This is an excerpt from the thesis "*Machine Learning-based Regression for Characterization of Urban Environments with Sentinel-2*".

Please contact Henrik Schrade for a full version of the thesis.

Extended works on this topic are documented in: Geiß, C., Schrade, H., Aravena Pelizari, P., and Taubenböck, H. (2020): Multistrategy Ensemble Regression for Mapping of Built-Up Height and Density with Sentinel-2 Data. ISPRS Journal of Photogrammetry and Remote Sensing, 170, 57–71.

Universität Augsburg
Institut für Geographie

Machine Learning-based Regression for Characterization of Urban Environments with Sentinel-2

Master Thesis

For the acquisition of the academic degree

Master of Science Geoinformatics

Supervisors:

Prof. Dr. Sabine Timpf

Dr. Christian Geiß (DLR)

Schrade, Henrik

Matriculation Number: 1411240

Master of Science Geoinformatics, 5th Semester

Submission date: 17.11.2017

Abstract

Knowing the characteristics of urban environments is crucial for managing cities and planning an infrastructure which satisfies the needs of its inhabitants. This knowledge is also essential to assess the effects of natural hazards and, amongst others, it allows estimations about the amount of people that could be affected. For that reason up-to-date information about urban environments is of high interest. The average built-up height and the share of built-up area are two basic parameters which allow conclusions to be drawn about existing buildings and about the amount of people living in an area. However, a manual recording of these parameters is often not possible due to too fast and uncontrolled urban growth, and many surveying techniques which provide a high accuracy, e.g. LIDAR, are too expensive for a comprehensive application. Thus, the aim of this master thesis was to develop a method which provides exhaustive, up-to-date and accurate information about both parameters mainly based on freely available data.

In this study, satellite imagery recorded by ESA's Sentinel-2 satellite was chosen as data basis, because it is freely accessible, globally available, has a high resolution of 10 meters and with a revisit time of five days it is always up-to-date. In the first step of the method presented several sets of features, like mathematical morphologies, textures and statistical features, were derived from Sentinel-2 scenes showing the urban areas of interest. Subsequently, based on these features the average built-up height and the share of built-up area were predicted with four different regression algorithms: Random forest, Gaussian process regression, neural network and support vector regression. Afterwards, the single predictions were combined into a final result via an ensemble learning technique. Within this study stacked generalization and local selection were applied as ensemble learning approaches and their performance was compared. Before the prediction the four regressors had been trained with the features calculated from the Sentinel-2 imagery and with reference data derived from TandDEM-X data, which is quite costly. However, since the aim of this study was to develop a low-priced method, the use of expensive training data for every prediction was contrary to the goal set. In order to overcome this drawback and reduce the necessary usage of TanDEM-X data, a domain adaptation procedure was integrated. In the domain adaptation process the four regressors were trained on a Sentinel-2 scene, the so-called source domain, where reference data were available. Afterwards, the regressors predicted the average built-up height and the share of built-up area for another scene, the target domain, where only Sentinel-2 imagery was present. In the end, the single predictions of the four different regressors were again combined via an ensemble learning procedure.

Experimental results were obtained for the cities of Berlin, Cologne, Hamburg and Munich for which reference data were available. Each city was separately used as source domain for the other three cities so that the accuracy of the presented method could be assessed via the available reference data. Finally, the domain adaptation approach developed in this study had a mean absolute error (MAE) of 3.97 meters on average regarding the average built-up height and a MAE of 9.05 % on average regarding the share of built-up area. However, depending on the combination of source and target domain the MAE can vary a lot and under optimal conditions a MAE of 1.23 meters or 3.73 % can be achieved.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AdaBoost	Boosting Algorithm that Adjusts Adaptively
ANN	Artificial Neural Network
ARE	Absolute Relative Error
CFS	Correlation-based Feature Selector
DA	Domain Adaptation
DEM	Digital Elevation Modell
DMP	Derivate of the Morphological Profile
ESA	European Space Agency
GLCM	Grey Level Co-occurrence Matrix
GP	Gaussian Process Regression
GUF	Global Urban Footprint
IQR	Interquartile Range
LIDAR	Light Detection And Ranging
LS	Local Selection
MAE	Mean Absolute Error
MBI	Morphological Building Index
ME	Mean Error
MLP	Multi-Layer Perceptron
MO	Morphological Operator
MSI	Morphological Shadow Index
NDVI	Normalized Vegetation Index
NIR	Near-Infrared
NN	Neural Network
RBF	Gaussain Radial Basis Function
RF	Random Forest
RMSE	Root Mean Square Error
RSM	Random Subspace Method

SAR	Synthetic Aperture Radar
SE	Structuring Element
SG	Stacked Generalization
SVM	Support Vector Machine
SVR / SV	Support Vector Rgeression
VHR	Very High Resolution

1 Introduction

In recent years, global urbanisation has increasingly posed a challenge for city planners worldwide and it can be assumed that the situation will not become less dynamic in the nearby future. It is expected that the bulk of the world's population growth over the next decades will take place in the urban areas of developing countries (Jiang, O'Neil 2017). Many of the cities affected, and especially their informal settlements, have grown too fast for a manual record of the changes of building stock and population increase. However, it is important to know the structural characteristics of urban environments, e.g. for analysing urban processes or estimating the actual population size. Exact information on the population number in an area is crucial for the administration of urban areas and for the provision of a sufficient infrastructure which meets the basic needs of the inhabitants, like clean water, food, shelter, basic health care or electricity. Moreover, in case of a natural disaster it is essential to know how many people are affected in order to estimate how many emergency shelters and how much emergency supply are necessary. Current events, like hurricane Irma which devastated parts of the Caribbean in September 2017 or the earthquake that hit Mexico City on the 19th of September 2017, illustrate the necessity of such data. For the characterization of urban environments as well as for the assessment of the damage potential of natural risks the average built-up height and the share of built-up area in a settlement area are important parameters. Since in-situ measurements are often not available or complete, remote sensing represents an appropriate way for a comprehensive and continuous data acquisition.

Since building density and height are two of the most important aspects of urban characterization, there have been various studies presenting a broad variety of methods for the derivation of these two parameters from remote sensing data. Information recorded by spaceborne Synthetic Aperture Radar (SAR) systems are a widely used data basis to detect building densities and heights (e.g. Brunner et al. 2010; Guida et al. 2010; Tison et al. 2004) or building density based on their polarimetric backscattering properties (Kajimoto, Susaki 2013). However, the complex scattering mechanisms of urban structures often have negative effects on the accuracy of building height estimations from SAR data (Zhu et al. 2012). Another approach to building height detection is the utilization of the airborne light detection and ranging (LIDAR) technique, which provides detailed height information of a surveyed area with a very high resolution (e.g. Meng et al. 2009; Rottensteiner et al. 2007; Verma et al. 2006; Dakowicz et al. 2005). Nowadays, the application of LIDAR is restricted by the very high provision costs for large-scale LIDAR data. Furthermore, high-resolution optical satellite imagery is used to detect building height, e.g. through detecting building shadows (Kim et al. 2007; Dare 2005) or using stereo satellite images (Alobeid et al. 2009). Both methods need expensive data pre-processing in the form of building detection or the correct alignment of images. A relatively new alternative to gain height information of urban areas is supplied by the TanDEM-X mission, a spaceborne radar interferometer which provides a global digital elevation model (DEM) with a resolution of approximately 12 meters (Krieger et al. 2007). However, opposing the direct provision of accurate global height data are the high acquisition costs, which restrict the usage of these data at a global scale.

In this study, a method is presented which combines information from selected TanDEM-X scenes with global available optical data from the Sentinel-2 mission in order to calculate the average built-up height and the share-of built up area in urban environments. ESA's Sentinel-2 satellite captures images from the surface of the earth with a ground resolution of 10 meters and bands covering the visible and near infrared (NIR) spectrum. Moreover, the relatively short revisit time of five days or even less provide up-to-date records which are freely available via a data hub (Drusch et al. 2012). In order to increase the amount of input information several spectral, morphological and textural features are derived from the initial Sentinel-2 imagery.

First of all, the four Sentinel-2 bands (Red, Green, Blue, NIR) and their spectral information can be seen as initial pointers for the derivation of building heights and the share of built-up area, since man-made structures have higher values in the visible spectrum whereas vegetation is more striking in NIR. This circumstance can be used to distinguish between sealed and non-sealed areas and thus help to determine the share of built-up area. Different combinations of the four bands, like the well-known Normalized Density Vegetation Index (NDVI), increase the amount of spectral information and thus have the ability to improve the final classification or regression result (Zhang et al. 2017). In an image morphological segmentation procedures extract the geometric shape of structures which are brighter or darker than their surroundings. Since buildings are generally bright structures which cast shadows, morphological segmentation is ideally suitable for the detection of buildings in optical images (Pesaresi, Benediktsson 2001). The third group of features used in this study are textures which are concerned with the statistical distribution of grey tones in an image. A widely accepted method to extract textural features from remote sensing images is the Grey Level Co-occurrence Matrix (GLCM) (Shanmugan 1981). Textural features extracted with the GLCM were already used for extracting quantitative information of building density and have improved the overall result (Karathanassi et al. 2000).

The previously presented features are used in combination with TanDEM-X data to train machine-learning algorithms for regressions which afterwards are used to predict the average built-up height and the share of built-up area for urban scenes solely based on the Sentinel-2 features. With this approach a possibility is given to provide comprehensive height and share of built-up information. Thus, the main drawback of the TanDEM-X data, their restricted availability, is tried to overcome. In previous studies with similar approaches different types of machine-learning algorithms were used to calculate characteristics of urban environments. Xian and Crane (2005) built regression trees trained with Landsat imagery for the prediction of percentage coverage of impervious surface in urban areas, whereas Hu and Weng (2009) utilized Artificial Neural Networks (ANN). Another approach was presented by Dell'Acqua and Gamba (2003), who estimated building density at SAR satellite images using co-occurrence matrices for texture extraction and neurofuzzy classifiers for density prediction. Zhang et al. (2017) trained a Support Vector Regression (SVR) algorithm for urban density estimation from optical data. In general, few studies can be found which deal with the application of machine-learning algorithms for the estimation of the share of built-up area or similar problems and even less studies attend to the prediction of average built-up height.

In contrast to related studies, the method presented in this work does not only fall back on one type of algorithm but combines the results of the four following different regression algorithms in an ensemble learning approach: Random Forest (RF), Gaussian Process Regression (GP), Artificial Neural Networks (NN) and Support Vector Regression (SV). All four algorithms were trained separately with features derived from Sentinel-2 imagery of Berlin, Cologne, Hamburg and Munich. From Tandem-X data and the global urban footprint (GUF) covering the same area the share of built-up area and the average built-up height were calculated with 200, 500 and 800 meter resolution with a method presented by Geiß et al. (2017). These parameter values were utilized to train the regressors and to test the accuracies of their predictions.

Subsequent to the training procedure, the hyper-parameters of the regressors were tuned via a grid search and the algorithms were boosted with an adapted version of the Adaboost.RT boosting algorithm (Solomatine, Shrestha 2004). The original AdaBoost.RT algorithm was designed to improve the prediction accuracy of weak learners. However, after tuning the regressors cannot be designated as weak learners anymore and thus the standard AdaBoost.RT approach is no longer suitable. In this study an adapted version of AdaBoost.RT was developed by introducing the random subspace method of Ho (1998) to it. This extended AdaBoost.RT algorithm is capable of improving the prediction accuracies of already well-working regressors. After boosting, predictions for the share of built-up area and the average built-up height are made for parts of each scene which were not recognised during training.

In this study the four regressors were trained separately on the Sentinel-2 scenes of Berlin, Cologne, Hamburg and Munich. Subsequently, each regressor predicted the average built-up height and the share of built-up area for every scene. Consequently, there were four different predictions available per scene and parameter, but only one is needed. In order to solve this problem without losing already gathered information, the predictions of the four single boosted regressors were aggregated for each Sentinel-2 scene into a single final outcome via an ensemble learning technique. Therefore, two different ensemble learning strategies were tested and compared within this study: local selection presented by Bruzzone and Melgani (2005) and stacked generalization which was developed by Wolpert in 1992. It was assumed that at least one of the tested ensemble learning procedures would lead to predictions which would be more robust than the single predictions of the boosted regressors and at the same time at least as accurate as the predictions of the boosted regressors.

In the method outlined so far the regressors have made predictions for the same scene they were trained and tested on. For the given four scenes Tandem-X data, and thus values for the average built-up height and the share of built-up area, are available. However, the long-term goal is to train the regressors on a scene where the prediction parameters are available and make predictions for scenes where this information is not at disposal. The transfer of a regressor from one scene, also called source domain, to another scene or target domain is called transfer learning or domain adaptation (Margolis 2011). In the second part of this study the presented method was executed a second time, in which the regressors were trained on one scene and predict the average built-up height and the share of built-up area for the other three scenes. Since reference data are available for all four scenes, the prediction accuracy of the domain adaptation can be assessed. However, during test runs it became obvious that the four domains were too different to apply a simple transfer of regressors. In order to align the input data of the source and target domain, different methods were utilized in previous studies, like selecting only very similar features for prediction (Uguroglu, Carbonell 2011) or transforming the data (Pan et al. 2011). Most of the time these methods are work- and time-consuming and thus the choice fell on histogram matching, which is an easy to implement but nevertheless suitable normalization method for domain adaptation (Matasci et al. 2015b). Consequently, a histogram matching between the source and target domains was conducted prior to the calculation of features. It was expected that a domain adaptation conducted with previously adjusted input data would provide sufficient results in combination with the developed method.

The method presented in this study combines several well-known and well-studied techniques in a new way which, so far, has not been presented. It represents an efficient tool for the prediction of the average built-up height and share of built-up area in urban environments, which are important parameters for a broad variety of scientific issues. Also, it combines the advantages of the data from the Sentinel-2 mission, free global availability, with those of the TanDEM-X data which is very accurate height information. Thus, the presented method can be seen as the basis for the future provision of comprehensive built-up height information.

The study at hand is structured in the following way: The first part deals with the study sites, the input TanDEM-X and Sentinel-2 data and the features derived from them. Section two describes the method in detail, the results of which are presented in section three and subsequently discussed in the fourth chapter. Finally, a conclusion is drawn in section five.

2 Methodology

The methodical approach presented in this study is subdivided into four major parts. In the first step different types of features are derived from Sentinel-2 imagery, including morphological image descriptors, textures and band statistics. Additional features are computed by subtracting Sentinel-2 bands from each other in various combinations (i.e. band ratios). This provision of a broad variety of features is expected to be a solid basis for the prediction of building density and building height in the following step.

Second, reference data for the average built-up height and the share of built-up area, which were computed according to the method of Geiß et al. (2017) from TanDEM-X data, are combined with the aforementioned features. This is the input for the training of multiple regression algorithms. The latter comprise random forest (RF), Gaussian process regression (GP), neural network (NN) and support vector regression (SV). Subsequently, the hyperparameters of the algorithms are optimized via a tuning procedure. Furthermore, a boosting algorithm is applied on each regressor in order to further improve their prediction abilities. After training and testing the accuracy of the regressors predictions on the built-up height and the share of built-up area are made for further labelled data which form the basis for the later ensemble learning.

The goal of the third section is to combine the four single predictions of the applied boosted regressors with an ensemble learning approach to a single final result. With Local Selection (LS) and Stacked Generalization (SG) two different ensemble learning procedures are applied and their results are compared.

In the final part of the methodology an approach for a domain adaptation (DA) process based on histogram matching is presented. The DA enables the transfer of regressors which were priorly trained on a source Sentinel-2 scene (the so-called source domain) on a spatially disjunct target domain.

This is an excerpt from the thesis “*Machine Learning-based Regression for Characterization of Urban Environments with Sentinel-2*”.

Please contact Henrik Schrade for a full version of the thesis.

Extended works on this topic are documented in: Geiß, C., Schrade, H., Aravena Pelizari, P., and Taubenböck, H. (2020): Multistrategy Ensemble Regression for Mapping of Built-Up Height and Density with Sentinel-2 Data. ISPRS Journal of Photogrammetry and Remote Sensing, 170, 57–71.

List of Literature

Alobeid A., Jacobsen K., Heipke C. (2009): Building height estimation in urban areas from very high resolution satellite stereo images. In: ISPRS Hannover Workshop, pp. 2-5.

Basak D., Pal S., Patranabis D. C. (2007): Support vector regression. In: Neural Information Processing-Letters and Reviews, 11(10), pp. 203-224.

Bazi Y., Bruzzone L., Melgani F. (2005): An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. In: IEEE Transactions on Geoscience and Remote Sensing, 43(4), pp. 874-887.

Ben-David S. (2009): On the training/Test Distributions Gap: A Data Representation Learning Framework, In: Quionero-Candela J., Sugiyama M., Schwaighofer A., Lawrence N. D. (2009): Dataset shift in machine learning. The MIT Press Scholarship Online. 2013.

Benediktsson J. A., Pesaresi M., Amason K. (2003): Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. In: IEEE Transactions on Geoscience and Remote Sensing, 41(9), pp. 1940-1949.

Bergmeir C. N., Benítez Sánchez J. M. (2012): Neural networks in R using the Stuttgart neural network simulator: RSNNS. American Statistical Association. In: Journal of Statistical Software, 46(7), pp. 1-26.

Bishop C. M. (1995): Neural networks for pattern recognition. Oxford university press.

Blaschke T. (2010): Object based image analysis for remote sensing. In: ISPRS journal of photogrammetry and remote sensing, 65(1), pp. 2-16.

Breiman L. (2001): Random forests. In: Machine learning, 45(1), pp. 5-32.

Breiman L. (1996a): Stacked regressions. In: Machine learning, 24(1), pp. 49-64.

Breiman L. (1996b): Bagging predictors. In: Machine learning, 24(2), pp. 123-140.

Briem G. J., Benediktsson J. A., Sveinsson J. R. (2002): Multiple classifiers applied to multisource remote sensing data. In: IEEE transactions on geoscience and remote sensing, 40(10), pp. 2291-2299.

Brunner D., Lemoine G., Bruzzone L., Greidanus H. (2010): Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique. In: IEEE Transactions on Geoscience and Remote Sensing, 48(3), pp. 1487-1504.

Bruzzone L., Melgani F. (2005): Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data. In: IEEE Transactions on Geoscience and Remote Sensing, 43(1), pp. 159-174.

Camps-Valls G., Tuia D., Bruzzone L., Benediktsson J. A. (2014): Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. In: IEEE Signal Processing Magazine, 31(1), pp. 45-54.

Coltuc D., Bolon P., Chassery J. M. (2006): Exact histogram specification. In: IEEE Transactions on Image Processing, 15(5), pp. 1143-1152.

Dalla Mura M., Benediktsson J. A., Waske B., Bruzzone L. (2010): Morphological attribute profiles for the analysis of very high resolution images. In: IEEE Transactions on Geoscience and Remote Sensing, 48(10), pp. 3747-3762.

Dakowicz M., Gold C. M., Kidner D. B. (2005): Building reconstruction using LIDAR data. In: In Proceedings 4th ISPRS Workshop on Dynamic and Multi-dimensional GIS, pp. 156-161.

Dare P. M. (2005): Shadow analysis in high-resolution satellite imagery of urban areas. In: Photogrammetric Engineering & Remote Sensing, 71(2), pp. 169-177.

Dayhoff J. E., DeLeo J. M. (2001): Artificial neural networks. In: Cancer, 91(S8), pp. 1615-1635.

Dell'Acqua F., Gamba P. (2003): Texture-based characterization of urban environments on satellite SAR images. In: IEEE Transactions on Geoscience and Remote Sensing, 41(1), pp. 153-159.

Dietterich T. G. (2002): Ensemble learning. In: The handbook of brain theory and neural networks, 2, pp. 110-125.

Do C. B. (2007): Gaussian processes. https://www.cs229.stanford.edu/section/cs229-gaussian_processes.pdf. Accessed 29 August 2017 (29.08.2017).

Drusch M., Del Bello U., Carlier S., Colin O., Fernandez V., Gascon F., Hoersch B., Isola C., Laberinti P., Martimort P., Meygret A., Spoto F., Sy O., Marchese F., Bargellini P. (2012): Sentinel-2: ESA's optical high-resolution mission for GMES operational services. In: Remote Sensing of Environment, 120, pp. 25-36.

Du S., Zhang F., Zhang X. (2015): Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach. In: ISPRS journal of photogrammetry and remote sensing, 105, pp. 107-119.

Freund Y., Schapire R. E. (1995): A desicion-theoretic generalization of on-line learning and an application to boosting. In: European conference on computational learning theory, Springer, Berlin, Heidelberg, pp. 23-37.

García-Pedrajas N., Ortiz-Boyer D. (2008): Boosting random subspace method. In: Neural Networks, 21(9), pp. 1344-1362.

Geiß C., Wurm M., Taubenböck H. (2017): Towards large-area morphologic characterization of urban environments using the TanDEM-X mission and Sentinel-2. In: IEEE Urban Remote Sensing Event (JURSE), 2017 Joint, pp. 1-4.

Geiß C., Klotz M., Schmitt A., Taubenböck H. (2016): Object-based morphological profiles for classification of remote sensing imagery. In: IEEE Transactions on Geoscience and Remote Sensing, 54(10), pp. 5952-5963.

Geiß C., Jilge M., Lakes T., Taubenböck H. (2015): Estimation of seismic vulnerability levels of urban structures with multisensor remote sensing. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(5), pp. 1913-1936.

Gonzalez R. C., Woods R. E. (2002): Digital Image Processing. Second Edition. Prentice Hall.

Grömping U. (2009): Variable importance assessment in regression: linear regression versus random forest. In: The American Statistician, 63(4), pp. 308-319.

Guida R., Iodice A., Riccio D. (2010): Height retrieval of isolated buildings from single high-resolution SAR images. In: IEEE Transactions on Geoscience and Remote Sensing, 48(7), pp. 2967-2979.

Gunn S. R. (1998): Support vector machines for classification and regression. In: ISIS technical report, 14, pp. 85-86.

Härdle W., Simar L. (2007): Applied multivariate statistical analysis. Springer Science & Business Media, Springer-Verlag Berlin Heidelberg.

Hall M. A. (1999): Correlation-based feature selection for machine learning (Doctoral dissertation, The University of Waikato).

Haralick R. M. (1979): Statistical and structural approaches to texture. In: Proceedings of the IEEE, 67(5), pp. 786-804.

Haralick R. M., Shanmugam K. (1973): Textural features for image classification. In: IEEE Transactions on systems, man, and cybernetics, 3(6), pp. 610-621.

Ho T. K. (1998): The random subspace method for constructing decision forests. In: IEEE transactions on pattern analysis and machine intelligence, 20(8), pp. 832-844.

Hornik K., Stinchcombe M., White H. (1989): Multilayer feedforward networks are universal approximators. In: Neural networks, 2(5), pp. 359-366.

Hu X., Weng Q. (2009): Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. In: Remote Sensing of Environment, 113(10), pp. 2089-2102.

Huang X., Lu Q., Zhang L. (2014): A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. In: ISPRS Journal of Photogrammetry and Remote Sensing, 90, pp. 36-48.

Hultquist C., Chen G., Zhao K. (2014): A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests. Remote sensing letters, 5(8), pp. 723-732.

Jiang L., O'Neill B. C. (2017): Global urbanization projections for the Shared Socioeconomic Pathways. In: Global Environmental Change, 42, pp. 193-199.

Jiang J. (2008): A literature survey on domain adaptation of statistical classifiers. http://www.sifaka. cs.uiuc.edu/jiang4/domainadaptation/survey (22.08.2017).

Kajimoto M., Susaki J. (2013): Urban density estimation from polarimetric SAR images based on a POA correction method. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6(3), pp. 1418-1429.

Karathanassi V., Iossifidis C. H., Rokos D. (2000): A texture-based classification method for classifying built areas according to their density. In: International Journal of Remote Sensing, 21(9), pp. 1807-1823.

Kim T., Javzandulam T., Lee T. Y. (2007): Semiautomatic reconstruction of building height and footprints from single satellite images. In: IEEE Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, pp. 4737-4740.

Kira K., Rendell L. A. (1992): A practical approach to feature selection. In: Sleeman D., Edwards P. (ED.)(1992): Proceedings of the ninth international workshop on Machine learning, San Mateo, Morgan Kaufman Publishers, Inc., pp. 249-256.

Kopal V., Holena V. K. M. (2015): Comparing Non-Linear Regression Methods on Black-Box Optimization Benchmarks. In: Yaghob J. (Ed.) (2015): ITAT 2015, Prague, Charles University in Prague pp. 135–142.

Krieger G., Moreira A., Fiedler H., Hajnsek I., Werner M., Younis M., Zink M. (2007): TanDEM-X: A satellite formation for high-resolution SAR interferometry. In: IEEE Transactions on Geoscience and Remote Sensing, 45(11), pp. 3317-3341.

Lee L. (1999): Measures of distributional similarity. In: Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, Association for Computational Linguistics, pp. 25-32.

Liaw A., Wiener M. (2002): Classification and regression by randomForest. In: R news, 2(3), pp. 18-22.

Lönnblad L., Peterson C., Rögnvalsson T. (1992): Pattern recognition in high energy physics with artificial neural networks—JETNET 2.0. In: Computer Physics Communications, 70(1), pp. 167-182.

Löw F., Conrad C., Michel U. (2015): Decision fusion and non-parametric classifiers for land use mapping using multi-temporal RapidEye data. In: ISPRS Journal of Photogrammetry and Remote Sensing, 108, pp. 191-204.

Lopez G., Rubio M. A., Martınez M., Batlles F. J. (2001): Estimation of hourly global photosynthetically active radiation using artificial neural network models. In: Agricultural and forest Meteorology, 107(4), pp. 279-291.

Martin M. E., Aber J. D. (1997): High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. In: Ecological applications, 7(2), pp. 431-443.

Matasci G., Volpi M., Kanevski M., Bruzzone L., Tuia,D. (2015a): Semisupervised Transfer Component Analysis for Domain Adaption in Remote Sensing Image Classification, In: IEEE Transactions on Geoscience and Remote Sensing, Vol.53(7), pp. 3550-3564.

Matasci G., Longbotham N., Pacifici F., Kanevski M., Tuia D. (2015b): Understanding angular effects in VHR imagery and their significance for urban land-cover model portability: A study of two multi-angle in-track image sequences. In: ISPRS Journal of Photogrammetry and Remote Sensing, 107, pp. 99-111.

Margolis A. (2011): A literature review of domain adaptation with unlabelled data. In: Tec. Report 2011, University of Washington, pp. 1-42.

Mellor A., Boukir S. (2017): Exploring diversity in ensemble classification: Applications in large area land cover mapping. In: ISPRS Journal of Photogrammetry and Remote Sensing, 129, (July 2017), pp. 151–161.

Melo J. (2012): Gaussian processes for regression: a tutorial. Technical Report 2012, Faculty of Engineering, University of Porto, Portugal.

Meng X., Wang L., Currit N. (2009): Morphology-based building detection from airborne LIDAR data. In: Photogrammetric Engineering & Remote Sensing, 75(4), pp. 437-442.

Meyer D., Wien F. T. (2001): Support vector machines. In: R News, 1(3), pp. 23-26.

Okujeni A., van der Linden S., Suess S., Hostert P. (2017): Ensemble Learning From Synthetically Mixed Training Data for Quantifying Urban Land Cover With Support Vector Regression. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(4), pp. 1640-1650.

Pacifici F., Chini M., Emery W. J. (2009): A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. In: Remote Sensing of Environment, 113(6), pp. 1276-1292.

Panov P., Džeroski S. (2007): Combining Bagging and Random Subspaces to Create Better Ensembles. In: R. Berthold M., Shawe-Taylor J., Lavrač N. (Dds)(2007): Advances in Intelligent Data Analysis VII. IDA 2007. Lecture Notes in Computer Science, 4723. Springer, Berlin, Heidelberg, pp. 118-129.

Persello C. (2013): Interactive domain adaptation for the classification of remote sensing images using active learning. In: IEEE Geoscience and Remote Sensing Letters, 10(4), pp. 736-740.

Pesaresi M., Gerhardinger A. (2011): Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4(1), pp. 16-26.

Pesaresi M., Gerhardinger A., Kayitakire F. (2008): A robust built-up area presence index by anisotropic rotation-invariant textural measure. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1(3), pp. 180-192.

Pesaresi M., Benediktsson J. A. (2001): A new approach for the morphological segmentation of high-resolution satellite imagery. In: IEEE transactions on Geoscience and Remote Sensing, 39(2), pp. 309-320.

Quinlan J. R. (1986): Induction of decision trees. In: Machine learning, 1(1), pp. 81-106.

Quinonero-Candela J., Rasmussen C. E., Williams C. K. (2007): Approximation methods for Gaussian process regression. In: Large-scale kernel machines, Neural Information Processing, Cambridge, MA, USA, MIT Press, pp. 203-224.

Rasmussen C. E., Williams C. K. (2006): Gaussian processes for machine learning (Vol. 1). London: Cambridge: MIT press.

Rasmussen C. E. (2003): Gaussian Processes in Machine Learning, In: Bousquet O., von Luxburg U., Rätsch G. (Eds.). (2011): Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany: August 4-16, 2003, Revised Lectures (Vol. 3176). Springer, pp. 63-71.

Richards J. A., Jia X. (2006): Remote sensing digital image analysis (Vol. 4). Berlin: Springer Verlag Berlin Heidelberg.

Robnik-Šikonja M., Kononenko I. (1997): An adaptation of Relief for attribute estimation in regression. In: Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97), pp. 296-304.

Rodriguez-Galiano V. F., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J. P. (2012): An assessment of the effectiveness of a random forest classifier for land-cover classification. In: ISPRS Journal of Photogrammetry and Remote Sensing, 67, pp. 93-104.

Rottensteiner F., Trinder J., Clode S., Kubik K. (2007): Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. In: ISPRS Journal of Photogrammetry and Remote Sensing, 62(2), pp. 135-149.

Rouse Jr J., Haas R. H., Schell J. A., Deering D. W. (1974): Monitoring vegetation systems in the Great Plains with ERTS. In: NASA. Goddard Space Flight Center 3d ERTS-1 Symp., 1, Sect. A pp. 309-317.

Schmidhuber J. (2015): Deep learning in neural networks: An overview. In: Neural networks, 61, pp. 85-117.

Shanmugan K. S., Narayanan V., Frost V. S., Stiles J. A., Holtzman J. C. (1981): Textural features for radar image analysis. In: IEEE Transactions on Geoscience and Remote Sensing, (3), pp. 153-156.

Shrestha D. L., Solomatine D. P. (2006): Experiments with AdaBoost. RT, an improved boosting scheme for regression. In: Neural computation, 18(7), pp. 1678-1710.

Skurichina M., Duin R. P. (2002): Bagging, boosting and the random subspace method for linear classifiers. In: Pattern Analysis & Applications, 5(2), pp. 121-135.

Solomatine D. P., Shrestha D. L. (2004): AdaBoost. RT: a boosting algorithm for regression problems. In: IEEE Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference, 2, pp. 1163-1168.

Sun S. (2013): A survey of multi-view machine learning. Neural Computing and Applications, 23(7-8), pp. 2031-2038.

Tao D., Tang X., Li X., Wu X. (2006): Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. In: IEEE transactions on pattern analysis and machine intelligence, 28(7), pp. 1088-1099.

Taubenböck H., Roth A., Esch T., Felbier A., Müller A., Dech S. (2011): The vision of mapping the global urban footprint using the TerraSAR-X and TanDEM-X mission. In Zlatanova, Ledoux, Fendel, Rumor (Eds.) (2012): Urban and regional data management, London, Taylor & Francis Group pp. 243–251.

Ting K. M., Witten I. H. (1997): Stacked Generalization: when does it work?, In: Working Paper Series 97/03, Hamilton, New Zealand: Department of Computer Science, University of Waik.

Tison C., Tupin F., Maitre H. (2004): Retrieval of building shapes from shadows in high resolution SAR interferometric images. In: Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International, 3, pp. 1788-1791.

Tooke T. R., Coops N. C., Webster J. (2014): Predicting building ages from LiDAR data with random forests for building energy modeling. In: Energy and Buildings, 68, pp. 603-610.

Tuia D., Persello C., Bruzzone L. (2016): Domain Adaption for the Classification of Remote Sensing Data – An Overview of recent advances, In: IEEE Geoscience and Remote Sensing Magazine, June 2016, pp. 41-57.

Tuia D., Munoz-Mari J., Gomez-Chova L., Malo J. (2013): Graph matching for adaptation in remote sensing. In: IEEE Transactions on Geoscience and Remote Sensing, 51(1), pp. 329-341.

Uguroglu S., Carbonell J. (2011): Feature selection for transfer learning. In: Gunopulos D., Hofmann T., Malerba D., Vazirgiannis M. (Eds) (2011): Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2011. Lecture Notes in Computer Science, 6913. Springer, Berlin, Heidelberg, pp. 430-442.

Vapnik V., Golowich S. E., Smola A. J. (1997): Support vector method for function approximation, regression estimation and signal processing. In: Advances in neural information processing systems, pp. 281-287.

Vapnik V. N., Chervonenkis A. J. (1974): Theory of pattern recognition. Nauka, In Russian

Verma V., Kumar R., Hsu S. (2006): 3d building detection and modeling from aerial lidar data. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, 2, pp. 2213-2220.

Verrelst J., Alonso L., Camps-Valls G., Delegido J., Moreno J. (2012): Retrieval of vegetation biophysical parameters using Gaussian process techniques. In: IEEE Transactions on Geoscience and Remote Sensing, 50(5), pp. 1832-1843.

Wang G., Ma J. (2011): Study of corporate credit risk prediction based on integrating boosting and random subspace. In: Expert Systems with Applications, 38(11), pp. 13871-13878.

Waske B., van der Linden S. (2008): Classifying multilevel imagery from SAR and optical sensors by decision fusion. In: IEEE Transactions on Geoscience and Remote Sensing, 46(5), pp. 1457-1466.

Wolpert D. H. (1992): Stacked generalization. In: Neural networks, 5(2), pp. 241-259.

Xian G., Crane M. (2005): Assessments of urban growth in the Tampa Bay watershed using remote sensing data. In: Remote Sensing of Environment, 97(2), pp. 203-215.

Yu L., Liu H. (2003): Feature selection for high-dimensional data: A fast correlation-based filter solution. In: ICML, 3, pp. 856-863.

Zhang T., Huang X., Wen D., Li J. (2017): Urban Building Density Estimation From High-Resolution Imagery Using Multiple Features and Support Vector Regression. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 1-16.

Zeileis A., Hornik K., Smola A., Karatzoglou A. (2004): kernlab-an S4 package for kernel methods in R. In: Journal of statistical software, 11(9), pp. 1-20.

Zhu Z., Woodcock C. E., Rogan J., Kellndorfer J. (2012): Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. In: Remote Sensing of Environment, 117, pp. 72-82.

Zhu G., Blumberg D. G. (2002): Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel. In: Remote sensing of Environment, 80(2), pp. 233-240.