Crop classification from Sentinel-2 derived vegetation indices using ensemble learning

Rei Sonobe, a,∗ Yuki Yamaya, b Hiroshi Tani, c Xiufeng Wang, c Nobuyuki Kobayashi, d Kan-ichiro Mochizuki e

aShizuoka University, Faculty of Agriculture, Shizuoka, Japan, 422-8529
bHokkaido University, Graduate School of Agriculture, Sapporo, Japan, 060-5859
cHokkaido University, Research Faculty of Agriculture, Sapporo, Japan, 060-5859
dSmart Link Hokkaido, Iwamizawa, Japan, 068-0034
ePASCO Corporation, Tokyo, Japan, 153-0043

Abstract. The identification and mapping of crops are important for estimating potential harvest as well as for agricultural field management. Optical remote sensing is one of the most attractive options because it offers vegetation indices and some data have been distributed free of charge. Especially, Sentinel-2A, which is equipped with a multispectral sensor (MSI) with blue, green, red and near-infrared-1 bands at 10 m; red edge 1 to 3, near-infrared-2 and shortwave infrared 1 and 2 at 20 m; and 3 atmospheric bands (Band 1, Band 9 and Band 10) at 60 m, offers some vegetation indices calculated to assess vegetation status. However, sufficient consideration has not been given to the potential of vegetation indices calculated from MSI data. Thus, 82 published indices were calculated and their importance were evaluated for classifying crop types. In this study, the two most common classification algorithms, random forests (RF) and support vector machine (SVM), were applied to conduct cropland classification from MSI data. Additionally, super learning was applied for more improvement, achieving overall accuracies of 90.2–92.2%. Of the two algorithms applied (RF and SVM), the accuracy of SVM was superior and 89.3-92.0% of overall accuracies were confirmed. Furthermore, stacking contributed to higher overall accuracies (90.2-92.2%) and significant differences were confirmed with the results of SVM and RF. Our results showed that vegetation indices had the greatest contributions in identifying specific crop types.

Keywords: crop, random forests, Sentinel-2, stacking, support vector machine, vegetation index.

*First Author, E-mail: reysnb@gmail.com

1 Introduction

From a land-planning perspective, cropland diversity is vital and crop cover maps provide information for estimating potential harvest and agricultural field management. To document field properties such as cultivated crops and locations, some local governments in Japan have been using manual methods. However, more efficient techniques are required to reduce the high expense of these methods. Thus, satellite data-based cropland mapping has gained attention. Some spectral indices, which are combinations of spectral measurements at different wavelengths, have been used to evaluate phenology or quantify biophysical parameters. As a
result, they have also made crop maps more accurate in previous studies \(^6\) and the abilities of optical remote sensing data have been improved for monitoring agricultural fields. The opportunities to obtain optical remote sensing data have improved due to the Sentinel-2A satellite launch on June 23, 2015. Now, it is collecting multispectral data including 13 bands covering the visible, SWIR wavelength regions. Sentinel-2B, which possesses the same specifications, was launched on March 7, 2017 and creates greater opportunities for monitoring agricultural fields. Furthermore, various spectral indices can be extracted including indices based on shortwave infrared bands (SWIR), which are influenced by plant constituents such as pigments, leaf water contents and biochemistries\(^7,8\). Furthermore, vegetation indices derived from reflectance data acquired from optical sensors have been widely used to assess variations in the physiological states and biophysical properties of vegetation \(^9\)\(^\text{--}\)\(^11\). Specifically, the Normalized Difference Vegetation Index (NDVI)\(^12\), Soil-Adjusted Vegetation Index (SAVI)\(^13\) and Enhanced Vegetation Index (EVI)\(^14\) have been used for monitoring vegetation systems or ecological responses to environmental change\(^15\). MSI data have been used for identifying crop types\(^16\)\(^\text{--}\)\(^18\), plastic-covered greenhouses\(^19\), water bodies\(^20\) and some previous studies showed the potential of VIs calculated from MSI data. However, it is possible to calculate a vast number of VIs from MSI data and most of them have been ignored in the previous studies. In this study 82 published indices and original reflectance data sources were evaluated to classify six crop types including beans, beetroot, grass, maize, potato and winter wheat, which are dominant crops on the western Tokachi plain, Hokkaido, Japan.

In addition to qualities of remote sensing data, classification algorithms are important to improve classification accuracies of crop maps. Recently, random forests (RF) is a widely used machine learning algorithm consisting of an ensemble of decision trees and it has been an
extremely successful machine learning algorithm for classification and regression method\textsuperscript{21}. It has been applied for generating land cover maps\textsuperscript{22, 23} and reached around 65\% (tree species identification)\textsuperscript{17}, 76\% (crop types identification)\textsuperscript{17} and 90\% (greenhouse detection)\textsuperscript{19} using MSI data in the previous studies.

Some studies showed that support vector machine (SVM) performed better than RF for this purpose and it has been widely applied for crop for crop classification\textsuperscript{22, 24-26}. Its robustness to outliers has been demonstrated and SVM is an excellent classifier when the number of input features is large\textsuperscript{27}.

The super learner (SL) methodology\textsuperscript{25}, also called stacking, is an ensemble learning method in which the user-supplied library of algorithms is combined through a convex weighted combination, with the optimal weights to make the cross-validated empirical risk smaller. Therefore, SL could be expected to classify crop types more accurately than the single use of RF or SVM, both considered in this study. Next, an ensemble approach based on SL was applied for improving classification accuracies.

Within this framework, the main objectives of the present study were to evaluate the potential of Sentinel-2 data for crop type classification and the potential of ensemble learning based on RF and SVM.

2. Materials and Methods

2.1. Study area

The study area was located in the western part of Tokachi plain, Hokkaido, Japan (Fig.1, \textdegree 142°42’51” to \textdegree 143°08’47” E, \textdegree 42°43’20” to \textdegree 43°07’24” N). Main cultivated crops types are beans,
beetroots, grasses, maize, potatoes and winter wheat. The average monthly temperatures were 8.3–21.8°C and monthly precipitation was 12.0–94.5 mm from May to October.

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. A total of 12639 fields (2265 beans fields, 1548 beetroot fields, 2110 grasslands (timothy and orchard grass), 1000 maize fields, 2452 potato fields and 3264 winter wheat fields) were observed. The fields ranged from 0.05 ha to 18.21 ha with an averaged value of 2.54 ha. Grasslands were located on the outskirts of the built-up area.

<Fig. 1 Study area and the distribution of croplands (background map shows Sentinel-2A data obtained on August 11, 2016, R: Band 4, G: Band 3, B: Band 2).>

2.2. Remote sensing data

The data acquired from Sentinel-2 Multispectral Imager (MSI) contained blue, green, red and near-infrared-1 bands at 10 m; red edge 1 to 3, near-infrared-2 and SWIR 1 and 2 at 20 m; and 3 atmospheric bands (Band 1, Band 9 and Band 10) at 60 m. In this study, the three atmospheric bands were removed because they were dedicated to atmospheric corrections and cloud screening.

Although Sentinel-2A imagery was gathered seven times from May to September 2016 for the whole site, all images were covered with clouds except for one acquired on 11 August. The Level 1C data acquired on August 11, 2016 were downloaded from EarthExplorer (https://earthexplorer.usgs.gov/). All bands were converted to 10 m resolution with a cubic convolution resampling method and average reflectance values of each band were calculated for
each field using the field polygons to compensate for spatial variability and to avoid problems related to uncertainty in georeferencing.

Some vegetation indices such as NDVI have been used for improving classification accuracies in previous studies \(^{16, 22, 29, 30}\). Eighty-two published vegetation indices for evaluating various vegetation properties were calculated in this study (Table 1).

Table 1 Vegetation indices calculated from Sentinel-2 MSI data.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Index</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFRI1.6(^{31})</td>
<td>Aerosol free vegetation index 1.6</td>
<td>(\frac{\text{Band8a} - 0.66 \times \text{Band11}}{\text{Band8a} + 0.66 \times \text{Band12}})</td>
</tr>
<tr>
<td>AFRI2.1(^{31})</td>
<td>Aerosol free vegetation index 2.1</td>
<td>(\frac{\text{Band8a} - 0.5 \times \text{Band12}}{\text{Band8a} + 0.5 \times \text{Band12}})</td>
</tr>
<tr>
<td>ARI(^{32})</td>
<td>Anthocyanin reflectance index</td>
<td>(\frac{\text{Band8a} - 0.66 \times \text{Band11}}{\text{Band8a} + 0.66 \times \text{Band12}})</td>
</tr>
<tr>
<td>ARVI(^{33})</td>
<td>Atmospherically resistant vegetation index</td>
<td>(\frac{\text{Band8a} - 0.5 \times \text{Band12}}{\text{Band8a} + 0.5 \times \text{Band12}})</td>
</tr>
<tr>
<td>ARVI2(^{33})</td>
<td>Atmospherically resistant vegetation index 2</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>ATSAVI(^{34})</td>
<td>Adjusted transformed soil-adjusted vegetation index</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>AVI(^{35})</td>
<td>Ashburn vegetation index</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>BNDVI(^{36})</td>
<td>Blue-normalized difference vegetation index</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>BRI(^{37})</td>
<td>Browning reflectance index</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>BWDRVI(^{38})</td>
<td>Blue-wide dynamic range vegetation index</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>CARI(^{39})</td>
<td>Chlorophyll absorption ratio index</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>CCCI(^{40})</td>
<td>Canopy chlorophyll content index</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>CRI550(^{41})</td>
<td>Carotenoid reflectance index 550</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>CRI700(^{41})</td>
<td>Carotenoid reflectance index 700</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
<tr>
<td>CVI(^{42})</td>
<td>Chlorophyll vegetation index</td>
<td>(\frac{\text{Band8a} - 0.18 \times \text{Band11}}{\text{Band8a} + 0.18 \times \text{Band11}})</td>
</tr>
</tbody>
</table>
Datt1\textsuperscript{43}  Vegetation index proposed by Datt 1
Datt2\textsuperscript{44}  Vegetation index proposed by Datt 2
Datt3\textsuperscript{44}  Vegetation index proposed by Datt 3
DVI\textsuperscript{45}  Differentiated vegetation index
EPIcar\textsuperscript{44}  Eucalyptus pigment index for carotenoid
EPIChla\textsuperscript{44}  Eucalyptus pigment index for chlorophyll a
EPIChlab\textsuperscript{44}  Eucalyptus pigment index for chlorophyll b
EPIChlb\textsuperscript{44}  Enhanced vegetation index
EV1\textsuperscript{44}  Enhanced vegetation index 2
EV12\textsuperscript{46}  Enhanced vegetation index 2.2
EV1.2\textsuperscript{47}  Green atmospherically resistant vegetation index
GBNDVI\textsuperscript{49}  Green-Blue normalized difference vegetation index
GDVI\textsuperscript{50}  Green difference vegetation index
GEMI\textsuperscript{51}  Global environment monitoring index
GLI\textsuperscript{52}  Green leaf index
GNDVI\textsuperscript{48}  Green normalized difference vegetation index
GNDVI2\textsuperscript{48}  Green normalized difference vegetation index 2
GOSAVI\textsuperscript{53}  Green optimized soil adjusted vegetation index
GRNDVI\textsuperscript{54}  Green-Red normalized difference vegetation index
GVMI\textsuperscript{55}  Global vegetation moisture index
Hue\textsuperscript{56}  Hue
IPVI\textsuperscript{57}  Infrared percentage vegetation index
LCl\textsuperscript{53}  Leaf chlorophyll index

\begin{align*}
\text{Band8} - \text{Band5} \\
\frac{\text{Band8} - \text{Band4}}{\text{Band4}} \\
\frac{\text{Band3} \ast \text{Band5}}{\text{Band8a}} \\
\frac{\text{Band3} \ast \text{Band5}}{\text{Band4}} \\
2.4 \ast \frac{\text{Band8} - \text{Band4}}{\text{Band4}} \\
0.0049 \ast \left( \frac{\text{Band4}}{\text{Band3} \ast \text{Band5}} \right)^{0.7488} \\
0.0161 \ast \left( \frac{\text{Band4}}{\text{Band3} \ast \text{Band5}} \right)^{0.7784} \\
0.0236 \ast \left( \frac{\text{Band4}}{\text{Band3} \ast \text{Band5}} \right)^{0.7954} \\
0.0337 \ast \left( \frac{\text{Band4}}{\text{Band3}} \right)^{1.8695} \\
\frac{\text{Band8} - \text{Band4}}{2.5 \ast \left( \frac{\text{Band8} + 6 \ast \text{Band4} - 7.5 + \text{Band2} + 1}{\text{Band8} - \text{Band4}} \right)} \\
2.4 \ast \left( \frac{\text{Band8} + \text{Band4} + 1}{\text{Band8} - \text{Band4}} \right) \\
2.5 \ast \left( \frac{\text{Band8} + 2.4 \ast \text{Band4} + 1}{\text{Band8} - \left( \text{Band3} - \left( \text{Band2} - \text{Band4} \right) \right)} \right) \\
\frac{\text{Band8} - \left( \text{Band3} + \text{Band2} \right)}{\text{Band8} + \left( \text{Band3} + \text{Band2} \right)} \\
\frac{\text{Band8} - \text{Band3}}{\text{n} \ast \left( 1 - 0.25 \ast \text{n} \right) - \text{Band4} - 0.125} \\
n = \frac{1 - \text{Band4}}{2 \ast \left( \text{Band5}^2 - \text{Band4}^2 + 1.5 \ast \text{Band5} + 0.5 \ast \text{Band4} \right)} \\
\frac{\text{Band5} + \text{Band4} + 0.5}{2 \ast \left( \text{Band3} - \text{Band5} - \text{Band2} \right)} \\
2 \ast \left( \text{Band3} + \text{Band5} + \text{Band2} \right) \\
\frac{\text{Band8} - \text{Band3}}{\text{Band8} + \text{Band3}} \\
\frac{\text{Band8} - \text{Band3}}{\text{Band8} + \text{Band3}} \\
\frac{\text{Band8} - \text{Band3}}{\text{Band8} - \text{Band3}} \\
\frac{\text{Band8} - \left( \text{Band3} + \text{Band5} \right)}{\text{Band8} + \left( \text{Band3} + \text{Band5} \right)} \\
\frac{\left( \text{Band8} + 0.1 \right) - \left( \text{Band12} + 0.02 \right)}{\left( \text{Band8} + 0.1 \right) + \left( \text{Band12} + 0.02 \right)} \\
\text{atan}\left( \frac{\text{Band8} - \left( 2 \ast \text{Band5} - \text{Band3} - \text{Band2} \right)^{30.5}}{\left( \text{Band7} + \text{Band3} \right)} \right) \\
\frac{\text{Band8} + \text{Band5} - \text{Band4} + \text{Band5}}{2} \\
\frac{\left( \text{Band5} + \text{Band5} \right) + 1}{\text{Band8} + \text{Band5}} \\
\frac{\left( \text{Band5} - \text{Band3} \right) + \text{Band2} + \text{Band4}}{30.5} \\
\text{atan}\left( \frac{\text{Band8} - \left( 2 \ast \text{Band5} - \text{Band3} - \text{Band2} \right)^{30.5}}{\left( \text{Band7} + \text{Band3} \right)} \right) \\
\frac{\text{Band8} + \text{Band5} - \text{Band4} + \text{Band5}}{2}
<table>
<thead>
<tr>
<th>Index Name</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCARI$^{59}$</td>
<td>$\frac{\text{Band7} - \text{Band5}}{\text{Band7} - \text{Band4}}$</td>
</tr>
<tr>
<td>MCARI/MTVI2$^{60}$</td>
<td>$\frac{(\text{Band5} - \text{Band4}) - 0.2 \times (\text{Band5} - \text{Band3})}{\text{Band5} - \text{Band4}}$</td>
</tr>
<tr>
<td>MCARI/OSAVI$^{61}$</td>
<td>$\frac{\text{MCARI/MTVI2}}{\text{MCARI/OSAVI}}$</td>
</tr>
<tr>
<td>MCARI1$^{61}$</td>
<td>$1.2 \times (2.5 \times (\text{Band8} - \text{Band4}) - 1.3 \times (\text{Band8} - \text{Band3}))$</td>
</tr>
<tr>
<td>MCARI2$^{61}$</td>
<td>$\frac{1.5 \times \frac{2.5 \times (\text{Band8} - \text{Band4}) - 1.3 \times (\text{Band8} - \text{Band3})}{\sqrt{(2 \times \text{Band8} + 1)^2 - (6 \times \text{Band8} - 5 \times \sqrt{\text{Band4}}) - 0.5}}}{\text{Band8} - \text{Band4}} - 2 \times \text{Band2}$</td>
</tr>
<tr>
<td>MGVI$^{62}$</td>
<td>$\frac{\text{Non such index proposed by Misra}}{\text{Modified normalized difference vegetation index}}$</td>
</tr>
<tr>
<td>mNDVI$^{63}$</td>
<td>$\frac{\text{Band8} + \text{Band4} - 2 \times \text{Band2}}{\text{Band8} + \text{Band4} - 2 \times \text{Band2}}$</td>
</tr>
<tr>
<td>MNSI$^{62}$</td>
<td>$0.404 \times \text{Band3} + 0.039 \times \text{Band4} - 0.505 \times \text{Band6} + 0.762 \times \text{Band8}$</td>
</tr>
<tr>
<td>MSAVI$^{64}$</td>
<td>$\frac{2 \times \text{Band8} + 1 - \sqrt{(2 \times \text{Band8} + 1)^2 - 8 \times (\text{Band8} - \text{Band5})}}{2}$</td>
</tr>
<tr>
<td>MSAVI2$^{64}$</td>
<td>$\frac{2 \times \text{Band8} + 1 - \sqrt{(2 \times \text{Band8} + 1)^2 - 8 \times (\text{Band8} - \text{Band4})}}{2}$</td>
</tr>
<tr>
<td>MSBI$^{62}$</td>
<td>$0.406 \times \text{Band3} + 0.600 \times \text{Band4} + 0.645 \times \text{Band6} + 0.243 \times \text{Band8}$</td>
</tr>
<tr>
<td>MSR670$^{65}$</td>
<td>$\frac{\text{Modified simple ratio}}{\text{Nir/Red}}$</td>
</tr>
<tr>
<td>MSRNir/Red$^{66}$</td>
<td>$\frac{\text{Band8} - \text{Band4} - 1}{\sqrt{\frac{\text{Band8} - \text{Band4} - 1}{\text{Band8} - \text{Band4} - 1}}} + 1$</td>
</tr>
<tr>
<td>MTVI2$^{61}$</td>
<td>$\frac{\text{Modified triangular vegetation index 2}}{\text{Nir/Swir normalized burn ratio}}$</td>
</tr>
<tr>
<td>NBR$^{67}$</td>
<td>$\frac{\text{Band8} - \text{Band12}}{\text{Band8} + \text{Band12}}$</td>
</tr>
<tr>
<td>NBR$^{68}$</td>
<td>$\frac{\text{Band7} - \text{Band4}}{\text{Band7} - \text{Band4}}$</td>
</tr>
<tr>
<td>ND774/677$^{68}$</td>
<td>$\frac{\text{Band7} + \text{Band4}}{\text{Band7} - \text{Band4}}$</td>
</tr>
<tr>
<td>NDII$^{69}$</td>
<td>$\frac{\text{Band8} + \text{Band11}}{\text{Band8} + \text{Band11}}$</td>
</tr>
<tr>
<td>NDRE$^{70}$</td>
<td>$\frac{\text{Band7} - \text{Band5}}{\text{Band7} - \text{Band5}}$</td>
</tr>
<tr>
<td>NDSI$^{71}$</td>
<td>$\frac{\text{Band11} + \text{Band12}}{\text{Band8} - \text{Band12}}$</td>
</tr>
<tr>
<td>NDVI$^{12}$</td>
<td>$\frac{\text{Band8} + \text{Band4}}{\text{Band8} - \text{Band4}}$</td>
</tr>
<tr>
<td>NDVI2$^{50}$</td>
<td>$\frac{\text{Band12} + \text{Band8}}{\text{Band8} + \text{Band4}}$</td>
</tr>
<tr>
<td>NGRDI$^{68}$</td>
<td>$\frac{\text{Band3} + \text{Band5}}{\text{Band8} - \text{Band4}}$</td>
</tr>
<tr>
<td>OSAVI$^{53, 72}$</td>
<td>$\frac{\text{Optimized soil adjusted vegetation index}}{\text{Nir/Swir normalized burn ratio}}$</td>
</tr>
<tr>
<td>OSAVI$^{53, 72}$</td>
<td>$1.16 \times \frac{\text{Band8} + \text{Band4} + 0.16}{\text{Band8} + \text{Band4}}$</td>
</tr>
</tbody>
</table>
### 2.3. Classification algorithm

All samples were divided into the following three groups using a stratified random sampling approach: training data (50%) for developing classification models, validation data (25%) for hyperparameter tuning and test data (25%) for evaluation of classification accuracies and table 2 shows the numbers of fields of each crop type.
Table 2 Crop type and number of fields.

<table>
<thead>
<tr>
<th>Crop type</th>
<th>Training data</th>
<th>Validation data</th>
<th>Test data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beans</td>
<td>1132</td>
<td>566</td>
<td>567</td>
</tr>
<tr>
<td>Beetroot</td>
<td>774</td>
<td>387</td>
<td>387</td>
</tr>
<tr>
<td>Grassland</td>
<td>1055</td>
<td>527</td>
<td>528</td>
</tr>
<tr>
<td>Maize</td>
<td>500</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>Potato</td>
<td>1226</td>
<td>613</td>
<td>613</td>
</tr>
<tr>
<td>Wheat</td>
<td>1632</td>
<td>816</td>
<td>816</td>
</tr>
</tbody>
</table>

SVM partitions data using maximum separation margins\textsuperscript{86} and the ‘kernel trick’ has frequently been applied instead of attempting to fit a non-linear model in previous studies\textsuperscript{29}. In this study, the Gaussian Radial Basis Function (RBF) kernel, which has mostly been used for classification purposes\textsuperscript{29}, was used as a kernel and two parameters were tuned to control the flexibility of the classifier, the regularization parameter $C$ and the kernel bandwidth $\gamma$. If the $C$ value is too large, there is a high penalty for no separable points and we may store many support vectors and overfit. If it is too small, there may be under-fitting. It controls the trade-off between errors of the SVM on training data and margin maximization ($C = \infty$ leads to hard margin SVM). The $\gamma$ value defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close.’

RF is an ensemble learning technique composed of multiple decision trees based on random bootstrapped samples of the training data\textsuperscript{87}. The output is determined by a majority vote of the results of decision trees. There are two user-defined hyperparameters including the number of trees ($ntree$) and the number of variables used to split the nodes ($mtry$). If $ntree$ is made larger, the generalization error always converges, and over-training will not be a problem. On the other hand, a reduction in $mtry$ makes each individual decision tree weaker.
The best combinations of these hyperparameters were determined using the Gaussian process, Bayesian optimization\textsuperscript{88}, which has been widely applied for hyperparameter tuning of machine learning algorithms\textsuperscript{1}.

Ensemble machine learning methods have been used to obtain better predictive performance than from single learning algorithms and the SL methodology has been proposed\textsuperscript{89}. In this method, given algorithms are combined through a convex weighted combination to minimize cross-validated errors. First, classification models based on RF or SVM were trained as the base algorithms using the training data. Next, a ten-fold cross-validation was performed on each and the cross-validated predicted results were obtained. \(N\) is the number of rows in the training data, cross-validated predicted results were combined and an \(N\) by two matrix was obtained as the “level-one” data and meta-learning model was generated. To predict the test data, the predictions from the base learners were feed into the meta-learning model to generate the ensemble prediction.

The data-based sensitivity analysis (DSA)\textsuperscript{90}, which performs a pure black box use of the fitted models by querying the fitted models with sensitivity samples and recording their responses, was applied for assessing the sensitivity of the classification models.

2.4. Accuracy assessment

Classification accuracies were evaluated based on the simple measures of quantity disagreement (QD) and allocation disagreement (AD)\textsuperscript{91}. They provide an effective summary of confusion matrices\textsuperscript{92}.

The proportion of fields that are classified as crop \(i\) and their actual classes are crop \(j\) \((P_{ij})\) is expressed in the following equation (1):

\[
P_{ij} = W_i \frac{n_{ij}}{n_{i+}}
\]
where $W_i$ are the fields classified as crop $i$, $n_{ij}$ is the number of fields classified as crop $i$ and their actual classes are crop $j$. $n_{i+}$ is the row totals of the confusion matrix. In this case, AD and QD are calculated using the following equations (2–5):

$$AD_i = 2 \min(p_{i+}, p_{+i}) - 2p_{ii} \quad (2)$$

$$AD = \frac{1}{2} \sum_{i=1}^{N_c} AD_i \quad (3)$$

$$QD_i = |p_{i+} - p_{+i}| \quad (4)$$

$$QD = \frac{1}{2} \sum_{i=1}^{N_c} QD_i \quad (5)$$

where $N_c$ is the number of classes (six in this study), $p_{i+}$ and $p_{+i}$ are the row and column totals of the confusion matrix, $AD_i$ is the allocation disagreement of crop $i$ and $QD_i$ is the quantity disagreement of crop $i$. The sum of $QD_i$ (QD) and $AD_i$ (AD) are calculated and the total disagreement can be evaluated by the sum of QD and AD.91

In addition, three indicators including overall accuracy (OA, equations (6)), producer’s accuracy (PA, equations (7)) and user’s accuracy (UA, equations (8)) were calculated because they have widely been applied for assessing classification accuracies.

$$OA = \frac{\sum_{i=1}^{N} p_{ii}}{N} \quad (6)$$

$$PA = \frac{p_{ii}}{R_i} \quad (7)$$

$$UA = \frac{p_{ii}}{C_i} \quad (8)$$

where $N$ is the number of fields, $R_i$ and $C_i$ represent the total number of crop $i$ in the correct data and the total number from the classification results, respectively. McNemar’s test93 has been used to judge whether the differences between two given classification results were significant94 and it was also applied in this study.
3. Results and Discussion

3.1. Classification accuracy

Crop classification maps are shown in Fig. 2, the maximum, minimum and averaged accuracies of ten repetitions and confusion matrices when all the repetitions were merged are shown in Table 3 and 4. Averaged OAs were 89.0% for RF, 90.6% for SVM and 91.6% for the ensemble machine learning method and the mean PAs and mean UAs derived using the machine learning algorithms were greater than 0.8, excepting those of RF (mean UA for maize was 0.797). All machine learning algorithms performed well in classifying croplands. Especially, the good accuracies were confirmed for the PAs and UAs for wheat (more than 93.8%) and beet (more than 89.9%). However, the chi-square values based on McNemar’s tests were 12.02 – 40.60, 27.78 – 62.43 and 17.00 – 51.60 for R – SVM, RF – SL and SVM – SL, respectively. As the results, significant differences were confirmed among the results of three machine learning algorithms ($p < 0.05$).

Classification results by SL had the best OA and AD+QD (8.5%) and SVM had a slightly better PA of wheat (97.1%). On the contrary, identifying maize fields was difficult due to the similarity in their reflectance. Grasses cultivation employs fewer controls and then a lot of weeds were mixed with timothy and orchard grass in grasslands. As a result, variation in reflectance features were larger than in other crop types, causing misclassifications of relatively larger fields.

<Fig. 2. Crop classification map generated by (a) RF, (b) SVM and (c) SL.>
Table 3 Classification accuracies of each algorithm.

<table>
<thead>
<tr>
<th></th>
<th>RF</th>
<th></th>
<th></th>
<th>SVM</th>
<th></th>
<th></th>
<th>SL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Maximum</td>
<td>Mean±std</td>
<td>Minimum</td>
<td>Maximum</td>
<td>Mean±std</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>PA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beans</td>
<td>80.6%</td>
<td>86.4%</td>
<td>83.4±1.6%</td>
<td>81.1%</td>
<td>90.5%</td>
<td>86.2±2.2%</td>
<td>84.7%</td>
<td>90.3%</td>
</tr>
<tr>
<td>Beet</td>
<td>89.9%</td>
<td>94.8%</td>
<td>93.0±1.3%</td>
<td>91.0%</td>
<td>96.4%</td>
<td>94.5±1.5%</td>
<td>93.8%</td>
<td>96.1%</td>
</tr>
<tr>
<td>Grassland</td>
<td>84.3%</td>
<td>88.3%</td>
<td>86.0±1.2%</td>
<td>86.7%</td>
<td>93.8%</td>
<td>89.4±2.5%</td>
<td>89.8%</td>
<td>94.3%</td>
</tr>
<tr>
<td>Maize</td>
<td>78.8%</td>
<td>84.8%</td>
<td>80.8±1.7%</td>
<td>78.8%</td>
<td>87.6%</td>
<td>83.0±3.1%</td>
<td>81.2%</td>
<td>87.6%</td>
</tr>
<tr>
<td>Potato</td>
<td>82.9%</td>
<td>89.7%</td>
<td>87.0±1.8%</td>
<td>83.5%</td>
<td>89.9%</td>
<td>87.6±1.9%</td>
<td>84.0%</td>
<td>89.7%</td>
</tr>
<tr>
<td>Wheat</td>
<td>96.4%</td>
<td>97.9%</td>
<td>97.0±0.5%</td>
<td>96.3%</td>
<td>97.5%</td>
<td>97.1±0.4%</td>
<td>95.7%</td>
<td>97.5%</td>
</tr>
<tr>
<td>UA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beans</td>
<td>84.9%</td>
<td>88.6%</td>
<td>86.8±1.1%</td>
<td>82.0%</td>
<td>91.4%</td>
<td>86.4±2.9%</td>
<td>83.4%</td>
<td>90.3%</td>
</tr>
<tr>
<td>Beet</td>
<td>94.5%</td>
<td>96.9%</td>
<td>95.6±0.8%</td>
<td>94.3%</td>
<td>97.3%</td>
<td>95.7±0.9%</td>
<td>95.1%</td>
<td>97.1%</td>
</tr>
<tr>
<td>Grassland</td>
<td>88.0%</td>
<td>93.3%</td>
<td>91.0±1.4%</td>
<td>89.9%</td>
<td>96.6%</td>
<td>94.0±2.3%</td>
<td>93.8%</td>
<td>97.7%</td>
</tr>
<tr>
<td>Maize</td>
<td>77.8%</td>
<td>82.0%</td>
<td>79.7±1.3%</td>
<td>78.4%</td>
<td>87.3%</td>
<td>81.9±2.2%</td>
<td>81.4%</td>
<td>85.2%</td>
</tr>
<tr>
<td>Potato</td>
<td>78.5%</td>
<td>83.1%</td>
<td>81.5±1.2%</td>
<td>82.1%</td>
<td>87.8%</td>
<td>85.2±1.9%</td>
<td>83.0%</td>
<td>86.8%</td>
</tr>
<tr>
<td>Wheat</td>
<td>93.8%</td>
<td>96.1%</td>
<td>95.0±0.7%</td>
<td>94.5%</td>
<td>97.2%</td>
<td>95.9±0.8%</td>
<td>95.1%</td>
<td>97.2%</td>
</tr>
<tr>
<td>OA</td>
<td>88.5%</td>
<td>89.4%</td>
<td>89.0±0.2%</td>
<td>89.3%</td>
<td>92.0%</td>
<td>90.6±0.9%</td>
<td>90.2%</td>
<td>92.2%</td>
</tr>
<tr>
<td>κ</td>
<td>85.9%</td>
<td>87.0%</td>
<td>86.5±0.3%</td>
<td>86.8%</td>
<td>90.2%</td>
<td>88.4±1.1%</td>
<td>88.0%</td>
<td>90.5%</td>
</tr>
<tr>
<td>AD</td>
<td>8.0%</td>
<td>9.9%</td>
<td>9.0±0.6%</td>
<td>6.5%</td>
<td>9.7%</td>
<td>7.9±1.0%</td>
<td>6.5%</td>
<td>8.8%</td>
</tr>
<tr>
<td>QD</td>
<td>1.3%</td>
<td>2.8%</td>
<td>2.0±0.5%</td>
<td>0.7%</td>
<td>2.5%</td>
<td>1.5±0.6%</td>
<td>0.6%</td>
<td>2.3%</td>
</tr>
</tbody>
</table>

Table 4 Confusion matrices for (a) RF, (b) SVM and (c) SL.

(a) RF

<table>
<thead>
<tr>
<th></th>
<th>Reference data</th>
<th>Beans</th>
<th>Beetroot</th>
<th>Grasslands</th>
<th>Maize</th>
<th>Potato</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classified data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beans</td>
<td>4726</td>
<td>59</td>
<td>247</td>
<td>100</td>
<td>287</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Beet</td>
<td>48</td>
<td>3599</td>
<td>23</td>
<td>28</td>
<td>65</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Grasslands</td>
<td>172</td>
<td>65</td>
<td>4543</td>
<td>52</td>
<td>116</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Maize</td>
<td>139</td>
<td>21</td>
<td>128</td>
<td>2019</td>
<td>177</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Potato</td>
<td>503</td>
<td>119</td>
<td>230</td>
<td>235</td>
<td>5332</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>82</td>
<td>7</td>
<td>109</td>
<td>66</td>
<td>153</td>
<td>7919</td>
<td></td>
</tr>
</tbody>
</table>

(b) SVM

<table>
<thead>
<tr>
<th></th>
<th>Reference data</th>
<th>Beans</th>
<th>Beetroot</th>
<th>Grasslands</th>
<th>Maize</th>
<th>Potato</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classified data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beans</td>
<td>4888</td>
<td>77</td>
<td>212</td>
<td>119</td>
<td>333</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Beet</td>
<td>61</td>
<td>3659</td>
<td>17</td>
<td>22</td>
<td>63</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Grasslands</td>
<td>110</td>
<td>34</td>
<td>4720</td>
<td>40</td>
<td>70</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Maize</td>
<td>112</td>
<td>14</td>
<td>130</td>
<td>2076</td>
<td>166</td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3 shows the relationship between field area and misclassified fields for each algorithm after ten repetitions (i.e. the total number is ten times of that of the test data). More than 75% of the misclassified fields were less than 200 a in area for all algorithms, and 95.1% (RF), 95.5% (SVM) and 96.1% (SL) of misclassified fields were below 450 a. Applying stacking made the model more robust for classifying smaller fields and the number of misclassified croplands decreased (813 fields for smaller than 50 a) compared with the results by RF (909 fields for smaller than 50 a) and SVM (855 fields for smaller than 50 a). It was especially useful for identifying beans fields. It was not effective for identifying small grasslands since grass cultivation employs fewer controls and many weeds were present in grasslands. However, stacking was useful for identifying grasslands more than 500 a, which had a certain homogeneity with *Dactylis glomerata* or *Phleum pretense* in the MSI image.

3.2. Sensitive factor analysis

Reflectance values obtained from Sentinel-2A are shown in Fig. 4. Differences in reflectance were particularly clear between wheat and beans since the wheat harvest was finished on 11
August and the reflectance of wheat fields was similar to that of bare soil. Beetroot had the steepest gradient between Bands 5 and 6 and some differences in the reflectance values at Band 11 were confirmed between maize and potato. Differences in the reflectance patterns between grass and beans were not clear.

<Fig. 4 Reflectance spectra of each crop.>

To clarify which variables contributed to identifying each crop type, DSA was conducted for each algorithm and their importance values were calculated.

For identifying beans fields, Datt3 (6.0%, 6.6% and 6.3% for RF, SVM and SL, respectively) and REIP (6.4%, 8.2% and 7.3% for RF, SVM and SL, respectively) played important roles in the three algorithms. Some variables (the reflectance values at Bands 2 and 3, AFRI2.1, CVI and NDSI) possessed importance values of more than 5.0% in the RF-based model, while no variables except for Datt3 and REIP had importance values of more than 5.0% for SVM and SL. Even though the importance values of GEMI, Maccioni and MNSI in SVM were less than 5.0%, they were more than 5 times those in RF. AFRI1.6 and SIWSI were useful for identifying beetroot fields and AFRI1.6 occupied 11.1%, 6.8% and 9.0% and SIWSI occupied 10.6%, 7.1% and 8.9% of the importance for RF, SVM and SL, respectively. GEMI and NDSI also had importance values of more than 10% for RF, but were less than 5% for the others. In contrast, REIP was useful in SVM and it occupied 9.1% of the importance in SVM. AFRI1.6, REIP and MNSI were effective for identifying grassland for all algorithms, while SIWSI played an important role (7.8%) for RF and the reflectance at Band 6 played an important role (8.2%) for SVM. For identifying maize fields, no variable had importance values more than 5.0% for any algorithm, but the importance value of REIP was 25.3% for SVM (2.9% for RF). CRI550, CRI700 and MSBI were 9.1%, 12.9% and 5.6% in RF, respectively (those in SVM were 2.4%,
2.2% and 3.6%, respectively). REIP played the greatest role for identifying potato fields in all algorithms (12.8%, 6.9% and 9.9% for RF, SVM and SL, respectively). The importance values of CCCI and CVI were also high in RF (9.9%) but those in SVM were less than 3.0%. In contrast, Maccioni had an importance of 6.9% in SVM but in RF was 1.4%. REIP also played a great role for identifying wheat fields in SVM but 1.2% of the importance value was confirmed in RF while AVI occupied 15.1% in RF (1.2% in SVM). However, the original reflectance values possessed importance values of less than 1.0%.

In this season, the photosynthetic activities of each crop type were different; maize is a C4 plant, beans and beetroot were in their growing season, grassland was after second harvest, potato growth was inhibited by chemicals for easy harvesting and wheat fields were cultivated. Besides indices related to chlorophyll content, the additional use of shortwave infrared data contributed to the estimation of photosynthetic pigments, water, nitrogen, cellulose, lignin, phenols, and leaf mass per area (e.g. NDSI). As a result, vegetation indices had greater influence on the classification results than the original reflectance. However, there were differences among algorithms in which vegetation indices were more important. The importance values in SL were near the averaged values of RF and SVM. So, the differences in importance between RF and SVM were useful when stacking was applied, and the modification contributed to identifying croplands with higher accuracies.

4. Conclusions and future work

Cropland classifications were conducted using a single image from Sentinel-2 MSI and the suitability and accuracy of vegetation indices from the original reflectance data from Sentinel-2 MSI were assessed.
Of the two algorithms applied (RF and SVM), the accuracy of SVM was superior and 89.3–92.0% of OAs were confirmed. Furthermore, stacking contributed to higher OAs (90.2–92.2%) and significant differences were confirmed with the results of SVM. Based on DSA, the vegetation indices calculated from the original reflectance from Sentinel-2 MSI data were useful to identify the specific crop types. Although the vegetation indices that played the largest roles were different between RF and SVM, stacking helped to modify and reduce the importance of specific variables, which might prevent overfitting. Stacking should be utilized to monitor agricultural fields for improving classification accuracies.

The field is used as a basic unit in classification and some problems related to the borders of fields remain to be resolved. We are planning to evaluate the potential of geographic object-based image analysis in conjunction with MSI data and address this question in future work.

Disclosures

No potential conflicts of interest are reported by the authors.

Acknowledgments

The authors would like to thank Tokachi Nosai for providing the field data.

References


19


67. C. Key, and N. Benson, "Landscape assessment: Ground measure of severity; The Composite Burn Index, and remote sensing of severity, the Normalized Burn Index. Landscape assessment: Ground measure of severity; The Composite Burn Index, and remote sensing of severity, the Normalized Burn Index. ," in FIREMON: Fire Effects Monitoring and Inventory System, D. Lutes, R. Keane, J. Caratti, C. Key, N. Benson, S. Sutherland, and L. Gangi, eds. (Rocky Mountains Research Station, USDA Forest Service, Fort Collins, CO, USA, 2005), pp. 1-51.


93. Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika* 12, 153-157 (1947).

**Caption List**

**Fig. 1** Study area and the distribution of croplands (background map shows Sentinel-2A data obtained on August 11, 2016, R: Band 4, G: Band 3, B: Band 2).
Fig. 2 Crop classification map generated by (a) RF, (b) SVM and (c) SL.
Fig. 3 Relationship between field area and misclassified fields.

(a) RF  (b) SVM  (c) SL
Fig. 4 Mean reflectance spectra and standard deviations of each crop.
Table 1 Vegetation indices calculated from Sentinel-2 MSI data.

Table 2 Crop type and number of fields.

Table 3 Classification accuracies of each algorithm.

Table 4 Confusion matrices for (a) RF, (b) SVM and (c) SL.