Crop classification using PerceptiveSentinel

Filip Koprivec
Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
filip.koprivec@ijs.si

Matej Čerin
Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
matej.cerin@ijs.si

Klemen Kenda
Jožef Stefan Institute
Jožef Stefan International Postgraduate School
Jamova 39, 1000 Ljubljana, Slovenia
klemen.kenda@ijs.si

ABSTRACT
Efficient and accurate classification of land cover and land usage can be utilized in many different ways: ranging from natural resource management, agriculture support to legal and economic processes support. In this article, we present an implementation of land cover classification using the PerceptiveSentinel platform. Apart from using base 13 bands, only minor feature engineering was performed and different classification methods were explored. We report an F_1 and accuracy score (80-90%) in range of state of the art when using pixel-wise classification and even comparable to time series based land cover classification.

Keywords
remote sensing, earth observation, machine learning, classification

1. INTRODUCTION
Specific aspects of earth observation (EO) data (huge amount of data, widespread usage, many different problem settings etc.), coupled with the recent launch of ESA Sentinel mission that provides a huge volume of data relatively frequently (every 5-10 days), present an environment that is suitable for current machine learning approaches.

Efficient and accurate land cover classification can provide an important tool for coping with current climate change trends. Classification of crops, their location and potentially their yield prediction provide various interested parties with information on crop resistance, adapting to changes in temperature and water level changes. Along with direct help, accurate crop classification tools can be used in a variety of other programs, from government based subsidies to various insurance schemes.

Along with previously highly promising features of EO data, data acquisition and processing pose some specific challenges. Satellite acquired data is highly prone to missing data due to various reasons; mostly cloud coverage, (cloud) shadows, atmospheric refraction due to changes in atmospheric conditions. Additionally, correct training data, either for classification or regression, is hard to come by, must be relatively recent, and regularly updated due to changes in land use. Furthermore, correct labels and crop values are almost impossible to verify and usually self-reported, which often means that quality of training data is not perfect. Valero et al. [13] raise the problem of incorrect (or partially correct) data labels, which will become apparent when interpreting results.

Another class of problems is posed by the spatial resolution of images. Since satellite images provided by the ESA Sentinel-2 mission have a resolution of $10\,\mathrm{m} \times 10\,\mathrm{m}$ on most granular bands and $60\,\mathrm{m} \times 60\,\mathrm{m}$ on bands used for atmospheric correction, land cover irregularities falling in this order of magnitude might not be detected and correctly learned on. This problem is especially prevalent in smaller and more diverse regions, where individual fields are smaller and land usage is more fragmented.

The current state of the art land classification focuses heavily on the temporal dimension of acquired data [1], [13], [14]. The time-based analysis offers clear advantages since it considers growth cycles of sample crops, enables continuous classification etc., and generally produces better results, with reported F_1 scores for crop/no-crop classification in a range from 0.80-0.93 [14]. One major drawback of time-based classification is the problem of missing data. In our test drive scenario, 70% of images are heavily obscured by clouds [9], a problem which removes a lot of the advantages of time-based classification and demands major compensations with missing data imputation.

In this paper, we present a possible approach on a land cover classification of single time image acquired using the PerceptiveSentinel [1] platform, using multiple classification methods for tulip field classification in Den Helder, Netherlands.

2. PERCEPTIVESENTINEL PLATFORM
2.1 Data
Data used in this article is provided by ESA Sentinel-2 mission. The Sentinel-2 mission comprises a constellation of two polar-orbiting satellites placed in the same orbit, phased at 180° to each other [2]. Sentinel-2A was launched on 23$^{\mathrm{rd}}$ June 2015, while the second satellite was launched on 7$^{\mathrm{th}}$ March 2017. Revisit time for equator is 10 days for each satellite, so since the launch of the second satellite, each data point is sampled at least every 5 days (a bit more frequently when away from the equator).

Each satellite collects data in 13 different wavelength bands presented in figure 1, with varying granularity. Data obtained for each pixel is firstly preprocessed by ESA where

[1] http://www.perceptivesentinel.eu/

atmospheric reflectance and earth surface shadows are corrected [4].

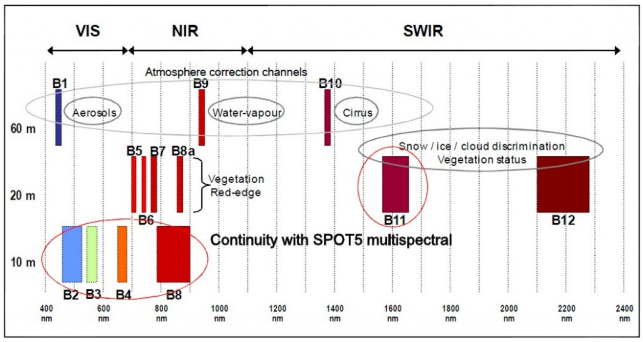

Figure 1: Sentinel 2 spectral bands [12]

2.2 Data Acquisition

Satellites provide around 1TB of raw data per day, which is provided for free on Amazon AWS. Images are then processed and indexed by Sinergise and subsequently provided for free along with their SentinelHub [11] library. As part of the PerceptiveSentinel project, a sample platform was developed on top of SH library, which eases data acquisition, cloud detection and further data analysis on acquired data.

The whole dataset currently consists of images captured from the end of June 2015 till August 2018. Images are available for use in a few hours after being taken. Since working with data for the whole world is impractical, smaller geographical regions are usually queried and analyzed on their own. One important aspect when analyzing larger regions that must be taken care of is the fact that EO data is acquired in swaths with the width of approximately 290 km [3]. Since the swaths overlap a bit, care must be taken when sampling larger areas (in a size of small state), as the area might be chopped into a few irregular tiles covering only part of an area of interest.

Corrected images are available using the SentinelHub library. PerceptiveSentinel platform provides an easy to use framework that combines satellite data acquisition, subsequent cloud detection enables an easy way to pipeline machine learning framework. They also provide an easy way to integrate (vectorized or rasterized) geopedia layers as a source of ground truth for classification.

2.3 Data Preprocessing

Most of the preprocessing is already done by ESA (atmospheric reflectance, projection . . .). The data is mostly clean and presented as a pixel array with dimensions H×W×B, where W and H are image dimensions (in our case 589 and 590) and B is number of bands selected (in our case 13, but we may individually preconfigure the Sentinelhub library to return variable number of bands and even custom calculations based on other bands).

When preprocessing we only need to consider one part of problematic data, namely clouded parts of images. ESA provides some sort of cloud detection, but our experiments proved it unsatisfactory, so we used the s2cloudless library developed by Sinergise for this task [10].

3. METHODOLOGY

3.1 Sample Data

For purpose of this article, a sample patch of fields in Den Helder, Netherlands, with coordinates: (4.7104, 52.8991), (4.7983, 52.9521) was analyzed. Three different datasets were considered: tulip fields in year 2016 and 2017 and arable land in 2017. For each of these datasets, the first observed date with no detected clouds was selected and binary classification (tulips vs no-tulips and arable vs non-arable land) was performed on the image from that date. The date selection was based on the fact that tulips' blooms are most apparent during late April and beginning of May [9].

3.2 Feature Vectors

Three additional earth observation indices that were used as features are presented in Table 1 as suggested by [8].

Name	Formula
NDVI	$\frac{B08 - B04}{B08 + B04}$
EVI	$\frac{2.5(B08 - B04)}{(B08 + 6B04 - 7.5B02 + 1)}$
SAVI	$(1 + 0.5)\frac{B08 - B04}{B08 + B04 + 0.5}$

Table 1: Additional indices

For each selected image, all 13 Sentinel2-bands were considered as feature vectors for each pixel, in the second experiment, additional land cover based classification indices from Table 1 were added.

3.3 Experiment

The experiment was conducted in the Den Helder region to asses the effectiveness of classification and improvement with additional features. The same region is also used as a test drive location for the PerceptiveSentintel project. One important characteristic to keep in mind is the fact that classification classes are not uniformly distributed. Tulip fields constitute 17.1% and 17.7% of all pixels in the year 2016 and 2017 respectively, while arable land accounts for 64% of pixels in 2017 data set. Care must, therefore, be taken when assessing the predictive power of a model.

For each dataset, multiple classification algorithms were tested on base band feature vectors and feature vectors enriched with calculated indices from Table 1. Experiments were carried out using python library `scikit-learn` and default parameters were used for each type of classifier. For each data set and each classifier (Ada Boost, Logistic regression, Random Forest, Multilayer perceptron, Gradient Boosting, Nearest neighbors and Naive Bayes), 3-fold cross-validation was performed. Folds were generated on a non-shuffled dataset with balanced classes ratios.

4. RESULTS

Results of selected classifiers are presented in Tables 2–4 (ind column indicates additional indices as features) are comparable with results from related works [5], [6] which report

accuracy results from 60-80%, although our experimental dataset was quite small and homogeneous, which might offer some advantage over larger plots of land.

Alg.	Ind	Prec	Rec	Acc	F_1	T
Logistic Regression	No	**0.895**	0.551	0.912	0.682	2.8
	Yes	0.877	0.564	0.912	0.686	3.6
Decision Tree	No	0.640	0.697	0.881	0.667	7.9
	Yes	0.629	0.698	0.878	0.662	11.3
Random Forest	No	0.870	0.675	**0.927**	0.760	15.0
	Yes	0.867	0.680	**0.927**	0.762	21.7
ML Perceptron	No	0.875	0.720	**0.935**	**0.790**	184.2
	Yes	0.835	0.740	**0.931**	**0.784**	241.3
Gradient Boosting	No	0.878	0.657	0.926	0.751	84.8
	Yes	0.856	0.657	0.923	0.743	120.6
Naive Bayes	No	0.343	**0.800**	0.704	0.480	0.4
	Yes	0.316	**0.808**	0.669	0.454	0.6

Table 2: Tulip fields in 2016 results

Alg.	Ind	Prec	Rec	Acc	F_1	T
Logistic Regression	No	0.514	0.561	0.829	0.537	2.8
	Yes	0.545	0.615	0.841	0.578	4.0
Decision Tree	No	0.574	0.633	0.852	0.602	7.3
	Yes	0.565	0.634	0.849	0.598	11.2
Random Forest	No	0.786	0.599	0.900	0.680	13.8
	Yes	**0.788**	0.604	0.901	0.683	20.5
ML Perceptron	No	**0.790**	0.673	**0.911**	0.727	375.9
	Yes	0.780	0.693	**0.911**	**0.734**	419.8
Gradient Boosting	No	0.786	0.613	0.902	0.689	84.4
	Yes	0.785	0.614	0.902	0.689	120.3
Naive Bayes	No	0.330	**0.861**	0.666	0.477	0.4
	Yes	0.318	**0.858**	0.649	0.464	0.6

Table 3: Tulip fields in 2017 results

For each test, precision, recall, accuracy, and F_1 score were reported along with the timing of the whole process. As seen from the tables, multilayer perceptron achieved best results when comparing F_1 score across all data sets, but its training was considerably slower than all other classification methods (in fact, it's training time was comparable to all other classification times combined). Multilayer perceptron was followed closely by random forest, which achieved just marginally worse results, but was way less expensive to train and evaluate, while still retaining score that was higher or comparable with related works.

Adding additional features to feature vector did not significantly improve classification score and has in some cases even hampered performance while having a significant impact on the training time.

Using classifier trained on 2016 tulips data and predicting data in 2017 yielded an F_1 score of 0.57, while classifier trained on 2017 data yielded an F_1 score of 0.67 on 2016 data, indicating the robustness of the classifier.

Graphical representation of classification errors can be seen in Figure 2 and 3 which show true positive (TP) pixels in purple color, false positive (FP) in blue color and false negative (FN) in red. Looking at the images it can easily be

Alg.	Ind	Prec	Rec	Acc	F_1	T
Logistic Regression	No	0.853	0.913	0.843	0.882	2.7
	Yes	0.854	0.908	0.841	0.880	3.2
Decision Tree	No	0.878	0.868	0.837	0.873	9.6
	Yes	0.885	0.868	0.842	0.876	14.5
Random Forest	No	0.928	0.889	0.884	0.908	17.3
	Yes	**0.934**	0.891	0.889	0.912	26.3
ML Perceptron	No	**0.929**	**0.932**	**0.911**	**0.931**	522.4
	Yes	0.926	**0.940**	**0.913**	**0.933**	586.2
Gradient Boosting	No	0.899	0.921	0.883	0.910	82.6
	Yes	0.905	0.926	0.890	0.915	118.4
Naive Bayes	No	0.823	0.830	0.776	0.827	0.4
	Yes	0.814	0.806	0.757	0.810	0.6

Table 4: Arable land in 2017 results

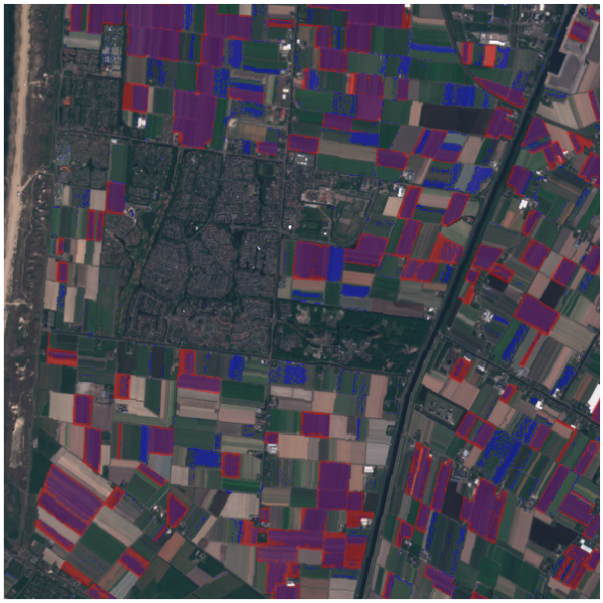

Figure 2: Graphical representation of errors in ML perceptron classification of tulip fields in 2017 (TP in purple, FP in blue, FN in red)

seen, that classification produced quite satisfactory results. An important thing to notice when inspecting Figure 2 is that the true positive pixels were classified in densely packed groups with clear and sharp edges, which correspond nicely to field boundaries seen with the naked eye (both RF and Gradient boosting decision trees produced visually very similar results). This might suggest that algorithms have detected another culture similar to tulips and classified it as tulips (or conversely, that the ground truth might not be that accurate). A lot of FN pixels can also be spotted on field boundaries, which may correspond to different quality or mixing of different plant cultures near field boundaries.

Similarly, observing results of arable land classification, one immediately notices small (false positive) blue patches and some red patches. Most notably, a long blue line is spotted in the left part of the image (near the sea), which may indicate some sort of wild culture near the sea that was not

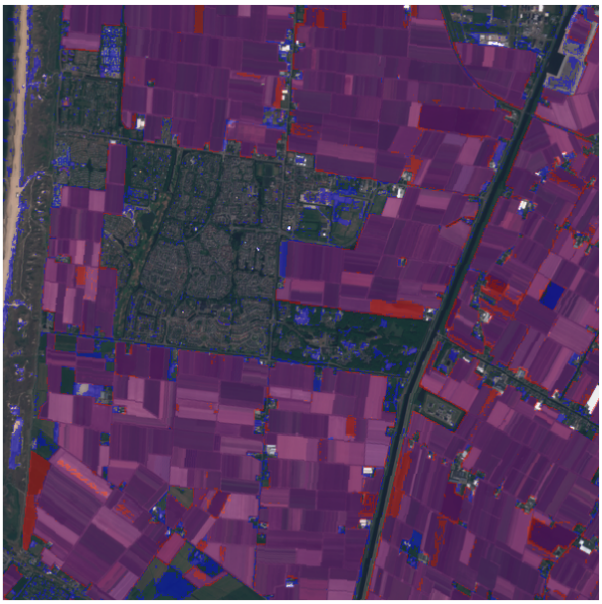

Figure 3: Graphical representation of errors in ML perceptron classification of arable land in 2017 (TP in purple, FP in blue, FN in red)

included in the original mask. Further manual observation of misclassified red patch in the middle of arable land suggests that this field is barren (easily seen in Figure 2) and possibly wrongly classified as arable land in training data.

5. CONCLUSIONS

In our work, we have examined the use of different classification methods and additional features for land cover classification problem on a single image. Our results are comparable with results from the related literature. We propose that classification strength and adaptability be further improved by considering time series and stream aggregates for each pixel as researched in [14] [7]. Additionally, pixels might be grouped together into logical objects to enable object (field) level classification as proposed by [13].

Furthermore, results have shown, that correct ground truth mask is essential for good classification performance. As seen from our results, even seemingly correct labels might miss some cultures or classify empty straits of land as crops.

6. ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency and the ICT program of the EC under project PerceptiveSentinel (H2020-EO-776115). The authors would like to thank Sinergise for it's contribution to sentinelhub and cloudless library along with all help with data analysis.

7. REFERENCES

[1] Belgiu, M., and Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment 204* (2018), 509 – 523.

[2] ESA. Satellite constellation / Sentinel-2 / Copernicus / Observing the Earth / Our Activities / ESA. `https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2/Satellite_constellation`. Accessed 13 August 2018.

[3] ESA. Sentinel-2 - Missions - Resolution and Swath - Sentinel Handbook. `https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath`. Accessed 13 August 2018.

[4] ESA. User Guides - Sentinel-2 MSI - Level-2 Processing - Sentinel Online. `https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/processing-levels/level-2`. Accessed 13 August 2018.

[5] Guida-Johnson, B., and Zuleta, G. A. Land-use land-cover change and ecosystem loss in the espinal ecoregion, argentina. *Agriculture, Ecosystems & Environment 181* (2013), 31 – 40.

[6] Gutiérrez-Vélez, V. H., and DeFries, R. Annual multi-resolution detection of land cover conversion to oil palm in the peruvian amazon. *Remote Sensing of Environment 129* (2013), 154 – 167.

[7] Gómez, C., White, J. C., and Wulder, M. A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing 116* (2016), 55 – 72.

[8] Jiang, Z., Huete, A. R., Didan, K., and Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment 112*, 10 (2008), 3833 – 3845.

[9] Kenda, K., Kažič, B., Čerin, M., Koprivec, F., Bogataj, M., and Mladenić, D. D4.1 Stream Learning Baseline Document. Reported 30. April 2018.

[10] Sinergise. sentinel-hub/sentinel2-cloud-detector: Sentinel Hub Cloud Detector for Sentinel-2 images in Python. `https://github.com/sentinel-hub/sentinel2-cloud-detector`. Accessed 14 August 2018.

[11] Sinergise. sentinel-hub/sentinelhub-py: Download and process satellite imagery in Python scripts using Sentinel Hub services. `https://github.com/sentinel-hub/sentinelhub-py`. Accessed 14 August 2018.

[12] Spaceflight 101. Sentinel-2 Spacecraft Overview. `http://spaceflight101.com/copernicus/wp-content/uploads/sites/35/2015/09/8723482_orig-1024x538.jpg`. Accessed 14 Aug. 2018.

[13] Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., and Koetz, B. Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. *Remote Sensing 8(1)* (2016), 55.

[14] Waldner, F., Canto, G. S., and Defourny, P. Automated annual cropland mapping using knowledge-based temporal features. *ISPRS Journal of Photogrammetry and Remote Sensing 110* (2015), 1 – 13.