A DEEP LEARNING APPROACH FOR URBAN LAND COVER CLASSIFICATION FROM HIGH-SPATIAL RESOLUTION IMAGERY AND GEOMORPHOMETRIC VARIABLES

I. Lizarazo[a,*], S. Ramirez[b]

[a] Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogota, D.C. - ializarazos@unal.edu.co
[b] Facultad de Ingenieria, Universidad Distrital Francisco Jose de Caldas, Bogota, D.C. - sramirez@udistrital.edu.co

KEY WORDS: Urban Land Cover, Deep Learning, Geomorphometry, Digital Surface Models, Extreme Gradient Boosting, Random Forests

ABSTRACT:

This paper describes a deep learning approach for urban land cover classification in the context of the ISPRS 2D semantic labelling benchmark. A high spatial resolution digital surface model (DSM) and a true ortho-image over the city Potsdam (Germany) was used as input dataset for obtaining six target classes. The proposed approach focuses on augmenting the original input dataset with a combined set of geo-morphometric variables extracted from DSM –including slope/aspect transformation, second derivate of elevation, compound topographic index and hierarchical slope position–. Furthermore, it uses advanced deep learning architecture provided by H2O framework which follows the model of multi-layer, feedforward neural networks for predictive modelling. Automatic hyper-parameter tuning with random search was conducted for model selection. The method comprises five steps: (i) spectral segmentation of ortho-images; (ii) extraction of relevant geo-morphometric variables from DSM; (iii) multivariate land cover classification; and (iv) accuracy assessment. The proposed approach was used for classifying a selected ISPRS benchmark tile where a reference map is available. Thematic accuracy of the proposed approach was assessed using the traditional error matrix and compared with thematic accuracy of a deep learning classification based only on the original data set (i.e. DSM and multispectral imagery). In addition, the deep learning classification approach was compared with a random forest (RF) classification using both original and augmented input dataset. It is shown that: (i) thematic accuracy improves only slightly when geomorphological variables are used to enhance the input dataset; and (ii) deep neural nets provide a similar predictive power than random forests for urban remote sensing applications.

1 INTRODUCTION

1.1 Machine Learning

Machine learning (ML) is an effective empirical approach for supervised or unsupervised land cover classification (supervised or unsupervised) . In ML, a comprehensive 'training dataset' of examples is constructed covering as much of the system parameter space as possible. Typically, a random subset of the data is put aside for a completely independent validation. ML is ideal for addressing those problems where our theoretical knowledge is still incomplete but for which we do have a significant number of observations and other data (Lary et al., 2016).

ML algorithms commonly used include artificial neural networks (ANN), support vector machines (SVM), self-organizing map (SOM), decision trees (DT), ensemble methods such as random forests (RF), case-based reasoning, neuro-fuzzy (NF), genetic algorithm (GA), multivariate adaptive regression splines (MARS). Application of these techniques in geosciences and remote sensing is relatively new but expanding (e.g. Atkinson and Tatnall, 1997; Brown et al., 2008; Lizarazo, 2008; Atzberger and Rembold, 2013; Hu et al., 2015; Langkvist et al., 2015; Liang and Li, 2016).

1.2 Deep Learning

Deep Learning (DL) techniques use neural networks with many hidden layers. Two options for applying DL are Deep Neural Networks (DNN) and Deep Belief Networks (DBN). A DBN has undirected connections between the top two layers (Hinton et al., 2012). This means that the topology of the DNN and DBN is different by definition. The undirected layers in a DBN are called Restricted Boltzmann Machines. This layers can be trained using an unsupervised learning algorithm (i.e., contrastive divergence) that is very fast.

In contrast to DBNs, a DNN is a feed-forward, artificial neural network that has more than one layer of hidden units between its inputs and its outputs – and there are no undirected connections (Hinton et al., 2012). Each hidden unit, j, typically uses the logistic function (the closely related hyberbolic tangent is also often used and any function with a well-behaved derivative can be used) to map its total input from the layer below, x_j, to the scalar state, y_j that it sends to the layer above (Hinton et al., 2012).

$$y_j = logistic(x_j) = \frac{1}{1+e^{-x_j}}; x_j = b_j + \sum_{i=1} y_i w_{ij}, \quad (1)$$

where b_j is bias of unit j, i is index over units in the layer below, and w_{ij} is weight on a connection to unit j from unit i in the layer below.

For multiclass classification, output unit j converts its total input, x_j, into a class probability, p_j, by using the 'softmax' nonlinearity

$$p_j = \frac{exp(x_j)}{\sum exp(x_k)}, \quad (2)$$

where k is an index over all classes (Hinton et al., 2012).

A DNN can be discriminatively trained (DT) by back-propagating derivatives of a cost function that measures the discrepancy between the target outputs and the actual outputs produced for each training case. When using the soft-max output function, the nat-

*Corresponding author

ural cost function C is the cross entropy between the target probabilities d and the outputs of the softmax, p:

$$C = -\sum_{j=1} d_j log p_j, \tag{3}$$

where the target probabilities, typically taking values of one or zero, are the supervised information provided to train the DNN classifier (Hinton et al., 2012).

For large training sets, it is typically more efficient to compute the derivatives on a small, random 'minibatch' of training cases, rather than the whole training set, before updating the weights in proportion to the gradient. This stochastic gradient descent method can be further improved by using a 'momentum' coefficient, $0 < \alpha < 1$, that smooths the gradient computed for minibatch t, thereby damping oscillations across ravines and speeding progress down ravines:

$$\Delta w_{ij}(t) = -\alpha \Delta w_{ij}(t-1) - \epsilon \frac{\partial C}{w_{ij}(t)}. \tag{4}$$

The update rule for biases can be derived by treating them as weights on connections coming from units that always have a state of one (Hinton et al., 2012).

1.3 Tree based Learning

Tree based learning algorithms are considered to be one of the best and mostly used supervised learning methods. Tree based methods empower predictive models with high accuracy, stability and ease of interpretation. Unlike linear models, they map non-linear relationships quite well. They are adaptable at solving any kind of problem at hand (classification or regression) (Breiman, 2001).

Methods like decision trees, random forest, gradient boosting are being popularly used in remote sensing problems (Belgiu and Dragut, 2016). Main advantages of such methods can be summarized as follows (Breiman, 2001): (i) Decision Trees (DTs) captures non-linear relationship and are robust to correlated features, feature distribution, and missing values. However, its performance is usually not very top-tiered. So there are two different categories of solutions for this problem, one is through bagging (RandomForest) and the second is through boosting (Gradient Boosting Machine); Random Forests (RF) are composed of many trees. Each tree is built on a sample of the features and on a sample of the observations (to increase variance of trees). Trees are independent with one another. It is easy to use since it has very few hyper-parameters to tune. And it runs pretty well with the default parameters. A main drawback is that it may be slow when it comes to scoring; and (iii) Gradient Boosting Machine (GBM) are also based on many trees. A GBM iteratively learns weak classifiers and adds them to a final strong classifier. After a weak learner is added, the data is re-weighted: examples that are misclassified gain weight and examples that are classified correctly lose weight. Thus, future weak learners focus more on the examples that previous weak learners misclassified. Trees are dependent with one another causing it to be not very robust to noisy data and outliers.

Results from several land cover classification studies (e.g. Pal, 2005; and Lizarazo, 2010) suggest that the RF classifier performs equally well to SVMs and other ML techniques in terms of classification accuracy and training time. Such studies also conclude that the number of user defined parameters required by RF classifiers is less than the number required for SVMs and easier to define. On the another hand, only a few number of studies have applied either DNN or DBN techniques for land cover classification (an exception being Lv et al, 2015 and Langkvist et al, 2016). Most studies focuses on 'scene classification', that is, giving each image an unique label according to a set of semantic categories (e..g. Castelluccio et al., 2015). Experimental results show that DNN and DBN based methods outperform other classification approaches and produce homogenous mapping results with preserved shape details.

1.4 Objectives

This article aims to establish whether urban land cover classification obtained from the ISPRS 2D semantic labelling benchmark (i.e a combination of a true color & infrared orthophoto (TO) plus a digital surface (DSM)) using a DNN approach is thematically more accurate than the one obtained using a RF technique. In addition, potential contribution of terrain-related variables, derived from DSM, to improve thematic accuracy of land cover classification is evaluated.

2 DATA AND METHODS

2.1 Data

2.1.1 Original data set Performance of the proposed deep learning classification approach was evaluated using the ISPRS 2D semantic labeling benchmark. A high spatial resolution digital surface models (DSM) and a true ortho-image over the city Potsdam (Germany), namely the $6-11$ tile, were used as input dataset for obtaining six target classes. The ortho-image comprises 3 spectral bands in the visible range and 1 in the near-infrared. Each band is 6000 x 6000 pixels with a spatial resolution of 0.05 m. The DSM has similar spatial extent and pixel size. Figure 1 shows: (a) a true colour composition from orthophoto; (b) a pseudo color representation of DSM.

(a) (b)

Figure 1. Input data set for this classification study: (a) Orthophoto's true colour composition; (b) DSM pseudo-color representation.

A Postdam reference map is available for the $6-11$ tile. It includes six different thematic classes:

1. Tree (green)
2. Low vegetation (cyan)
3. Clutter/background (red)
4. Building (blue)
5. Impervious surfaces (white)
6. Car (yellow)

2.1.2 Training and validation data sets Two stratified random samples of approx. equal-sized number of points at each thematic class were selected from the reference map. For the training step, about 10000 sampling points were drawn (see Table 1). For validation of classification results, about 30000 samples were selected (see Table 1). At each sampling point a 3x3 window was considered for both establishing the response value (i.e. majority thematic class) and input data set values (i.e. mean of spectral and elevation bands).

Class	Training	Testing
1	1763	3778
2	1612	3639
3	1654	3659
4	1913	3912
5	1791	3798
6	1272	3302

Table 1. Number of training and testing points at each thematic category.

2.1.3 Augmented data set The original input data set was augmented using a number of geo-morphometric variables extracted from DSM. Such variables include slope/aspect transformation (SAT), second derivate of elevation (CURV), compound topographic index (CTI), heat load index (HLI), and hierarchical slope position (HSI). Figure 2 depicts these terrain variables in a small window of $6 - 11$ tile.

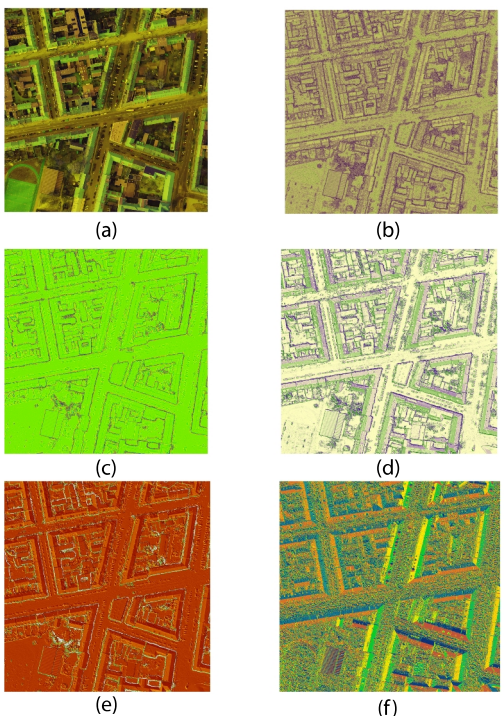

Figure 2. Geo-morphometric variables obtained from DSM: (a) RGB123 ortophoto colour composition ; (b) slope/aspect transformation (SAT), (c) second derivate of elevation (CURV), (d) compound topographic index (CTI), (e) heat load index (HLI), and (f) hierarchical slope position (HSI).

Slope/aspect transfomation was calculated using a simple ratio (Wilson, 2012):

$$SAT = \frac{slope}{aspect}, \tag{5}$$

Curvature is second derivative of elevation and was calculated according to Shary (1995):

$$CURV = \frac{d^2(dz)}{dh^2} \tag{6}$$

CTI is a steady state wetness index. CTI is a function of both the local slope ($tan(B)$ in radians) and the upstream contributing area per unit width orthogonal to the flow direction (As in m_2) (Hjerdt et al., 2004):

$$CTI = ln\frac{A_s}{tan(B)}, \tag{7}$$

HLI is an aspect and slope based terrain metric. McCune and Keon (2002) 'fold' aspect and steepness so that 'heat' highest values are southwest and lowest values are northeast

$$HLI = \frac{aspect - slope}{aspect + slope}, \tag{8}$$

HSI is an index which captures terrain attributes in a hierarchical way by averaging slope over a number of windows at different scales (Florinsky, 2012):

$$HSI = \frac{\sum_{j=1} d_j slope_j}{n} \tag{9}$$

2.2 Methods

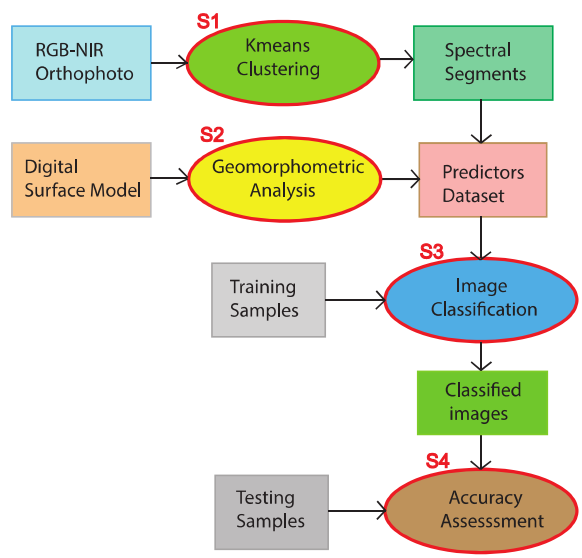

Figure 3. Flowchart of land cover mapping stages in this study.

Figure 3 depicts the four-stage method followed in this study: (S1) Image Segmentation; (S2) Terrain analysis; (S3) Supervised image classification; and (S4) Accuracy assessment. In the S1 stage, the *Kmeans* algorithm was applied to produce spectral segmentation into 48 clusters. In the S2 stage, five (5) geo-morphometric variables (i.e. SAT, CURV2, CTI, HLI and HSI) were obtained from the DSM. In the S3 stage, DNN and RF classification models were built and applied using as predictors features extracted from either the original input data set or the augmented data set. Original features include mean of Red (R), Green (G), Blue (B)

and Infra-Red (IR) bands within spectral segments as well as pixel-based DSM values. Augmented features include original features plus pixel-based geo-morphometric values. After class prediction, a 13 x 13 majority filter was applied to smooth classified images. In the S4 stage, accuracy assessment of obtained thematic images was conducted using the traditional error matrix.

The *R* system (R Core Team, 2015), a free software environment for statistical computing and graphics, was used to implement every stage of the proposed method. Raster and vector data sets reading, processing and analysis tasks were conducted using *rgdal*, *sp*, *raster* and *rasterVis* libraries. DNN and RF model building and prediction was accomplished using *h2o* library. Tuning of DNN hyper-parameters was accomplished using a random search technique. Optimal values obtained include a deep neural net with five (5) hidden layers using 500, 1000, 2000, 1000 and 500 nodes per respective layer. RF application was conducted with no tuning technique, that is, using 500 trees, the default setting.

3 RESULTS

Figure 4 shows classification results for different methods corresponding to a small window near to the lower left zone of $6-11$ tile. This zoom in allows for a better visualization of results. However, interested readers may access the complete classified images at *https://goo.gl/KogDX0*. As reported in Tables 2, 3, 4, 5 the DNN approach using the augmented input data set obtained better results. Accuracy metrics show that most elusive categories to map are *Tree* (class 1) and *Low vegetation* (class 2) and that most reachable ones are *Clutter / background* (class 3) and *Building* (class 4).

A close look at Figure 4 and Tables 6, 7, 8, and 9 shows that DNN and RF techniques obtained similar results from both the original input dataset and the augmented input dataset. Results also show that inclusion of geo-morphometric variables in the input data set improves thematic accuracy only slightly.

Accuracy assessment was based on the computation of pixel-based error matrices per classification using the testing sample data set described in Section 2.1.2 and the reference data set with no boundary (i.e. testing points eventually falling on a boundary were ignored). From those matrices different measures are derived: per class we compute completeness (recall), correctness (precision) and $F1$ score, and through the normalisation of the trace from error matrices overall accuracy (OA) values were derived.

Class	1	2	3	4	5	6
1	**3043**	273	20	18	117	55
2	475	**2712**	7	25	96	18
3	138	70	**2803**	54	177	90
4	40	39	12	**3721**	32	15
5	136	186	33	85	**3188**	133
6	18	5	44	0	57	**2435**

Table 2. Error matrix for DNN classification from original data set. Overall accuracy (OA) is 0.878.

In confusion matrices row direction gives reference values, while column direction shows the prediction. In precision tables cells show normalised values with respect to reference values. The True Positive t_p pixels are derived from the main diagonal elements, while the False Positive f_p is computed from the sum per column, excluding the main diagonal element. Likewise, the False Negative f_n is the sum along the row, excluding the main diagonal element. From those values following measures are derived:

$$Precision = \frac{t_p}{t_p + f_p} \tag{10}$$

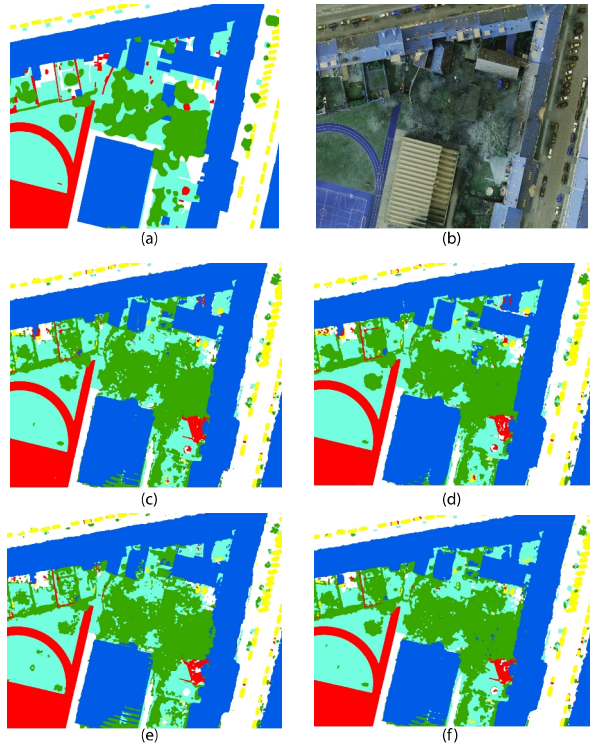

Figure 4. Classification results for a spatial subset of 611 tile of Potsdam data set: (a) reference; (b) orthophoto; (c) result for DNN on original data set; (d) result for DNN on augmented data set; (e) result for RF on original data set; (f) result for RF on augmented data set.

Class	1	2	3	4	5	6
1	**3038**	274	15	29	119	50
2	448	**2735**	9	30	92	19
3	119	74	**2782**	50	208	96
4	34	29	10	**3726**	42	18
5	121	192	31	64	**3224**	130
6	16	7	50	1	57	**2429**

Table 3. Error matrix for DNN classification from augmented data set. Overall accuracy (OA) is 0.880.

Class	1	2	3	4	5	6
1	**2972**	317	19	42	125	53
2	635	**2563**	1	46	74	13
3	123	78	**2814**	68	164	86
4	67	41	10	**3707**	25	8
5	173	149	25	143	**3175**	98
6	11	13	28	2	76	**2428**

Table 4. Error matrix for RF classification from original data set. Overall accuracy (OA) is 0.866.

Class	1	2	3	4	5	6
1	**2993**	262	13	34	165	54
2	539	**2662**	3	26	89	12
3	155	82	**2786**	54	181	74
4	43	40	9	**3727**	30	10
5	162	199	28	98	**3164**	115
6	19	10	41	0	69	**2418**

Table 5. Error matrix for RF classification from augmented data set. Overall accuracy (OA) is 0.872.

$$Recall = \frac{t_p}{t_p + f_n} \tag{11}$$

The F1-score is defined as the harmonic mean of precision and recall:

$$F_1 = \frac{Precision.Recall}{Precision + Recall} \tag{12}$$

Class	Precision	Recall	F1
1	0.790	0.863	0.825
2	0.825	0.814	0.819
3	0.960	0.841	0.897
4	0.953	0.964	0.959
5	0.869	0.848	0.858
6	0.887	0.951	0.918

Table 6. Precision, Recall and F1 statistics for DNN classification from original data set (DNN_{or}).

Class	Precision	Recall	F1
1	0.804	0.862	0.832
2	0.826	0.820	0.823
3	0.960	0.836	0.894
4	0.955	0.965	0.960
5	0.861	0.857	0.859
6	0.886	0.949	0.916

Table 7. Precision, Recall and F1 statistics for DNN classification from augmented data set (DNN_{au}).

Class	Precision	Recall	F1
1	0.746	0.842	0.791
2	0.811	0.769	0.789
3	0.971	0.844	0.903
4	0.924	0.961	0.942
5	0.872	0.844	0.858
6	0.904	0.949	0.926

Table 8. Precision, Recall and F1 statistics for RF classification from original data set (RF_{or}).

Class	Precision	Recall	F1
1	0.765	0.850	0.805
2	0.818	0.799	0.808
3	0.967	0.836	0.897
4	0.946	0.966	0.956
5	0.855	0.840	0.848
6	0.901	0.946	0.923

Table 9. Precision, Recall and F1 statistics for RF classification from augmented data set (RF_{au}).

4 DISCUSSION

It is worth to have a closer look at classifications obtained in this study in order to understand why *Tree* (green) and *Low vegetation* (cyan) categories have the lowest accuracies. Figure 5 shows a small window near top right corner of $6 - 11$ tile. It shows: (a) reference map; (b) DNN classification from augmented input data set; (c) ortho-photo; (d) DSM. From the ortho-image, it can be seen that trees have dropped most of their leaves and confound easily with grass and low vegetation. In fact, by looking through tree branches it is possible to identify ground objects. Figure 5 also shows that cars (yellow) are well mapped and that small buildings (red) are poorly mapped.

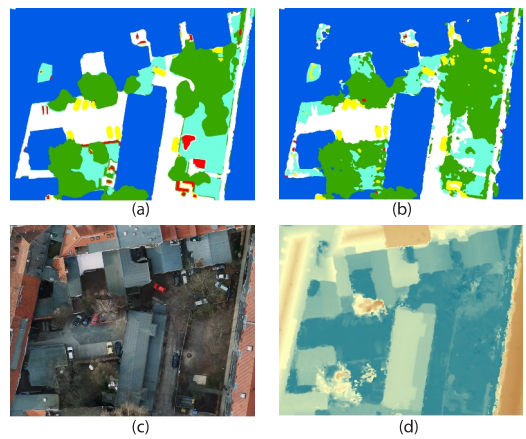

Figure 5. A closer look at the DNN land cover classification from the augmented data set: (a) reference map; (b) DNN classification from augmented input data set; (c) orthophoto; (d) DSM

Method	OA	mPrecision	mRecall	mF1
DNN_{or}	0.878	0.881	0.880	0.8794
DNN_{au}	0.880	0.882	0.881	0.881
RF_{or}	0.866	0.872	0.868	0.868
RF_{au}	0.872	0.875	0.873	0.873

Table 10. Overall Accuracy (OA), Precision and Recall macro statistics for all classified images

Table 10 summarizes thematic accuracy of results. From matrices presented in Section 3, it may be stated that, for practical applications, any of the DNN and RF techniques evaluated here allows for accurate classifications. However, a relevant discussion topic is to examine how expensive is each method regarding its use of computational resources and human effort. On such a matter, it should be said that this study confirms that DNN needs careful tuning of hyper-parameters as suggested from previous work (e.g. Hinton et al., 2012). Moreover, DNN seems to be a very unstable technique as minor variations in, for example, number of hidden layers or number of nodes, causes unexpected changes in model's behaviour and performance. While this happens to DNN, RF needs no significant work on parameterisation and works in a very stable way.

On the another hand, inclusion of geo-morphometric variables in the input data set did not improve accuracy of results significantly. This is confirmed by examination of variable importance ranking reported in Table 11. It should noted that elevation from DSM and Mean of IR band are the most important variables in both DNN and RF models. Geo-morphometric variables are the less important ones while their ranking is different in each

Variable	DNN_{au}	RF_{au}
DSM	1	1
Mean of IR band	2	2
Mean of G band	3	4
Mean of R band	4	3
CURV	5	8
HSI	6	5
Mean of B band	7	6
CTI	8	10
HLI	9	7
SAT	10	9

Table 11. Overall Accuracy (OA), Precision and Recall macro statistics for all classified images

method. It should also noted that DSM is a pixel-wise variable –and this holds for every geo-morphometric variable in this study.

5 CONCLUSIONS

In this article, a new classification method that uses deep learning techniques for land cover classification from ortho-images and digital surface models (DSM) was investigated. The classification results indicate that deep neural networks (DNN) and random forest (RF) techniques can accurately classify target thematic categories. Furthermore, it is shown that neither DN nor RF outperforms each other method results in terms of thematic accuracy. It was also investigated the influence of geo-morphometric variables, obtained from DSM, on classification. It is shown that such variables increase thematic accuracy only slightly. Machine learning techniques examined in this study give very promising results, particularly the RF algorithm which is very robust, consistent and efficient. Future work will be focused on investigating causes of unstable performance of the deep learning implementation used in this study.

REFERENCES

Atkinson P. M. and Tatnall, A. R. L., 1997. Introduction: neural networks in remote sensing. *International Journal of Remote Sensing*, 18 (4), pp. 699-709.

Atzberger, C. and Rembold, F., 2013. Mapping the Spatial Distribution of Winter Crops at Sub-Pixel Level Using AVHRR NDVI Time Series and Neural Nets. *Remote Sensing*, 5(3), pp. 1335-1354.

Belgiu, M. and Dragut, L. 2016. Random forest in remote sensing: A review of applications and future directions *ISPRS Journal of Photogrammetry and Remote Sensing*, 114 (4), pp.24-31

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), pp. 5-32.

Brown, M. E., Lary, D. J., Vrieling, A. , Stathakis, D. and Mussa H., 2008. Neural networks as a tool for constructing continuous NDVI time series from AVHRR and MODIS. *International Journal of Remote Sensing*, 29 (24), pp. 7141-7158.

Castelluccio, M, Poggi, G., Sansone C. and Verdoliva, L., 2015. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. *arXiv preprint*, arXiv:1508.00092.

Florinsky I. 2012. Digital Terrain Analysis in Soil Science and Geology. Elsevier Academic Press: Oxford, UK.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N. and Kingsbury, B., 2012. Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Processing Magazine*, 29(6), pp. 82-97.

Hjerdt, K. N.,McDonnell, J. J.,Seibert, J. and Rodhe, A., 2004. A new topographic index to quantify downslope controls on local drainage. *Water Resources Research*, 40(5), CiteID W05602.

Hu, F., Xia, G-S., Hu, J., and Zhang, L., 2015. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sensing*, 7(11), pp. 14680-14707.

Langkvist, M., Kiselev, A., Alirezaie, M. and Loutfi, A., 2016. Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks. *Remote Sensing*, 8(4), pp. 329-350.

Lary, D. J., Alavi, A. H., Gandomi, A. H., Walker, A. L., 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*,7(1), pp. 3-10.

Liang, H. and Li, Q., 2016. Hyperspectral Imagery Classification Using Sparse Representations of Convolutional Neural Network Features. *Remote Sensing*, 8(2), pp. 99-118.

Lizarazo, I., 2008. SVM based segmentation and classification of remotely sensed data. *International Journal of Remote Sensing*, 29 (24), pp. 7277-7283.

Lizarazo, I., 2010. Fuzzy image regions for estimation of impervious surface areas. *Remote Sensing Letters*, 1(1), pp. 19-27.

Lv, Q., Dou, Y., Niu, X., Xu, J., Xu, J. and Xia, F., 2015. Urban Land Use and Land Cover Classification Using Remotely Sensed SAR Data through Deep Belief Networks. *Journal of Sensors*, 2015, 10 pages.

McCune, B. and Keon, D., 2002. Equations for potential annual direct incident radiation and heat load index. *Journal of Vegetation Science*, 13, pp. 603-606.

Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26 (1), pp. 217-222.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Viena, Austria.

Shary, P.A., 1991. The second derivative topographic method. In: *Stepanov I. (Ed.) The Geometry of the Earth Surface Structures*. Pushchino Research Centre Press: Pushchino, USSR.

Wilson J. P. , 2012. Digital terrain modeling. *Geomorphology*, 137, pp. 107-121.

ACKNOWLEDGEMENTS

Authors thank ISPRS Working group III/4 for organizing benchmark on urban object detection. Acknowledgements also go for *BSF Swissphoto* for providing ISPRS with the Postdam input data set.

Article submitted on 10 July 2016