LARGE SCALE CROP CLASSIFICATION USING GOOGLE EARTH ENGINE PLATFORM

Andrii Shelestov[1,2], Mykola Lavreniuk[1,2], Nataliia Kussul[1,2], Alexei Novikov[2], Sergii Skakun[3,4]

[1]Space Research Institute NASU-SSAU, Kyiv, Ukraine
[2]National Technical University of Ukraine “Igor Sikorsky Kiev Polytechnic Institute”, Kyiv, Ukraine
[3]Department of Geographical Sciences, University of Maryland, College Park, MD, USA
[4]NASA Goddard Space Flight Center Code 619, Greenbelt, MD, USA

ABSTRACT

For many applied problems in agricultural monitoring and food security it is important to provide reliable crop classification maps in national or global scale. Large amount of satellite data for large scale crop mapping generate a “Big Data” problem. The main idea of this paper was comparison of pixel-based approaches to crop mapping in Ukraine and exploring efficiency of the Google Earth Engine (GEE) cloud platform for solving “Big Data” problem and providing high resolution crop classification map for large territory. The study is carried out for the Joint Experiment of Crop Assessment and Monitoring (JECAM) test site in Ukraine covering the Kyiv region (North of Ukraine) in 2013. We found that Google Earth Engine (GEE) provided very good performance in enabling access to remote sensing products through the cloud platform, but our own approach based on ensemble of neural networks outperformed SVM, decision tree and random forest classifiers that are available in GEE.

***Index Terms* —** Google Earth Engine, big data, classification, optical satellite imagery, image processing.

1. INTRODUCTION

The information on land use/land cover distribution over the big areas during the long-time periods is extremely important for many environmental and monitoring tasks, including climate change, ecosystem dynamics analysis, food security and others. Reliable crop maps can be used for more accurate agriculture statistics estimation [1–3], stratification purposes [4], better crop yield prediction [5–6] and drought risk assessment [7–8]. During the last decades, satellite imagery became the most promising data source for solving such important tasks as land use/land cover mapping. Yet, there are no globally available satellite-derived crop specific maps at present. Only coarse-resolution imagery (> 250 m spatial resolution) has been utilized to derive global cropland extent (e.g. GlobCover, MODIS) [9]. Now wide range of satellites provides objective, open and free high special resolution data on a regular basis. These new opportunities allow to build high-resolution LULC maps on regular basis and to assess land use/land cover changes for large territories [10]. But with launches of Sentinel-1, Sentinel-2, Proba-V and Landsat-8 satellites, there will be generated up to petabyte of raw images per year. The increasing volume of remote sensing data, dubbed as a “Big Data” problem, creates new challenges in handling datasets that require new approaches to extracting relevant information from remote sensing (RS) data from data science perspective [11–13]. Generation of high resolution crop maps for large areas (>10,000 sq. km) using Earth observation data from space requires processing of large amount of satellite images. Images acquired at different dates during crop growth period are usually required to discriminate certain crop types. The following issues should be addressed while providing classification of multi-temporal satellite images for large areas: (i) non-uniformity of coverage of ground truth data and satellite scenes; (ii) seasonal differentiation of crop groups (e.g. winter and summer varieties) and the need for incremental classification; (iii) the need to store, manage and seamlessly process large amount of data (big data issues).

Therefore, in this paper we propose to investigate Google Earth Engine (GEE) cloud platform capabilities for large scale crop mapping using Landsat-8 optical imagery, compare different classification methods from GEE to our own neural network approach [14–18]. Results are presented for the Joint Experiment of Crop Assessment and Monitoring (JECAM) test site in Ukraine with the area of more than 28,000 km^2.

2. STUDY AREA AND MATERIALS DESCRIPTION

Ukraine is one of the most developed agricultural countries in the world. According to the U.S. Department of Agriculture (USDA) Foreign Agricultural Service (FAS) statistics, Ukraine was the largest sunflower producer (11.6 MT) and exporter, and the ninth largest wheat producer (22.2 MT) in the world in 2013. The proposed investigation is evaluated for the Joint Experiment for Crop Assessment and Monitoring (JECAM) test site in Ukraine. The JECAM test site in Ukraine was established in 2011 and covers the administrative region of Kyiv region with the geographic area of 28,100 km^2 and almost 1.0 M ha of cropland (Fig. 1). Major crops include: winter wheat, maize, soybeans, vegetables, sunflower, barley, winter rapeseed, and sugar beet. The crop calendar is September-July for winter crops, and April-October for spring and summer crops. Fields in Ukraine are quite large with size generally up to 250 ha.

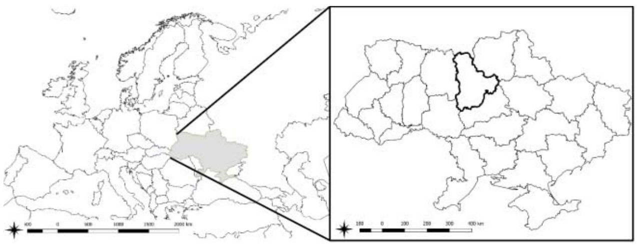

Fig. 1. Location of Ukraine and JECAM test site in Ukraine (Kyiv region, marked with bold boundaries)

Ground surveys to collect data on crop types and other land cover classes were conducted in 2013 in Kyiv region. European Land Use and Cover Area frame Survey (LUCAS) nomenclature was used in this study as a basis for land cover / land use types. In total, 386 polygons were collected covering the area of 22,700 ha. Data were collected along the roads using mobile devices with built-in GPS. All surveyed fields are randomly divided into training set (50%) to train the classifier and testing set (50%) for testing purposes (Fig. 2). Fields are selected in such a way so there is no overlap between training and testing sets. All classification results, in particular overall accuracy (OA), user's (UA) and producer's (PA) accuracies are reported for testing set. The input features are classified into one of the 13 classes.

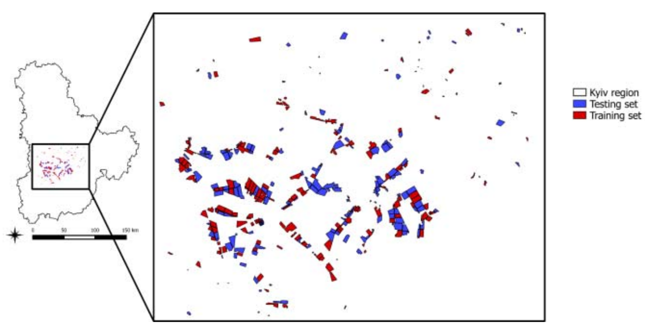

Fig. 2. Location of fields surveyed during ground measurements in 2013.

Remote sensing images acquired by the Operational Land Imager (OLI) sensor aboard Landsat-8 satellite were used for crop mapping over the study region. Only bands 2 through 7 were used for crop classification maps providing [10]. Three scenes with path/row coordinates 181/24, 181/25 and 181/26 covered the test site region.

2.1 Composite products available at GEE

Different composites derived from Landsat-8 imagery and available in GEE were analyzed in the study. Landsat 8 8-Day Top-of-atmosphere (TOA) Reflectance Composites were used from GEE. As to the time of composition, 8 day composites were selected over 32 day composites. The reason for that is that 32 composites are composed based on the latest image, and this latest image can be of not the best quality. All the images from each 8-day period are included in the composite, with the most recent pixel on top.

2.2 Landsat-8 data pre-processing (outside GEE)

The following pre-processing steps are applied for all Landsat-8 images: conversion of digital numbers (DNs) values to the top-of-atmosphere (TOA) reflectance values, conversion from the TOA reflectance to the surface reflectance (SR) using the Simplified Model for Atmospheric Correction (SMAC), detection of clouds and shadows were done using Fmask algorithm, reconstruction from missing pixel values (clouds and shadows) using self-organizing Kohonen maps (SOMs) [14, 19].

These pre-processing steps were performed outside GEE platform. After these products were generated they were uploaded in the GEE cloud platform for the further classification using classification algorithm available in GEE.

3. METHODOLOGY

In the study classification was done on a per-pixel basis. Multiple classification techniques were evaluated in the study. At first, all classification algorithms available in GEE were analyzed and used for classifying multi-temporal 8-days Landsat-8 TOA composites from GEE. Then the best classification algorithms in terms of overall classification accuracy were compared with the neural network classifier that used multi-temporal SR values generated outside GEE. The presented approaches were compared in terms of classification accuracy at pixel level. GEE offers several classification algorithms among which are decision trees and random forests, support vector machine (SVM), and Naïve Bayes classifier. It should be noted that neural network (NN) classifiers are not available in GEE.

3.1 Ensemble of Neural Networks

Our proposed neural network approach based on committee of NNs, in particular Multi-Layer Perceptron (MLPs), is utilized to improve performance of individual classifiers. The MLP classifier has a hyperbolic tangent activation function for neurons in the hidden layer and logistic activation function in the output layer. A committee of MLPs was used to increase performance of individual

classifiers. The committee is formed using four MLPs with different number of hidden neurons (10, 20, 30, and 40) trained on the same training data within 250 epochs. Outputs from different MLPs were integrated using the technique of average committee. Under this technique, the average class probability over classifiers is calculated, and the class with the highest average posterior probability for the given input sample is selected. In order to prevent NN overfitting, we exploited early stopping and weight decay (L2 regularization) techniques. Coefficient of regularization was selected from 0.00005, 0.0001, 0.0003, 0.001, 0.003, 0.01, and 0.03 using a validation set (20% of training set).
This approach is used as a benchmark for assessing classification techniques available in GEE.

4. RESULTS

4.1 Input (product) selection

The first set of experiments was carried out to select the best input (TOA 8-day composites or restored values) and evaluating different classifiers available in GEE. Table 1 shows the derived OA on polygons from a testing set using TOA 8-day composites as inputs. The best performance was achieved for CART at 75%. Somewhat surprisingly, an ensemble of DTs, i.e. RF, was outperformed by CART and yielded only 68%. Logistic regression (GMO Max Entropy) gave 72% accuracy. Linear classifiers, MultiClassPerceptron and Winnow, provided up to 60% accuracy, while variants of SVM achieve moderate accuracy of 57%. Unfortunately, it was unable to produce stable classification results for SVM classifiers which usually resulted into the Internal Server Error on invocation from Python.

Table 1. Overall classification accuracy achieved by GEE classifiers for TOA 8-day composites as an input

Classifier	OA, %
CART	75
GMO Max Entropy	72
Random Forest	68
MultiClassPerceptron	60
IKPamir	57
Winnow	49
FastNaiveBayes	32
Pegasos	-
VotingSvm	-
MarginSvm	-

4.2 Classifier selection

One of the best GEE classifiers (CART and RF) on atmospherically corrected Landsat-8 imagery were compared to the committee of NN that was implemented outside GEE, in the Matlab environment using a Netlab toolbox. CART provided overall accuracy 76.9%, RF achieved 69.9% and committee of MLPs considerably outperformed DT-based classifiers: by +14.8% RF and by +7.8% DT (Fig. 3).

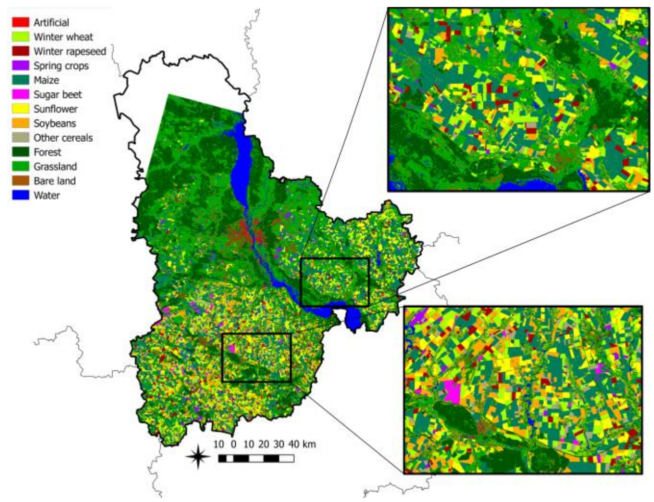

Figure 3. Final map obtained by classifying multi-temporal Landsat-8 imagery using a committee of MLP classifiers.

5. DISCUSSION AND CONCLUSIONS

The GEE platform offers powerful capabilities in handling large volumes of remote sensing imagery that can be used, for example, for classification purposes and crop mapping for large territories. The best OA achieved on composites from the GEE was 75%, while on atmospherically corrected and restored images the achieved accuracy was almost 77%. The GEE platform provides a set of classification algorithms. The best results in the GEE were obtained for the DT-based classifiers, namely CART and RF. At the same time, committee of neural networks considerably outperformed DT-based classifiers: by +14.8% RF and by +7.8% DT. Research within this paper were targeted on the comparison of pixel-based approaches to crop mapping in Ukraine and exploring efficiency of the Google Earth Engine (GEE) cloud platform for large scale crop mapping. In general, GEE provided very good performance in enabling access to remote sensing products through the cloud platform and powerful computational resources that could help users to deal with “Big Data” problem in crop mapping for large territory.

6. ACKNOWLEDGEMENTS

This research was conducted in the framework of the “Large scale crop mapping in Ukraine using SAR and optical data fusion” Google Earth Engine Research Award funded by the Google Inc.

7. REFERENCES

[1] J. Gallego, E. Carfagna, and B. Baruth, "Accuracy, objectivity and efficiency of remote sensing for agricultural statistics," *Agricultural survey methods*, pp. 193-211, 2010.

[2] J. Gallego, A. N. Kravchenko, N. N. Kussul, S. V. Skakun, A. Y. Shelestov, and Y. A. Grypych, "Efficiency assessment of different approaches to crop classification based on satellite and ground observations," *Journal of Automation and Information Sciences*, vol. 44, no. 5, 2012.

[3] F. J. Gallego, N. Kussul, S. Skakun, O. Kravchenko, A. Shelestov, and O. Kussul, "Efficiency assessment of using satellite data for crop area estimation in Ukraine," *International Journal of Applied Earth Observation and Geoinformation*, vol. 29, pp. 22-30, 2014.

[4] C. G. Boryan and Z. Yang, "Deriving crop specific covariate data sets from multi-year NASS geospatial cropland data layers," *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4225-4228, 2013.

[5] F. Kogan et al., "Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models," *International Journal of Applied Earth Observation and Geoinformation*, vol. 23, pp. 192-203, 2013.

[6] A. Kolotii, N. Kussul, A. Shelestov, S. Skakun, B. Yailymov, R. Basarab, M. Lavreniuk, T. Oliinyk, and V. Ostapenko, "Comparison of biophysical and satellite predictors for wheat yield forecasting in Ukraine," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XL-7/W3, pp. 39-44, 2015. DOI: 10.5194/isprsarchives-XL-7-W3-39-2015.

[7] S. Skakun, N. Kussul, A. Shelestov, and O. Kussul, "The use of satellite data for agriculture drought risk quantification in Ukraine," *Geomatics, Natural Hazards and Risk*, vol. 7, no. 3, pp. 901-917, 2016.

[8] N. N. Kussul, B. V. Sokolov, Y. I. Zyelyk, V. A. Zelentsov, S. V. Skakun, and A. Y. Shelestov, "Disaster risk assessment based on heterogeneous geospatial information," *Journal of Automation and Information Sciences*, vol. 42, no. 12, 2010.

[9] S. Fritz et al., "The need for improved maps of global cropland," *Eos, Transactions American Geophysical Union*, vol. 94, no. 3, 31-32, 2010.

[10] D. P. Roy, M. A. Wulder, T. R. Loveland, C. E. Woodcock, R. G. Allen, M. C. Anderson, ... and T. A. Scambos, "Landsat-8: Science and product vision for terrestrial global change research," *Remote Sensing of Environment*, vol. 145, pp. 154-172, 2014.

[11] Y. Ma, et al. "Remote sensing big data computing: challenges and opportunities," *Future Generation Computer Systems*, vol. 51, pp. 47-60, 2015.

[12] Y. Ma, et al. "Towards building a data-intensive index for big data computing–A case study of Remote Sensing data processing," *Information Sciences*, vol. 319, pp. 171-188, 2015.

[13] N. Kussul, A. Shelestov, R. Basarab, S. Skakun, O. Kussul, and M. Lavrenyuk, "Geospatial Intelligence and Data Fusion Techniques for Sustainable Development Problems," *In ICTERI*, pp. 196-203, 2015.

[14] S. Skakun, N. Kussul, A. Y. Shelestov, M. Lavreniuk, and O. Kussul, "Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine," *IEEE J. of Select. Topics in Applied Earth Obser. and Rem. Sens.*, vol. 9, no. 8, pp. 3712-3719, 2016.

[15] F. Waldner, et al., "Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity," *International Journal of Remote Sensing*, vol. 37, no. 14, pp. 3196-3231, 2016.

[16] S. V. Skakun, E. V. Nasuro, A. N. Lavrenyuk, and O. M. Kussul, "Analysis of applicability of neural networks for classification of satellite data," *Journal of Automation and Information Sciences*, vol. 39, no. 3, 2007.

[17] M. S. Lavreniuk, S. V. Skakun, A. J. Shelestov, B. Y. Yalimov, S. L. Yanchevskii, D. J. Yaschuk, and A. I. Kosteckiy, "Large-Scale Classification of Land Cover Using Retrospective Satellite Data," *Cybernetics and Systems Analysis*, vol. 52, no. 1, pp. 127–138, 2016.

[18] N. Kussul, G. Lemoine, F. J. Gallego, S. V. Skakun, M. Lavreniuk, and A. Y. Shelestov, "Parcel-Based Crop Classification in Ukraine Using Landsat-8 Data and Sentinel-1A Data," *IEEE J. of Select. Topics in Appl. Earth Observ. and Rem. Sens.*, vol. 9, no. 6, pp. 2500–2508, 2016.

[19] N. Kussul, S. Skakun, A. Shelestov, M. Lavreniuk, B. Yailymov, and O. Kussul, "Regional Scale Crop Mapping Using Multi-Temporal Satellite Imagery," Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-7/W3, pp. 45–52, 2015. DOI:10.5194/isprsarchives-XL-7-W3-45-2015.