beans fields, 569 beet fields, 640 grasslands, 317 maize fields, 783 potato fields and 1429 winter wheat fields) covered the area in 2016. Field size was 0.25–9.70 ha (median 2.04 ha) for beans, 0.21–9.98 ha (median 2.46 ha) for beetroot, 0.30–17.50 ha (median 2.21 ha) for grassland, 0.18–8.42 (median 1.67 ha) for maize, 0.25–8.48 ha (median 2.17 ha) for potato, and 2.00–14.6 ha (median 1.92 ha) for wheat.

2.3. Satellite data

Sentinel-1A follows a sun-synchronous, near-polar, circular orbit at a height of 693 km with a 12-day repeat cycle. The satellite is equipped with a C-band imager (C-SAR) at 5.405 GHz with an incidence angle between 20° and 45°. There are four imaging modes: Strip Map (SM), Interferometric Wide-swath (IW), Extra Wide-swath (EW), and Wave (WV). C-SAR also supports operation in dual polarisation (HH + HV, VV + VH) (Torres *et al.*, 2012).We used data acquired during descending passes on 13 May, 6 June, 30 June, 24 July, and 17 August, 2016 (Table 1(a)). Data were downloaded from the ESA Data Hub (https://scihub.copernicus.eu/dhus/) as Ground Range Detected (GRD) products, which have already been focused, multi-looked, calibrated, and projected to ground range. Data were converted to gamma nought (γ^0 dB), which are equally spaced radiometrically calibrated power images, and then orthorectified using the 10 m mesh DEM produced by the Geospatial Information Authority of Japan (GSI) and the Earth Gravitational Model 2008 (EGM2008).

Sentinel-2A is equipped with a MultiSpectral Instrument (MSI). Table 2 shows the spatial and spectral resolution of MSI bands. The three atmospheric bands were not used in this study because they are mainly dedicated to atmospheric corrections and cloud screening (Drusch *et al.*, 2012). The only MSI data that available for the study area in 2016 was acquired on 11 August (Table 1(b)). The Level 1C top-of-atmosphere reflectance data were downloaded from EarthExplorer (http://earthexplorer.usgs.gov/). All bands are converted to 10 m resolution using Sentinel-2 Toolbox version 5.0.4. To compensate for spatial variability and to avoid problems related to uncertainty in georeferencing, average values of satellite data from multiple images were calculated for each field.

<Table 1>

<Table 2>

2.4. Classification algorithm

A stratified random-sampling approach was used to divide the dataset into three parts: a training set (50%), which was used to fit the models; a validation set (25%) used to estimate prediction error for model selection; and a test set (25%) used for assessing generalisation error in the final selected model (Hastie *et al.*, 2009). The stratified random-sampling procedure was repeated ten times for more robust results. The following classification algorithms were used: support vector machine (SVM), random forests (RF), multilayer feedforward neural networks (FNN), and kernel-based extreme learning machine (KELM). All processes were implemented using R version 3.3.1 (R Core Team 2016).

SVM partitions data using maximum separation margins (Cortes and Vapnik, 1995). Since few real systems are linear, the ‘kernel trick’ was applied instead of attempting to fit a non-linear model (Aizerman *et al.*, 1964). We applied the Gaussian Radial Basis Function (RBF) kernel which has two hyperparameters that control the flexibility of the classifier: the regularization parameter C and the kernel bandwidth γ. High C values lead to high penalties for inseparable points, which may result in overfitting. In contrast, low C values lead to under-fitting. The γ value defines the reach of a single training example, with low values indicating 'far' and high values indicating 'close' reach.

RF is an ensemble learning technique that builds multiple trees based on random bootstrapped samples of the training data (Breiman, 2001). Nodes are split using the best split variable from a group of randomly selected variables (Liaw and Wiener, 2002). This strategy allows robustness against over-fitting and can handle thousands of dependent and independent input variables without variable deletion. The output is determined by a majority vote for the classification tree. The original RF has two hyperparameters: the number of trees (*ntree*) and the number of variables used to split the nodes (*mtry*). However, the best split for a node can increase classification accuracy (Ishwaran and Kogalur, 2007; Ishwaran *et al.*, 2008; Sonobe *et al.*, 2017b). Thus, three additional hyperparameters were considered: the minimum number of unique cases in a terminal node (*nodesize*), the maximum depth of tree growth (*nodedepth*), and the number of random splits (*nsplit*).

FNN, which are neural networks trained to a back-propagation learning algorithm, are the most popular neural networks and are composed of neurons that are ordered into layers. The first is called the input layer, the last, the output layer, and the layers in between are hidden layers (Svozil *et al.*, 1997). In the model, each neuron in a particular layer is connected with all neurons in the next layer. This connection is characterised by a weight (w_i) and a threshold coefficient (b). Thus, the basic unit is described as follows:

$$f(\sum_i w_i x_i + b), \qquad (1)$$

where function f represents the activation function used throughout the network. As the rectified linear activation function demonstrated high performance in image recognition tasks and is, biologically, an accurate model of neuron activations (LeCun *et al.*, 2015), it was applied in the present study. Dropout, a regularization method, was also used, as it was shown to be able to provide classifications. Tuning the learning rate and momentum is useful for overcoming poor convergence of standard back-propagation (Svozil *et al.*, 1997). The training mode begins with an arbitrary sample size (batch size) and proceeds iteratively. Each iteration of the complete training set is called an epoch, and the network adjusts the weights in the direction that reduces the error in each epoch. During the iterative process, the weights gradually converge on a locally optimal set of values. Finally, the softmax function without an activation function or bias is applied to the net inputs. In the present study we used the following parameters: number of hidden layers (*num_layer*), number of units (*num_unit*), dropout ratio (*dropout*) for each layer, learning rate (*learning.rate*), momentum (*momentum*), batch size (*batch.size*), and number of iterations of training data needed to train the model (*num.round*).

For extreme learning machine (ELM; Huang *et al.*, 2004), it is not necessary to tune the initial parameters of the hidden layer, and almost all non-linear piecewise continuous functions can be used as hidden neurons. Therefore, if for N arbitrary distinct samples $\{(x_i, t_i | x_i \in R_n, t_i \in R_m, i = 1, \ldots, N)\}$, the output function in an ELM with hidden neurons is

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta, \qquad (2)$$

where $\beta = \{\beta_1, \dots \beta_L\}$ is the vector of the output weights between the hidden layer of L neurons and the output neuron, and $h(x) = \{h_1(x), \dots h_L(x)\}$ is the output vector of the hidden layer with respect to input x. This maps the data from the input space to ELM feature space. To decrease training error and improve the generalization performance of neural networks, the training error and the output weights are simultaneously minimized using Karush-Kuhn-Tucker (KKT) conditions (Fletcher, 1981):

$$\beta = H^T \left(\frac{1}{C_r} + HH^T\right)^{-1} T, \tag{3}$$

where H is the hidden layer output matrix, Cr is the regulation coefficient, and T is the expected output matrix of samples. When the feature mapping $h(x)$ is unknown and the kernel matrix of ELM is based on Mercer's conditions, the output function $f(x)$ of the KELM can be written as follows:

$$f(x) = [k(x, x_i), \dots, k(x, x_N)] \left(\frac{1}{c} + HH^T\right)^{-1} T, \tag{4}$$

where *k*() is the kernel function of hidden neurons (here we applied the Radial Basis Function (RBF) kernel). In our study, the regulation coefficient (*Cr*) and the kernel parameter (*Kp*) were tuned.

Bayesian optimisation was applied for optimising the hyperparameters of the machine learning algorithms.

2.5. Accuracy assessment

As a first step, the ability to separate the six crop types statistically was evaluated using Jeffries-Matusita (J-M) distances (Richards, 1999). J-M distance values range from 0 to 2.0, with values greater than 1.9 indicating good separation, and values between 1.7 and 1.9 fairly good separation.

The classification results were evaluated according to the two simple measures of quantity disagreement (QD) and allocation disagreement (AD), which provide an effective summary of a cross-tabulation matrix. QD is defined as the difference between the reference data and the classified data based on a mismatch of class proportions, while AD is the difference between the classified data and the reference data due to incorrect spatial allocations of pixels in the classification. The sum of QD and AD indicates the total disagreement (Pontius and Millones, 2011). The results were further evaluated regarding their overall accuracy (OA), producer's accuracy (PA), and user's accuracy (UA). OA is the total classification accuracy. PA is obtained by dividing the number of correctly classified fields for each crop type by the number of reference fields. UA is computed by dividing the number of correctly classified fields for each crop type by the total number of fields classified as that crop type. McNemar's test was applied to identify whether there were significant differences between the two classification results (McNemar, 1947). This test takes the lack of independent samples into account by comparing how each point was either correctly or incorrectly classified in two compared classifications. A chi-square value above 3.84 indicates a significant difference between the two classification results at a 95% significance level.

The sensitivity of the classification models was determined using data-based sensitivity analysis (DSA). This simple method performs a pure black box use of the fitted models by querying the fitted models with sensitivity samples and recording their responses. DSA is similar to a computationally efficient one-dimensional sensitivity analysis (Kewley *et al.*, 2000), where only one input is changed at a time and the others are kept at their average values. However, this method uses several training samples instead of a baseline vector (Cortez and Embrechts, 2013).

3. Results and discussion

3.1. Acquired data and separability assessments

Figure 3 shows the time series of gamma nought values (γ^0) acquired from Sentinel-1A. The γ^0 values of beetroot crops increased as crop height increased throughout the season, while germinations of beans and maize remained unconfirmed by 13 May. After 6 June, the increases in γ^0 were confirmed with the growth of the crops. However, differences between bean and beetroot γ^0 values decreased with plant growth. In potato fields, direct reflections from the pronounced furrow ridges (30–35 cm in height) resulted in a simple scattering pattern after 30 June, which led to high γ^0 values (Figure 3).

The main scattering pattern of wheat changed from double-bounce scattering to volume scattering from mid-May to June. Correspondingly, the γ^0 values were relatively stable until harvesting. Initially the scattering pattern of grass was similar to that of wheat, however γ^0 increased after the first harvest conducted between 30 June and 24 July (Figure 3). Sentinel-1A data were thus useful for identification based on crop structure, since the total backscattering strength of the cropland is expressed as a function of direct backscattering strength from the ground, the stem-ground, the stem, the canopy-ground, and the canopy including multiple scattering within the canopy.

<Figure 3>

In contrast, reflectance from Sentinel-2A is shown in Figure 4. Significant differences in mean reflectance were found, except for the pairs of maize-beans for band 4, beetroot-maize for band 2, grass-beetroot for band 2, 3, 4, 5, 11 and 12, potato-beetroot for band 11and potato-grass for band 6, 7, 8 and 11 ($p < 0.05$, based on Tukey-Kramer test). Differences in reflectance were particularly clear between wheat and beans. Wheat harvesting was completed by 11 August, and thus wheat reflectance was similar to that of bare soil (although some residues were left in wheat fields), i.e., relatively high in bands 2–5, 10, and 11. Other crops had similar spectral patterns but peaked around bands 7–8a. This feature was particularly obvious for beans, beetroot, and grass, which are late growing-season crops or crops that ripen early.

<Figure 4>

Separability analysis is important to assess the performance of training data. The separability levels of the two classes were evaluated based on the J-M values. Figure 5(a) shows crop pairs with a J-M distance greater than 1.7 in at least one Sentinel-1A data set or one Sentinel-2A band, and Figure 5(b) shows pairs with a J-M distance below 1.0 in every data set and band. Distances above 1.7 were found between beans and wheat, beetroot and grass, beetroot and maize, beetroot and wheat, grass and potato, grass and wheat, maize and wheat, and potato and wheat. Distinguishing beetroot from wheat was particularly straightforward since ten data types illustrated the distinction, including VV polarisation (Sentinel-1A data) on 30 June and 24 July, and reflectance in bands 2, 4, and 6–12. The VV polarisation on 24 July was useful for discriminating between beans and wheat, beetroot and grass, beetroot and wheat, grass and potato, maize and wheat, and potato and wheat. In contrast, VV polarisation on 13 May and 6 June and reflectance in band 3 were unsuitable for distinguishing crop types.

3.2. Accuracies and statistical comparison

Optimal values for combinations of parameters were (C, γ) = (2^9, 2^{-7}) for SVM, (*ntree*, *mtry*, *nodesize*, *nodedepth*, *nsplit*) = (864, 6, 6, 21, 4) for RF, and (Cr, Kp) = (2^{21}, 2^{14}) for KELM. Two hidden layers were suitable for FNN with (*num_unit of first layer*, *num_unit of second layer*, *dropout*, *learning.rate*, *momentum*, *batch.size*, *num.round*) = (107, 257, 0.270, 135, 0.959, 21, 0.194). Accuracy results are tabulated in Table 3 and McNemar’s test results are shown in Table 4.

<Table 3>

<Table 4>

Although J-M distance values between some crop combinations were lower than 1.0, the PAs and UAs derived using the machine learning algorithms were greater than 0.9, excepting those of SVM (PA and UA for maize were 0.849 and 0.882, respectively). OAs were 96.0% for SVM, 95.7% for RF, 96.0% for FNN, and 96.8% for KELM; thus all machine learning algorithms performed well in classifying agricultural crops. However, the classification results were significantly different from each other based on McNemar's tests ($p < 0.05$, Table 4). Classification results by KELM (Figure 6) had the best OA and AD+QD, although FNN had a better QD value. FNN performed well for identifying wheat, which covered approximately 30% of the cropland, while showing relatively poor performance when identifying grass (UA of grass was 0.939). This led to a mismatch of class proportions between the reference data and the classification data. Figure 7 shows the relationship between field area and misclassified field for each algorithm. More than 90% of the misclassified fields were less than 700 a in area. and 50.9% (FNN) –78.1% (RF) of misclassified fields were below 200 a. Except for use with grasslands, KELM was the most robust algorithm for classifying smaller fields.

Since grasses cultivation employs fewer controls, a lot of weeds were present in grasslands. As a result, variation in spectral features were larger here than in other crop types, causing misclassifications of relatively larger fields. FNN in particular performed unsatisfactorily when identifying grasslands, with 84.2 % of misclassified fields consisting of grasslands. This percentage was much lower for the other algorithms, from 35.5% (KELM) to 26.3% (RF).

Overall, identifying maize fields was difficult due to the small number of fields and the similarity in their reflectance and γ^0 to those of bean fields (Figure 5). SVM classified 62.5% of omissions in maize fields as beans; KELM, 75.0%; RF, 71.4%; and FNN, 25% (here maize fields were mostly classified as grassland).

In contrast, identifying wheat fields was straightforward due to the large differences between growth stages when compared to other crops; in addition, cultivated wheat fields were already present at the acquisition date of Sentinel-2A. As a result, only 1.1% (FNN) –7.9 % (SVM) of the misclassified fields were wheat fields, the lowest error rate for each algorithms Beetroot was also easy to identify because it had high productivity in mid-August and was the only vegetation present during its growing season. In addition, the structure of beetroot (leaf rosettes) produced a simple scattering pattern easy to identify from VV polarization data. Therefore, crop pairs with J-M distances above 1.7 always involved beetroot, and beetroot was responsible for only 2.3% (FNN) –13.2 % (SVM) of the misclassified fields.

<Figure 6>

Table 5 shows the accuracy results achieved by KELM using three different satellite datasets: I) five Sentinel-1A images, II) one Sentinel-2A image and III) merged data. When using only Sentinel-1A data (in the present study, only VV polarization data), it was impossible to identify maize fields and most were misclassified as bean fields, which is also shown by the pair's low J-M distance (0.015–0.161). Although dataset II was already much superior to dataset I, classification results were further and significantly improved ($p < 0.05$, based on McNemar's test) when both were combined into dataset III.

Table 6 summarises many of the studies that have been undertaken for classification of crop types using satellite data of medium spatial resolution (less than 30 m). Although conditions such as study area and crop type in the present study differ from those in previous studies, study areas had similar cultivation styles and included the same crops (maize [corn], soybean, beetroot [sugar beet], potato, grass and wheat). Compared to those studies that used the same algorithms as those evaluated in the present study, our OA values were larger. This indicates the large potential of the combination of Sentinel-1A and 2A data and particularly of KELM. The approach proposed in the present study may thus be useful for other agricultural regions.

Some studies have reported that the integration and comparison of microwave and optical remote sensing images is useful for land use/land cover classifications (Villa *et al.*, 2015; Hutt *et al.*, 2016). This conclusion was confirmed in the present study; however, we used C-band SAR data while the above authors used X-band SAR data. The dependence of these conclusions on the specific type of optical data should be explored in future research.

Classification problems related to the borders of fields remain to be resolved. To make good use of remote sensing data in geographic object-based image analysis (GEOBIA), very fine resolutions of less than 1 m are required (Baker *et al.*, 2013). Some recent studies have however shown the potential of GEOBIA in conjunction with Landsat-8 OLI or Sentinel-2A MSI data (Immitzer *et al.*, 2016; Novelli *et al.*, 2016). With the available information, it is difficult to evaluate the degree of certainty related to the edges of the provided shape files provided. Future research is planned to address this question.

<Table 5>

3.3. Sensitivity analysis

To clarify which variables contributed to the high overall accuracy of each algorithm, a data-based sensitivity analysis (DSA) was conducted. The VV polarisation data acquired on 24 July and band 4 showed the greatest potential for crop classification, corroborating the results of J-M distance analyses (Figure 8). There was also support for the strong dependence identified between the two datasets for RF (Figure 8). Excluding the VV polarisation data reduced the OA from 95.7 to 94.8%, a significant difference ($p < 0.05$, McNemar’s test). There was an increase in the importance of band 4 (from 16.2 to 21.7%) and the VV polarisation data (from 9.5 to 19.4%). A similar tendency was identified for FNN; in this case, OAs decreased from 96.0 to 95.2%. While the VV polarisation data acquired on 24 July also had some influence on the KELM classification, high performance could still be yielded in its absence (OAs decreased from 96.8% to 96.5%). Excluding it did not substantially influence KELM classification accuracy. The most notable change was observed within band 6 (importance increased from 6.3 to 8.9%). However there was little dependence on this band (which had a more important role for SVM classification), and OA was still 96.0% when the VV polarisation data were excluded.

These results suggest some vulnerabilities of RF in cross-year training and classification, which is required for saving some manual effort related to collecting training data. However, the other algorithms, especially KELM, might show high performances in this area.

4. Conclusions

Sentinel-1A and 2A data are available free of charge and could be a valuable tool for managing agricultural fields. Some local governments in Japan are already investigating alternatives to manual documentation of field properties (including crop types and locations) in the interest of reducing labour costs. This study investigates the differences in classification accuracies among four classification algorithms (SVM, RF, FNN, and KELM) using five Sentinel-1A images and one Sentinel-2A image with the aim of determining the best method to generate crop maps.

We found that KELM generated the most accurate crop classification map for the study area, with an overall accuracy of 96.8%. VV polarisation data acquired on 24 July played the most important role in the RF and KELM classifications. In contrast, FNN was mostly dependent on band 4 data and SVM on band 6 data. KELM showed high flexibility in allowing for crop classification of almost undiminished quality (as determined by OA) even under data reduction by exclusion of the VV polarisation data. This implies that use of this algorithm would confer some robustness towards possible future sensor degradation in the satellites.

The results of this study verify the validity of this remote sensing method, demonstrate Sentinel-1A and 2A's remarkable potential for crop classification and suggest a great potential for expanded future use of data from both satellites.

Acknowledgments

The authors would like to thank Tokachi Nosai for providing the field location and attribute data used in this study.

Disclosure statement

No potential conflicts of interest are reported by the authors.

References

Ahmadian, N., Ghasemi, S., Wigneron, J.P. & Zolitz, R., 2016. Comprehensive study of the biophysical parameters of agricultural crops based on assessing Landsat 8 OLI and Landsat 7 ETM+ vegetation indices. *Giscience & Remote Sensing,* 53**,** 337-359.

Aizerman, M., Braverman, E. & Rozonoer, L., 1964. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control,* 25**,** 821-837.

Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M. & Brivio, P.A., 2016. Assessing in-season crop classification performance using satellite data: a test case in Northern Italy. *European Journal of Remote Sensing,* 49**,** 361-380.

Baker, B.A., Warner, T.A., Conley, J.F. & Mcneil, B.E., 2013. Does spatial resolution matter? A multi-scale comparison of object-based and pixel-based methods for detecting change associated with gas well drilling operations. *International Journal of Remote Sensing,* 34**,** 1633-1651.

Bergstra, J. & Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research,* 13**,** 281-305.

Biau, G. & Scornet, E., 2016. A random forest guided tour. *Test,* 25**,** 197-227.

Breiman, L., 2001. Random forests. *Machine Learning,* 45**,** 5-32.

Brown, K.M., Foody, G.M. & Atkinson, P.M., 2009. Estimating per-pixel thematic uncertainty in remote sensing classifications. *International Journal of Remote Sensing,* 30**,** 209-229.

Burges, C.J.C., 1998. A tutorial on Support Vector Machines for pattern recognition. *Data Mining and Knowledge Discovery,* 2**,** 121-167.

Cooner, A.J., Shao, Y. & Campbell, J.B., 2016. Detection of Urban Damage Using Remote Sensing and Machine Learning Algorithms: Revisiting the 2010 Haiti Earthquake. *Remote Sensing,* 8**,** 17.

Cortes, C. & Vapnik, V., 1995. Support-vector networks. *Machine Learning,* 20**,** 273-297.

Cortez, P. & Embrechts, M.J., 2013. Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences,* 225**,** 1-17.

Costa, M.P.F., 2004. Use of SAR satellites for mapping zonation of vegetation communities in the Amazon floodplain. *International Journal of Remote Sensing,* 25**,** 1817-1835.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. & Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment,* 120**,** 25-36.

Eberhardt, I.D.R., Schultz, B., Rizzi, R., Sanches, I.D., Formaggio, A.R., Atzberger, C., Mello, M.P., Immitzer, M., Trabaquini, K., Foschiera, W. & Luiz, A.J.B., 2016. Cloud Cover Assessment for Operational Crop Monitoring Systems in Tropical Areas. *Remote Sensing,* 8**,** 14.

Eitel, J.U.H., Long, D.S., Gessler, P.E. & Smith, A.M.S., 2007. Using in-situ measurements to evaluate the new RapidEye (TM) satellite series for prediction of wheat nitrogen status. *International Journal of Remote Sensing,* 28**,** 4183-4190.

Fieuzal, R. & Baup, F., 2016. Estimation of leaf area index and crop height of sunflowers using multi-temporal optical and SAR satellite data. *International Journal of Remote Sensing,* 37**,** 2780-2809.

Fletcher, R., 1981. *Practical methods of optimization. Volume 2: Constrained optimization (v. 2)* New York: John Wiley & Sons Ltd.

Foody, G. & Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing,* 42**,** 1335-1343.

Foody, G.M., 2000. Mapping land cover from remotely sensed data with a softened feedforward neural network classification. *Journal of Intelligent and Robotic Systems,* 29**,** 433-449.

Goodin, D.G., Anibas, K.L. & Bezymennyi, M., 2015. Mapping land cover and land use from object-based classification: an example from a complex agricultural landscape. *International Journal of Remote Sensing,* 36**,** 4702-4723.

Guarini, R., Bruzzone, L., Santoni, M. & Dini, L., 2015. Analysis on the Effectiveness of Multi-temporal COSMO-SkyMed (R) Images for crop classification. *Conference on Image and Signal Processing for Remote Sensing XXI,* Toulouse, FRANCE.

Haldar, D., Rana, P., Yadav, M., Hooda, R.S. & Chakraborty, M., 2016. Time series analysis of co-polarization phase difference (PPD) for winter field crops using polarimetric C-band SAR data. *International Journal of Remote Sensing,* 37**,** 3753-3770.

Hartfield, K., Marsh, S., Kirk, C. & Carriere, Y., 2013. Contemporary and historical classification of crop types in Arizona. *International Journal of Remote Sensing,* 34**,** 6024-6036.

Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition* The United States: Springer-Verlag New York.

Heine, I., Jagdhuber, T. & Itzerott, S., 2016. Classification and Monitoring of Reed Belts Using Dual-Polarimetric TerraSAR-X Time Series. *Remote Sensing*, 8.

Huang, G.B., Zhou, H.M., Ding, X.J. & Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics,* 42**,** 513-529.

Huang, G.B., Zhu, Q.-Y. & Siew, C.-K., 2006. Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. *International Joint Conference on Neural Networks (IJCNN2004),* Budapest, Hungary, 985--990.

Huang, X., Wang, J., Shang, J., Liao, C. & Liu, J., 2017. Application of polarization signature to land cover scattering mechanism analysis and classification using multi-temporal C-band polarimetric RADARSAT-2 imagery. *Remote Sensing of Environment,* 193**,** 11-28.

Hutt, C., Koppe, W., Miao, Y.X. & Bareth, G., 2016. Best Accuracy Land Use/Land Cover (LULC) Classification to Derive Crop Types Using Multitemporal, Multisensor, and Multi-Polarization SAR Satellite Images. *Remote Sensing,* 8.

Immitzer, M., Vuolo, F. & Atzberger, C., 2016. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing,* 8**,** 27.

Inglada, J., Vincent, A., Arias, M. & Marais-Sicre, C., 2016. Improved Early Crop Type Identification By Joint Use of High Temporal Resolution SAR And Optical Image Time Series. *Remote Sensing,* 8.

Ishwaran, H. & Kogalur, U.B., 2007. Random survival forests for R. *R News,* 7**,** 25-31.

Ishwaran, H., Kogalur, U.B., Blackstone, E.H. & S., L.M., 2008. Random survival forests. *The Annals of Applied Statistics,* 2**,** 841-860.

Kewley, R.H., Embrechts, M.J. & Breneman, C., 2000. Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks,* 11**,** 668-679.

Kim, H.O. & Yeom, J.M., 2015. Sensitivity of vegetation indices to spatial degradation of RapidEye imagery for paddy rice detection: a case study of South Korea. *Giscience & Remote Sensing,* 52**,** 1-17.

Larranaga, A. & Alvarez-Mozos, J., 2016. On the Added Value of Quad-Pol Data in a Multi-Temporal Crop Classification Framework Based on RADARSAT-2 Imagery. *Remote Sensing,* 8**,** 19.

Lecun, Y., Bengio, Y. & Hinton, G., 2015. Deep learning. *Nature,* 521**,** 436-444.

Liaw, A. & Wiener, M., 2002. Classification and regression by random Forest. *R News,* 2**,** 18-22.

Mcnemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika,* 12**,** 153-157.

Ministry of Agriculture, Forestry and Fisheries. 2016. "Remote sensing application for the policies of agriculture and fisheries" Accessed 1 June 2016. http://www8.cao.go.jp/space/comittee/dai36/siryou3-5.pdf

Nicodemus, K.K., Malley, J.D., Strobl, C. & Ziegler, A., 2010. The behaviour of random forest permutation-based variable importance measures under predictor correlation. *Bmc Bioinformatics,* 11.

Nigam, R., Vyas, S.S., Bhattacharya, B.K., Oza, M.P., Srivastava, S.S., Bhagia, N., Dhar, D. & Manjunath, K.R., 2015. Modeling temporal growth profile of vegetation index from Indian geostationary satellite for assessing in-season progress of crop area. *Giscience & Remote Sensing,* 52**,** 723-745.

Novelli, A., Aguilar, M.A., Nemmaoui, A., Aguilar, F.J. & Tarantino, E., 2016. Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: A case study from Almeria (Spain). *International Journal of Applied Earth Observation and Geoinformation,* 52**,** 403-411.

Ozdarici-Ok, A., Ok, A.O. & Schindler, K., 2015. Mapping of Agricultural Crops from Single High-Resolution Multispectral Images-Data-Driven Smoothing vs. Parcel-Based Smoothing. *Remote Sensing,* 7**,** 5611-5638.

Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing,* 26**,** 217-222.

Pal, M., Maxwell, A.E. & Warner, T.A., 2013. Kernel-based extreme learning machine for remote-sensing image classification. *Remote Sensing Letters,* 4**,** 853-862.

Pontius, R. & Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing,* 32**,** 4407-4429.

Puertas, O., Brenning, A. & Meza, F., 2013. Balancing misclassification errors of land cover classification maps using support vector machines and Landsat imagery in the Maipo river basin (Central Chile, 1975-2010). *Remote Sensing of Environment,* 137**,** 112-123.

Richards, J.A., 1999. *Remote Sensing Digital Image Analysis*: Berlin: Springer-Verlag.

Rosina, K. & Kopecka, M., 2016. Mapping of urban green spaces using Sentinel-2A data: Methodical aspect. *6th International Conference on Cartography and GIS,* Albena, BULGARIA: Bulgarian Cartographic Assoc, 562-568.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymbumer, L., Masek, J.G., Mccorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H. & Zhu, Z., 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment,* 145**,** 154-172.

Roychowdhury, K., 2016. Comparison Between Spectral, Spatial and Polarimetric Classification of Urban and Periurban Landcover Using Temporal Sentinel - 1 Images. *Xxiii ISPRS Congress, Commission Vii,* 41**,** 789-796.

Ryu, C., Suguri, M., Iida, M., Umeda, M. & Lee, C., 2011. Integrating remote sensing and GIS for prediction of rice protein contents. *Precision Agriculture,* 12**,** 378-394.

Sheoran, A. & Haack, B., 2013. Classification of California agriculture using quad polarization radar data and Landsat Thematic Mapper data. *Giscience & Remote Sensing,* 50**,** 50-63.

Sonobe, R., Tani, H., Wang, X., Kobayashi, N. & Shimamura, H., 2014. Random forest classification of crop type using multi- temporal TerraSAR- X dual- polarimetric data. *Remote Sensing Letters,* 5**,** 157-164.

Sonobe, R., Tani, H. & Wang, X.F., 2017a. An experimental comparison between KELM and CART for crop classification using Landsat-8 OLI data. *Geocarto International,* 32**,** 128-138.

Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N. & Mochizuki, K.-I., 2017b. Mapping crop cover using multi-temporal Landsat 8 OLI imagery. *International Journal of Remote Sensing,* 38**,** 4348-4361.

Svozil, D., Kvasnicka, V. & Pospichal, J., 1997. Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems,* 39**,** 43-62.

Tokachi Subprefecture. 2016. "Growing condition of crops in 2016" Accessed 1 June 2016. http://www.tokachi.pref.hokkaido.lg.jp/ss/num/kairyou/sakkyou27.htm

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I.N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'abbate, M., Croci, R., Pietropaolo, A., Huchler, M. & Rostan, F., 2012. GMES Sentinel-1 mission. *Remote Sensing of Environment,* 120**,** 9-24.

Villa, P., Stroppiana, D., Fontanelli, G., Azar, R. & Brivio, P.A., 2015. In-Season Mapping of Crop Type with Optical and X-Band SAR Data: A Classification Tree Approach Using Synoptic Seasonal Features. *Remote Sensing,* 7**,** 12859-12886.

Xue, Z.H., Du, P.J., Li, J. & Su, H.J., 2017. Sparse graph regularization for robust crop mapping using hyperspectral remotely sensed imagery with very few in situ data. *ISPRS Journal of Photogrammetry and Remote Sensing,* 124**,** 1-15.

Zhang, H., Li, Q., Liu, J., Shang, J., Du, X. & Zhao, L., 2017. Crop classification and acreage estimation in North Korea using phenology features. *Giscience & Remote Sensing,* 54**,** 381-406.

Tables

Table 1. Characteristics of the satellite data used in this study

(a) Sentinel-1A

Acquisition date	Incidence angle (°)		Pass direction	Cycle number	Orbit number
	Near	Far			
13-May-16	30.68	45.86	DESCENDING	78	11245
6-Jun-16	30.67	45.87	DESCENDING	80	11595
30-Jun-16	30.67	45.86	DESCENDING	82	11945
24-Jul-16	30.67	45.86	DESCENDING	84	12295
17-Aug-16	30.67	45.86	DESCENDING	86	12645

(b) Sentinel-2A

Acquisition date	Sun Zenith Angle (°)	Sun Azimuth Angle (°)	Orbit number
11-Aug-16	30.35	151.29	74

Table 2. Spatial and spectral resolution of MSI data.

Band	Spatial Resolution (m)	Central Wavelength (nm)	Bandwidth (nm)
Band 1	60	443	20
Band 2	10	490	65
Band 3	10	560	35
Band 4	10	665	30
Band 5	20	705	15
Band 6	20	740	15
Band 7	20	783	20
Band 8	10	842	115
Band 8a	20	865	20
Band 9	60	945	20
Band 10	60	1380	30
Band 11	20	1610	90
Band 12	20	2190	180

Table 3. Accuracy results for four classification algorithms: support vector machine (SVM), random forests (RF), multilayer feedforward neural networks (FNN), and kernel-based extreme learning machine (KELM). PA: producer's accuracy; UA: user's accuracy; OA: overall accuracy; AD: allocation disagreement; QD: quantity disagreement

	SVM	RF	FNN	KELM
PA				
Beans	0.974	0.948	0.963	0.953
Beetroot	0.959	0.967	0.967	0.992
Grassland	0.933	0.913	0.933	0.926
Maize	0.849	0.868	0.849	0.925
Potato	0.946	0.962	0.946	0.962
Wheat	0.990	0.994	0.994	0.997
UA				
Beans	0.916	0.923	0.944	0.958
Beetroot	0.967	0.992	0.975	0.984
Grassland	0.972	0.971	0.939	0.972
Maize	0.882	0.902	0.900	0.907
Potato	0.961	0.926	0.939	0.962
Wheat	0.994	0.981	0.994	0.978
OA	0.960	0.957	0.960	0.968
AD	2.714	2.818	3.445	2.401
QD	1.253	1.461	0.522	0.835

Table 4. Chi-square values from McNemar's test performed on results of four classification algorithms: support vector machine (SVM), random forests (RF), multilayer feedforward neural networks (FNN), and kernel-based extreme learning machine (KELM)

	SVM	RF	FNN	KELM
SVM	X	12.17	6.62	19.47
RF		X	12.30	13.15
FNN			X	18.20
KELM				X

Note: A chi-square value ≥ 3.84 indicate a significant difference ($p < 0.05$) between two classification results.

Table 5. Comparison of accuracy for six crop types achieved by KELM using three different satellite dataset. PA: producer's accuracy; UA: user's accuracy; OA: overall accuracy; AD: allocation disagreement; QD: quantity disagreement

	Sentinel-1A	Sentinel-2A	Sentinel-1A+2A
PA			
Beans	0.817	0.911	0.953
Beetroot	0.746	0.992	0.992
Grassland	0.779	0.933	0.926
Maize	0.038	0.943	0.925
Potato	0.808	0.962	0.962
Wheat	0.965	0.990	0.997
UA			
Beans	0.768	0.972	0.958
Beetroot	0.645	0.992	0.984
Grassland	0.823	0.979	0.972
Maize	0.500	0.926	0.907
Potato	0.772	0.880	0.962
Wheat	0.907	0.972	0.978
OA	0.806	0.959	0.968
AD	13.466	2.088	2.401
QD	5.950	1.983	0.835

Table 6. Summary of overall accuracy in reviewed studies

Sensor	Algorithm	Location	Class	Best overall accuracy	Reference
Landsat 8 OLI, COSMO-SkyMed	Classification and regression tree	Northern Italy	Maize, Rice, Soybean, Winter crop, Double crop, Forages, Forestry-woodland	0.918	(Villa *et al.*, 2015)
Kompsat-2	Support vector machine	Northwest Turkey	Corn, Pasture, Rice, Sugar Beet, Wheat, Tomato	0.9332	(Ozdarici-Ok *et al.*, 2015)
TerraSAR-X	Random forests	Northeastern Germany	Reed, Water, Meadow, Deciduous, Coniferous forest	0.9190	(Heine *et al.*, 2016)
TerraSAR-X, FORMOSAT-2	Optimized Maximum Likelihood	Northeast China	Coniferous Forest, Decideous Forest, Maize, Pumpkin, Rice, Soya, Urban, Concrete, Water	0.92	(Hutt *et al.*, 2016)
RADARSAT-2	MTSBTCS-MDPS	Southwestern Ontario, Canada	Corn, Soybean, Wheat, Grass, Forest, Urban	0.875	(Huang *et al.*, 2017)
COSMO-SkyMed	Support vector machine	Lower Austria	Carrot, Corn, Potato, Soybean, Sugar beet	0.845	(Guarini *et al.*, 2015)
Landsat 8 OLI	Support vector machine	Ukraine-Poland border	Artificial/urban, Bare, Grassland or Herbaceous cover, Woodland, Wetland, Water	0.89	(Goodin *et al.*, 2015)
Landsat Thematic Mapper	Classification and regression tree	Arizona	Alfalfa, Cotton, Grain, Fallow, Corn, Melon, Orchards/citrus, Sorghum	0.92	(Hartfield *et al.*, 2013)
Landsat 8 OLI	Maximum Likelihood	Northern Italy	Maize, Rice, Soybean, Winter crops, Forage crops	0.927	(Azar *et al.*, 2016)
TerraSAR-X	Random forests	Japan	Beans, Beet, Grass, Maize, Potato, Winter wheat	0.929	(Sonobe *et al.*, 2014)

MTSBTCS-MDPS: Multi-temporal supervised binary-tree classification scheme - Maximum power difference of polarization signature (MDPS)

Figures

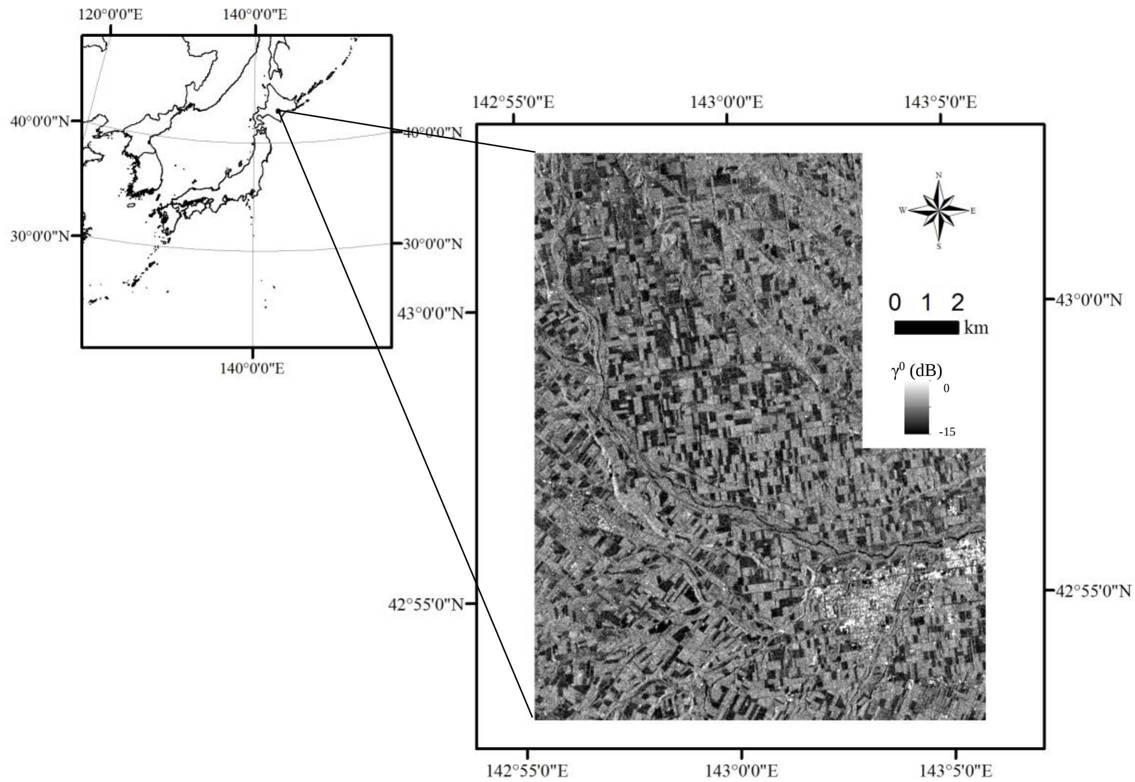

Figure 1. The study area in Hokkaido, Japan. Enlarged map shows Sentinel-1A VV polarization data acquired on 24 July, 2016.

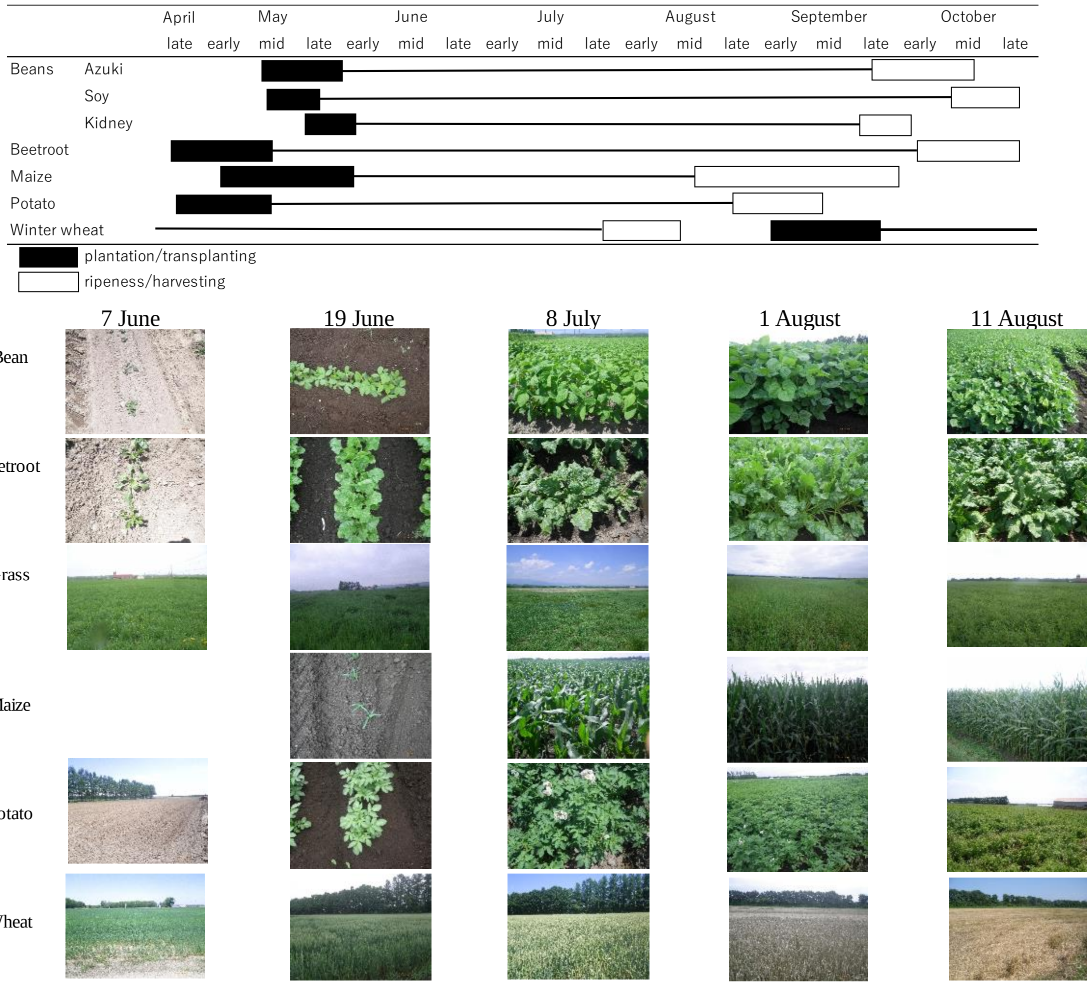

Figure 2. Crop growth stages in the study area.

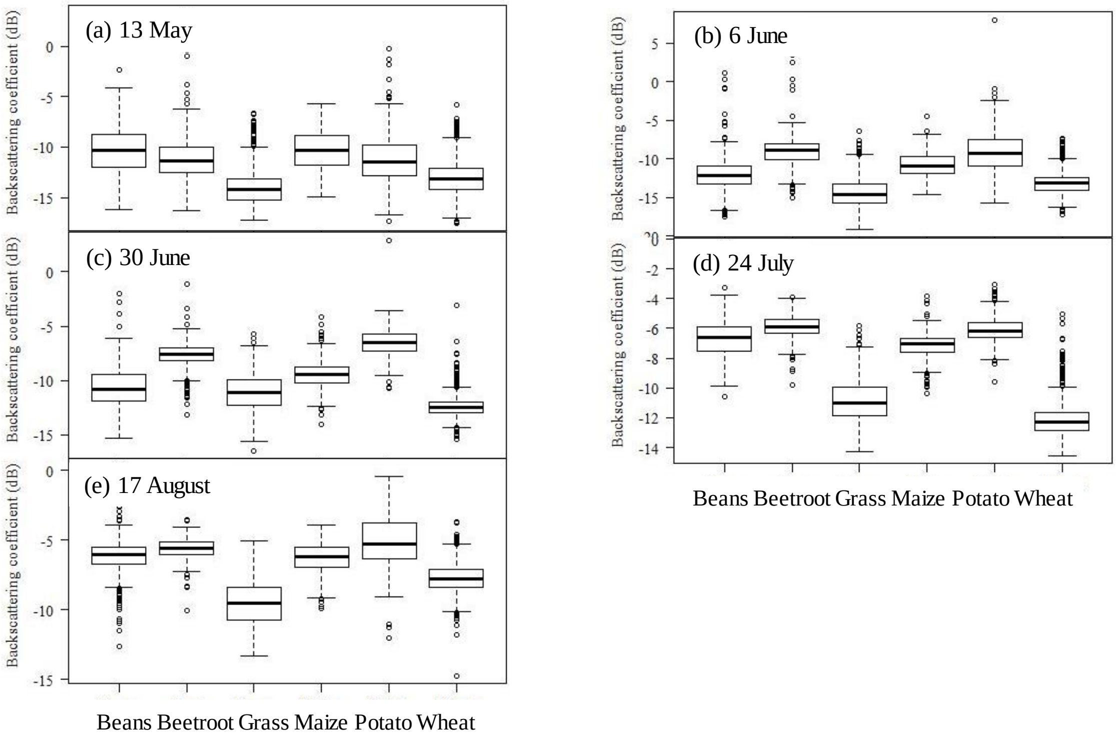

Figure 3. Boxplots of gamma nought (γ^0) values acquired from Sentinel-1A on (a) 13 May, (b) 6 June, (c) 30 June, (d) 24 July, and (e) 17 August.

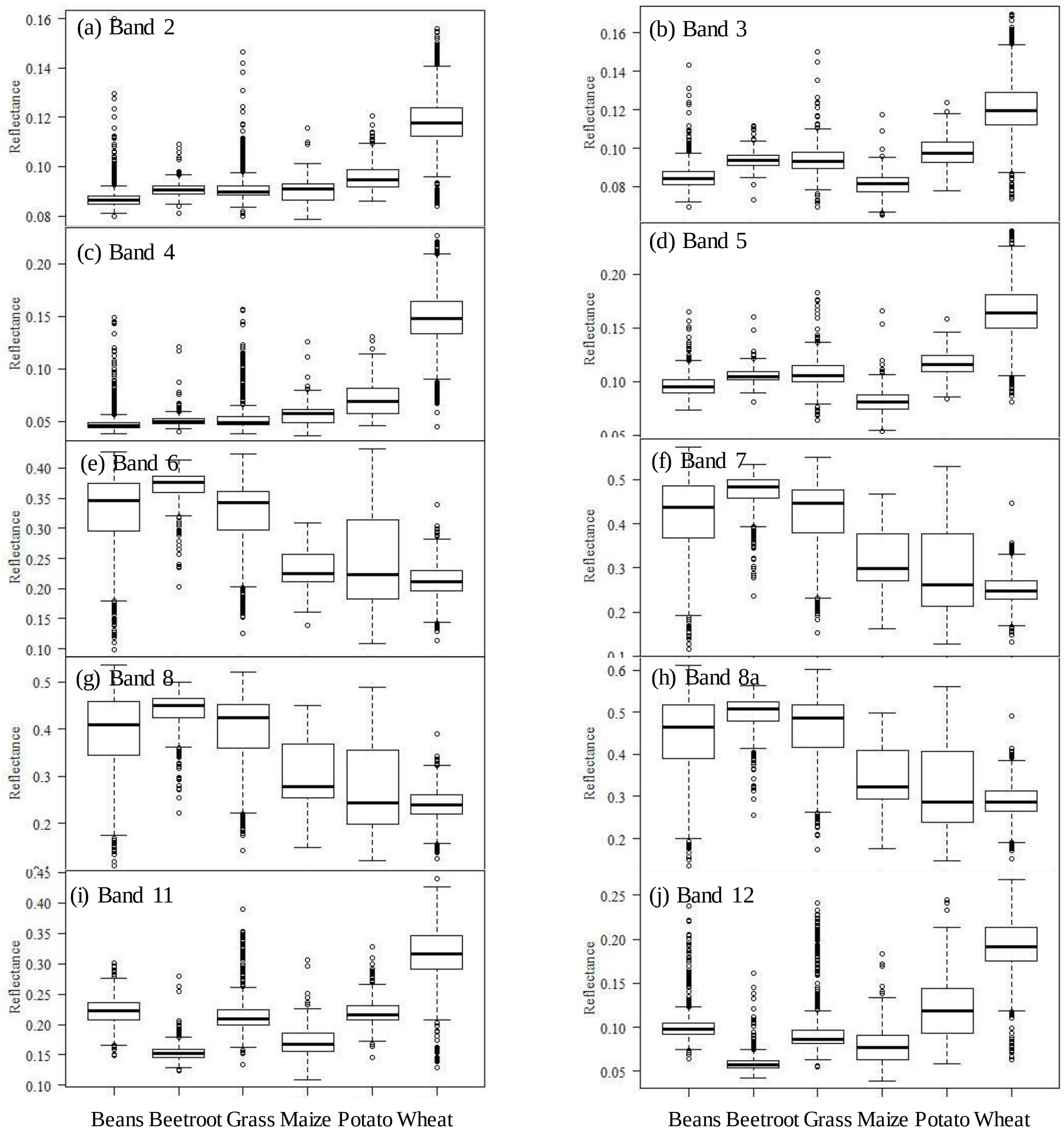

Figure 4. Boxplots of reflectance for each crop in (a) band 2, (b) band 3, (c) band 4, (d) band 5, (e) band 6, (f) band 7, (g) band 8, (h) band 8a, (i) band 11, and (j) band 12. The data for these plots were obtained from Sentinel-2A, taken on 11 August 2016.

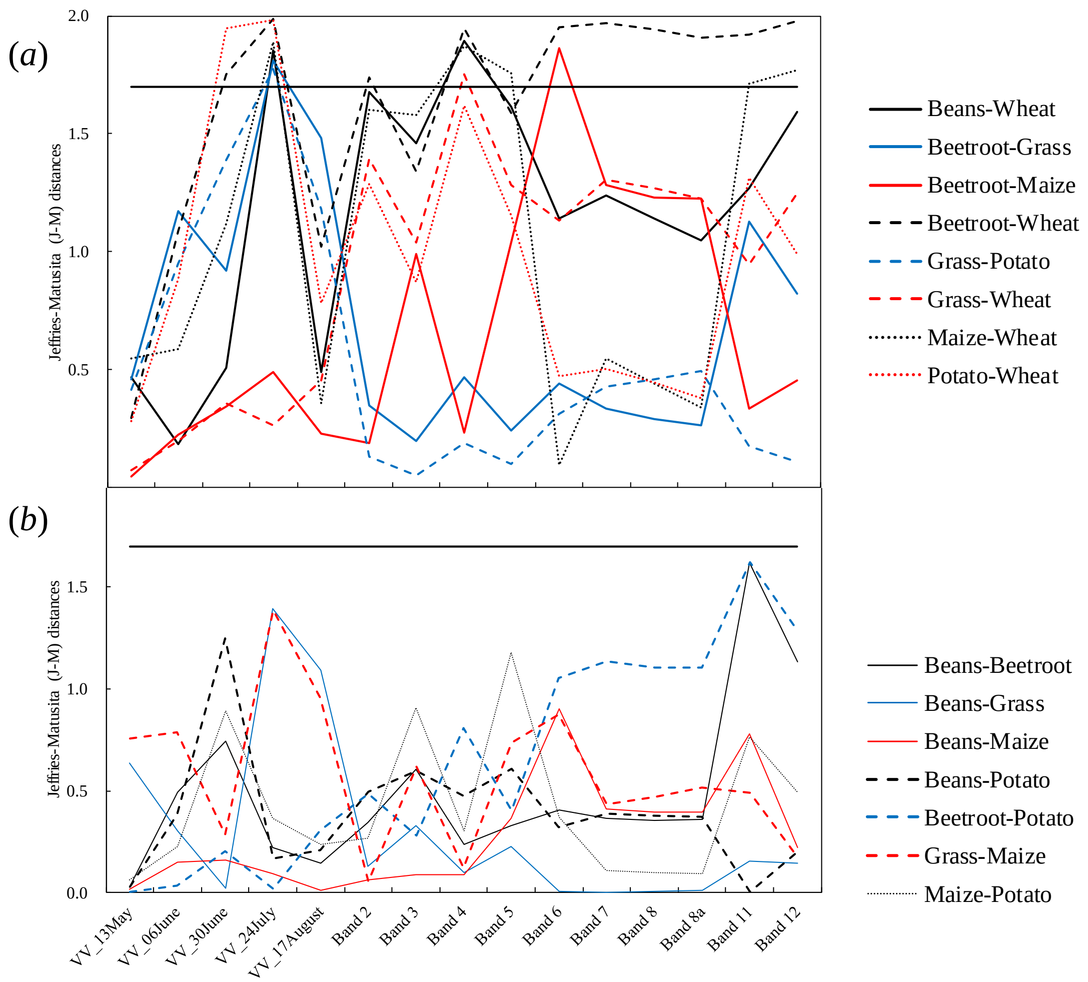

Figure 5. Jeffries-Matusita (J-M) distance values calculated for all potential crop pairs using all available data. The heavy horizontal line represents the J-M distance value of 1.7, the solid lines indicate J-M distance values greater than 1.7, and the dotted lines represent J-M distance values less than 1.7.

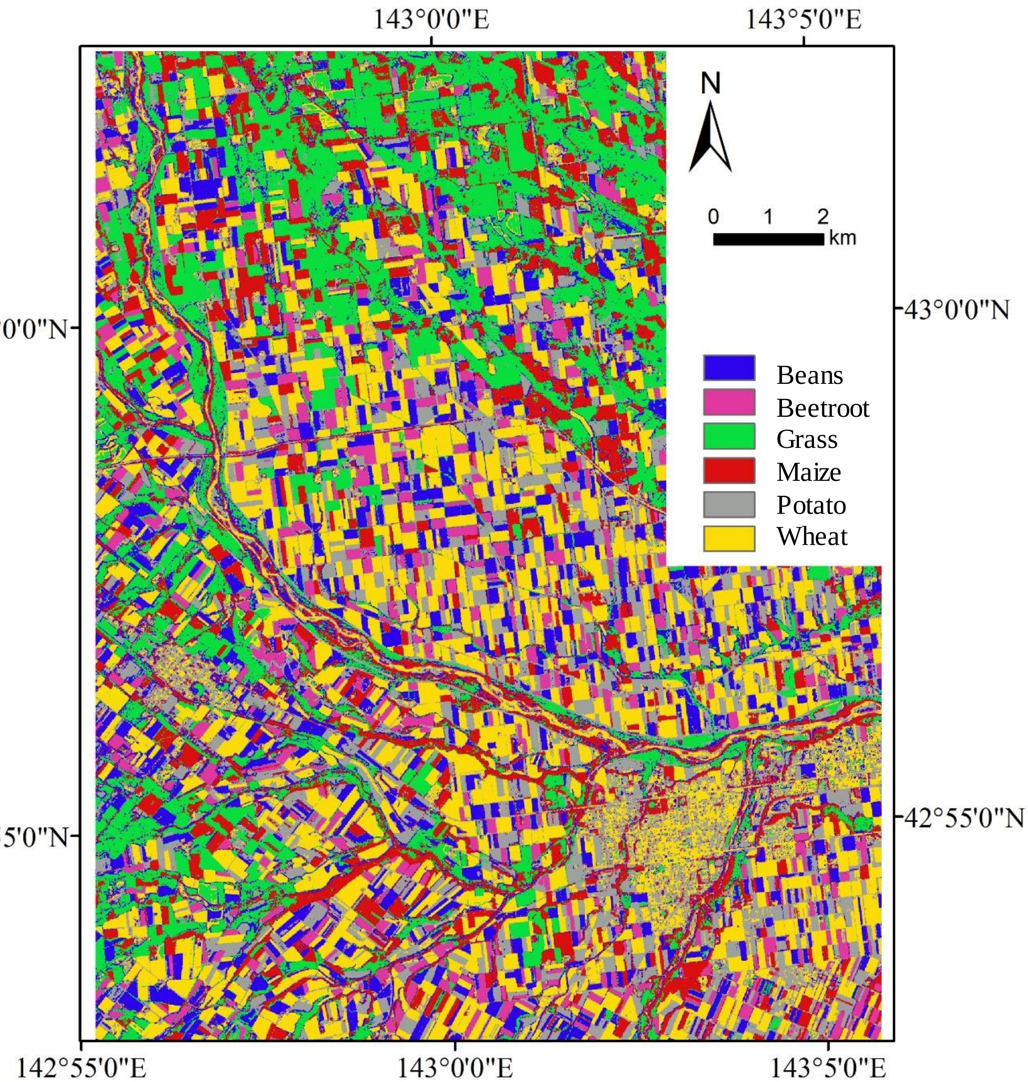

Figure 6. Crop classification map generated by KELM.

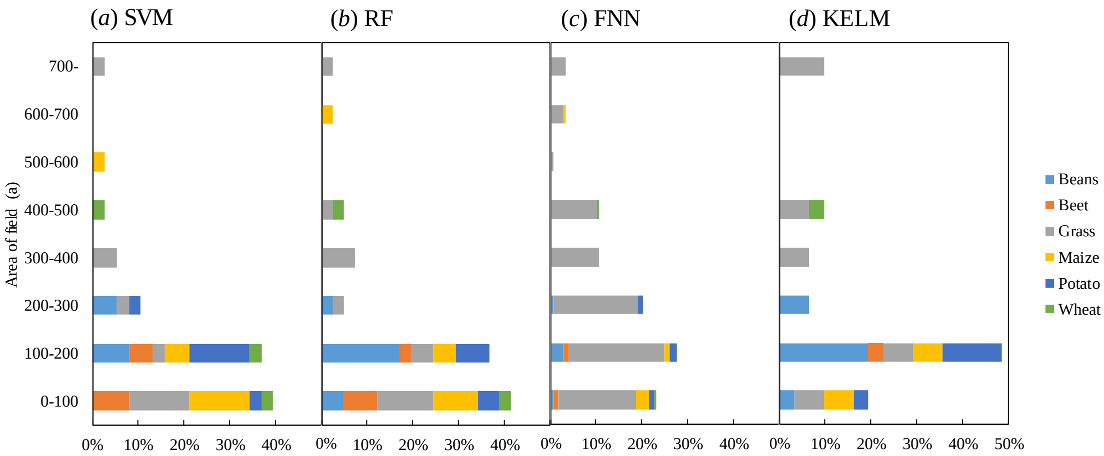

Figure 7 Relationship between field area and misclassified fields.

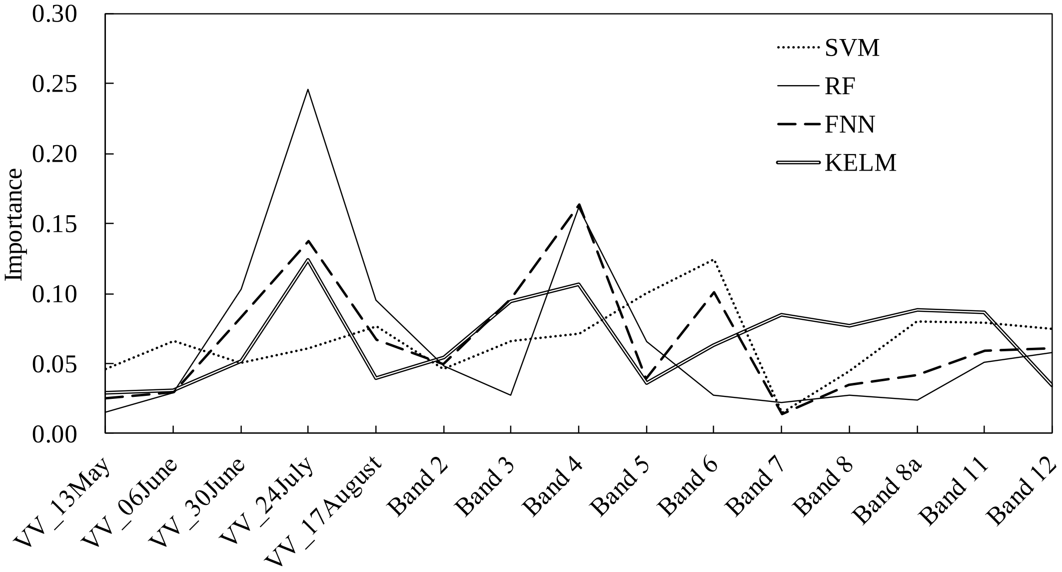

Figure 8. Data-based sensitivity analysis (DSA) results for each classification algorithm.