Semantic labelling of building types. A comparison of two approaches using Random Forest and Deep Learning

ARIANE DROIN[1,2], MICHAEL WURM[2] & WOLFGANG SULZER[1]

Abstract: In the context of sustainable planning, knowledge about building type is crucial. Yet, this information is scarce and mostly inhomogeneous. In regard to the big-data era, two approaches for building type classification are presented based on different data basis. The first approach shows semantic classification of building footprints using a set of features (simple geometric, morphological and topological features) and the machine learning algorithm Random Forest. Very high accuracies for the federal states of Germany could be achieved with Kappa Coefficients between 0.87 and 0.98. The second framework presents the possibility to conduct semantic labelling of aerial images using Fully Convolutional Neural Networks. The gained accuracy in this case is a Kappa of 0.73 for the federal state of Berlin.

1 Introduction

In the general discourse of energy consumption, buildings make up for about 30-40% of the global final energy use consumption (INTERNATIONAL ENERGY AGENCY AND THE UNITED NATIONS ENVIRONMENT PROGRAMME 2018). However, the building sector is also known to have huge potentials in regard to energy savings and hence to officiate as key holder to meet the defined energy saving goals (STEEMERS & YUN 2009). Consequently, knowledge about building types is crucial in the context of heat demand calculations and predictions. In spite of the importance of this data, the availability of this information is scarce, not up-to-date and heterogeneous (HECHT et al. 2015).
Derivation and classification of building types based on remote sensing and Geographic Information System (GIS) data has been researched and conducted by several studies (e.g. MEINEL et al. 2009; BELGIU et al. 2014; WURM et al. 2016). The researched approaches however vary broadly depending on data basis, on local conditions, on the desired output and focus mostly on city applications.
The development and evolution of artificial intelligence makes it possible to emerge further into the field of big data and to solve tasks which were not feasible up to now. Hence, the presented study shows two approaches for semantic labelling of building types based on the assumption of the availability of different data basis. The first approach uses the machine learning algorithm Random Forest (RF) to derivate building types on two semantic stages (see Figure 1) based on Level of Detail 1 (LoD1) and census data. The second approach relies solely on remote sensing data and hence on the classification of different building types based on spectral data when

[1] Karl-Franzens Universität Graz, Department of Geography and Regional Sciences, Heinrichstraße 36, A-8010 Graz, Austria, E-Mail: wolfgang.sulzer@uni-graz.at
[2] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Department of Geo-Risks and Civil Security, Münchener Straße 20, D-82234 Weßling, Germany
E-Mail: [Ariane.Droin, Michael.Wurm]@dlr.de

building footprint data is not available. Up to now this task was not feasible, based on the broad intra-spectral variability of buildings, but with the emergence and evolution of Neural Networks, which can detect non-linear and complex relationships in the data, new fields arise (ZHANG et al. 2016; ZHU et al. 2017). Consequently, the second approach focuses on the semantic segmentation of building types based on aerial images.

The nomenclature for the semantic stages (see Fig. 1) is based on the nomenclature of the Institute of Housing and Environment (IWU), as these related types are used as basis for energy consumption modelling.

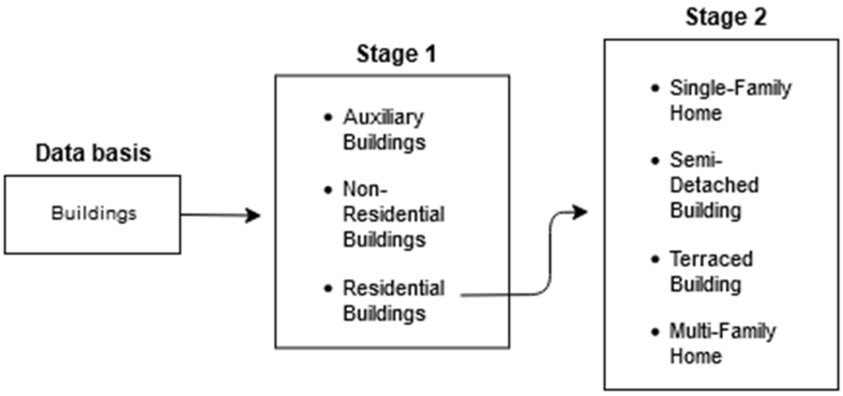


Fig. 1: Two-stage approach for semantic labelling of building types

2 Methods and data

2.1 Semantic labelling of building types using Random Forest

The building footprint dataset (LoD1) is acquired from the Federal Agency of Cartography and Geodesy (BKG) of Germany and contains ~51 million buildings. Additionally, the census data from 2011, provided in 100 m grid cells, is used for the generation of training and reference datasets as building type information is contained within the census (see Fig. 2).

The RF algorithm is based on the majority vote of a multitude of decision trees which are built using bootstrap aggregating (BREIMAN 2001). The efficiency of RF, and hence making it a state-of-the-art machine learning algorithm, lies in the fact that it takes random subsamples of data and features for building each single tree (bootstrap aggregating) and thus making it prune to overfitting. The remaining data that was not used in the building process of the trees is used to estimate the performance of each single tree (Out-Of-Bag (OOB) estimate) as well as the ensemble of trees. Furthermore, the importance of each feature can be estimated, which is useful information for restricting the calculation performance (RODRIGUEZ-GALLIANO 2012; LIAW & WIENER 2002).

For the classification task, different sets of features are chosen (see Tab. 1). It is assumed that the different building types have different morphological and topological properties based on their type (see Tab. 2).

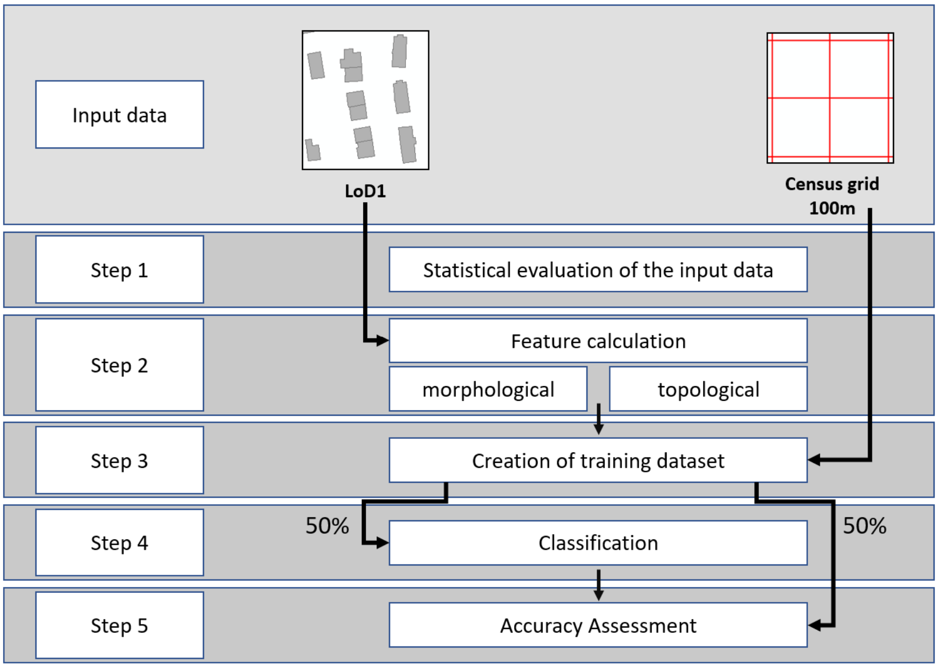


Fig. 2: General layout of the developed methodology for semantic labelling of LoD1 data

The feature sets are calculated for every single building and used for the classification task. The information of the building types of the first semantic stage is already included officially for almost every federal state of Germany (apart from Bremen, Saxony and Thuringia) and is hence used for prediction of the latter three. Using the defined features and RF, the information of the first semantic stage is predicted for those three states.

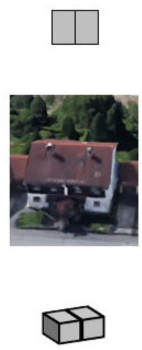
Tab. 1: Feature set used for Random Forest classification of building types

Simple Geometric Features:	Perimeter [m], Area [m²], Height [m], Proportion between Height and Area, Volume [m³]
Morphological Features	Detour, Range, Exchange, Cohesion, Proximity, Spin, Shape Index, Fractal Dimension
Topological Features	Consecutive neighbors, Dissolved area of neighboring buildings [m²], Dissolved perimeter [m], Relative area [%] of the building area compared to the dissolved area

For the prediction of the second semantic stage (see Fig. 1) training data is generated using the census grid cells. Buildings which are located in cells with only one building type are selected. However, as it is assumed that certain errors can occur, additional thresholds are defined, to ensure that only the desired building type is included (e.g. Single-Family Homes have no consecutive neighbors or Semi-Detached buildings must have exactly one neighbor and at least a share of 25% on the dissolved area).
For the classification procedure the default values of RF classifiers are used (500 trees and $\sqrt{n}$ features) and to reduce calculation time each federal state is classified separately. A random subsample of 50% is taken from the training data for the classification process while the other 50% are used for Accuracy Assessment. Furthermore, several set-ups are constructed to assess

the quality, applicability and conformity of the developed approach. The first set-up evaluated the performance of census data as training dataset. Therefore, a manually labelled and randomly selected training sample of 3,000 buildings footprints for the federal state of Berlin is generated. The second set-up assesses the influence and importance of the different feature sets on the classification result.

Tab. 2: Examples of different morphological properties for different residential building types (Image Source: Google Earth 2018)

	Single-Family Home	**Semi-Detached Building**	**Terraced Building**	**Multi-Family Home**	
Schema					
Perimeter Index	0.85	0.87	0.84	0.69	
Detour Index	0.89	0.87	0.84	0.69	
Range Index	0.80	0.76	0.72	0.53	

2.2 Semantic segmentation of building types using Deep Learning

Semantic segmentation of aerial images is the task of assigning each pixel a semantic meaning. Deep learning architectures have proven to be far more successful in this sense than conventional machine learning algorithms (LONG et al. 2015). The advantage of Deep Learning is that it can automatically learn abstract and discriminative features from the input data. Hence, over-specification or incompleteness of features is hindered (ZHU et al. 2017).
LONG et al. (2015) brought up Fully Convolutional Networks (FCN) where no fully connected layers are needed for semantic segmentation of images. For semantic segmentation of building types based on aerial images the FCN-vgg19 from the Visual Group of Oxford University (SIMONYAN & ZISSERMAN 2014) is used and for the architecture of the network, we refer to WURM et al. (2019) who deployed the network for semantic segmentation of informal settlements. In the proposed approach, we use orthophotos with a resolution of 40 cm in the visible spectrum.

For semantic segmentation, rasterization of the LoD1 data (using the first and the second semantic stage respectively) is carried out for the generation of training data. Additionally, training data using only the building footprints with no further semantic differentiation is also generated. This is done to be able to assess if the Random-Forest approach can also be applied on footprint data derived using Deep Learning algorithms. The general layout of the semantic segmentation approach using Deep Learning is shown Fig. 3.

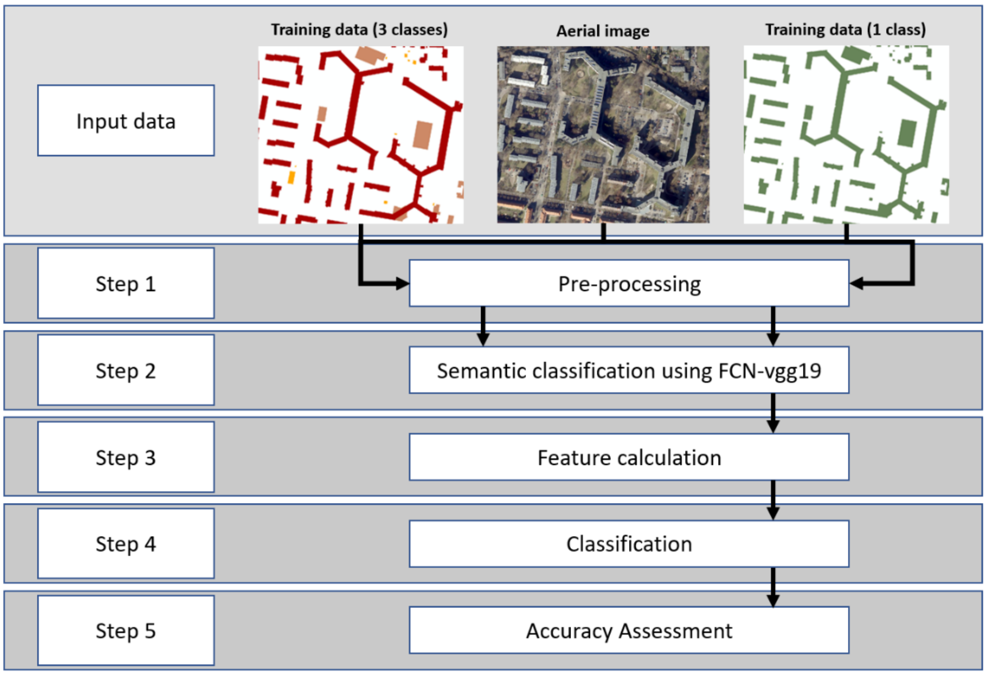


Fig. 3: General layout of the developed methodology for semantic segmentation of aerial images

The FCN algorithm works with image tiles with a size of 224×224 pixels, resulting in a training data set with 55,843 tiles without overlap. For assessing the accuracy, a 4-fold cross validation is carried out using the columns of the tiling process.

While the results of the semantic segmentation process on the first stage only need to be post-processed due to tiling-effects, the semantic segmented building footprints need to be further processed to gain a higher semantic meaning. Therefore, the morphological features from the RF-approach are calculated for every derived footprint. Based on the fact, that the resulting footprints are always one polygon and that the housing units cannot be separated, no topological features can be calculated. Furthermore, due to missing elevation data no features including height information could be included in the feature set. The accuracy of the results from semantic classification on the first stage on the Deep Learning derived footprints is assessed in two ways. On the one hand, 4,700 building footprints are annotated manually for the generation of a reference dataset. On the other hand, the LoD1 data with the included semantic information which intersect the derived building footprints are dissolved into one building and assigned the class with the biggest share on the dissolved area. Hence, 81,000 reference buildings for Berlin could be generated (see Fig. 4).

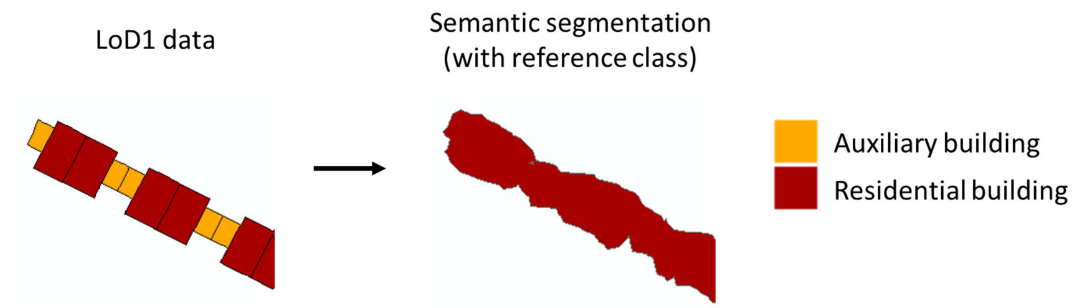


Fig. 4: Generation of reference data using the class with the largest share on the intersecting building

3 Results

For the RF approach, Kappa accuracies of 0.86 (Bremen) and 0.91 (Saxony and Thuringia) could be reached for the first semantic stage. The other federal states already had the first semantic stage information included. For the second semantic stage Overall Accuracy (OA) for every singly federal state is above 95% and Kappa is around 0.9 for the states of Baden-Württemberg and Saxony and above 0.9 for the other federal states (see Fig. 5).

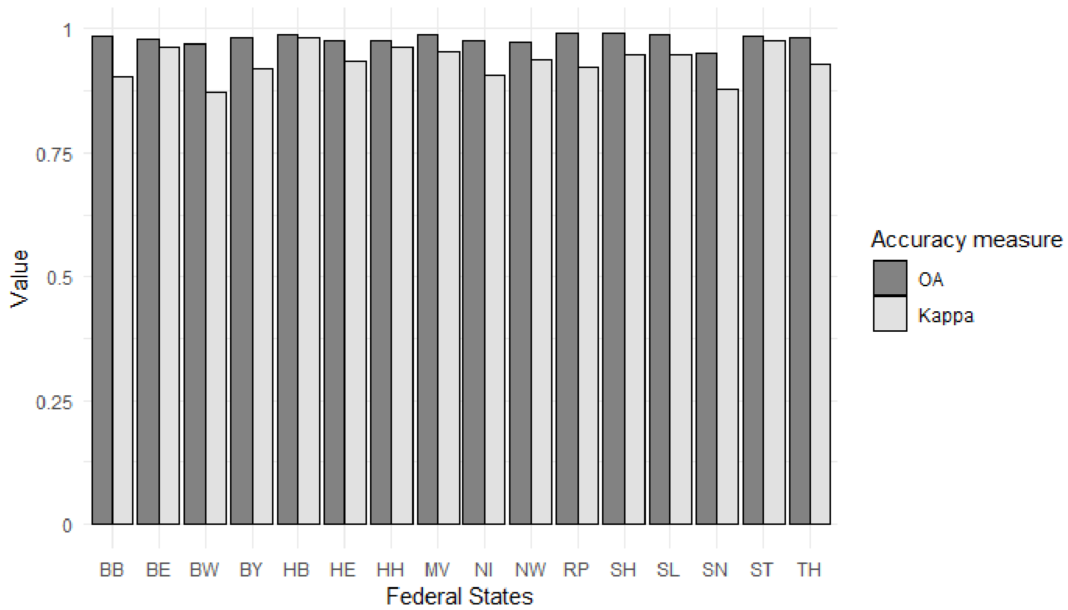


Fig. 5: Accuracy results for each federal state

Regarding the feature importance (see Fig. 6) differences between the different sets can be deduced. Simple geometric features, especially those containing height information, and topological features score the highest importance between the federal states. The indexed morphological features show big variations in their importance. The importance of the feature "Relative area" has the lowest variation and has therefore approximately the same importance in the classification process of each federal state.

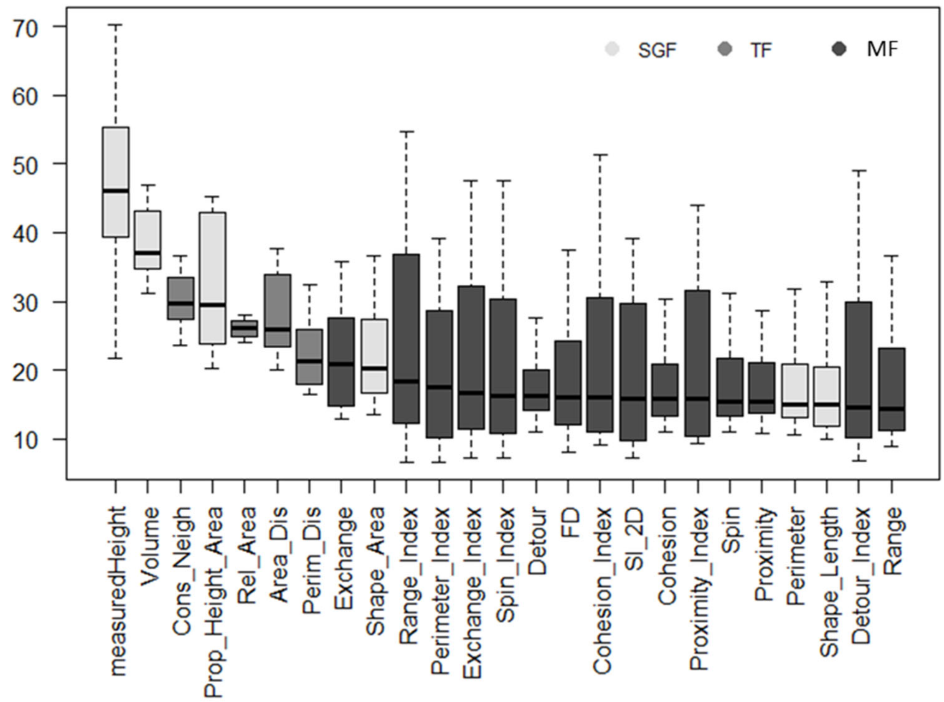


Fig. 6: Feature importance for the RF approach for residential building type classification (SGF = Simple geometric features, TF = Topological features and MF = morphological features)

The accuracy results for the evaluation of the usability of census data using a manually labelled reference data set for comparison yields quite resembling results, with a slightly lower accuracy for the manual data set (see Fig. 7).

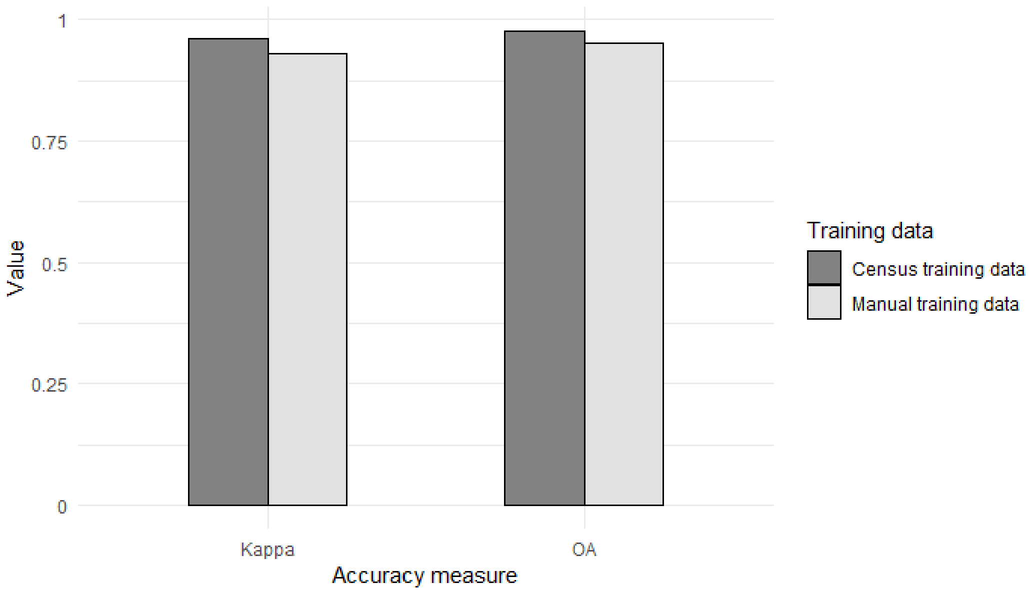


Fig. 7: Comparison of accuracy measures for reference data using different sets of training data

The impact on the accuracy when different sets of features are used (first set is with all features, second set is only with simple geometric and topological features combined and third set is only morphological features) show clear tendencies. The accuracy results show that there is a clear decrease in accuracy for every federal state when only morphological features are used, and only a slight decrease in accuracy when no morphological features are included (see Fig. 8). Furthermore, it can be stated, that the lowering of accuracy is considerably less in urban federal states than in the rural ones.

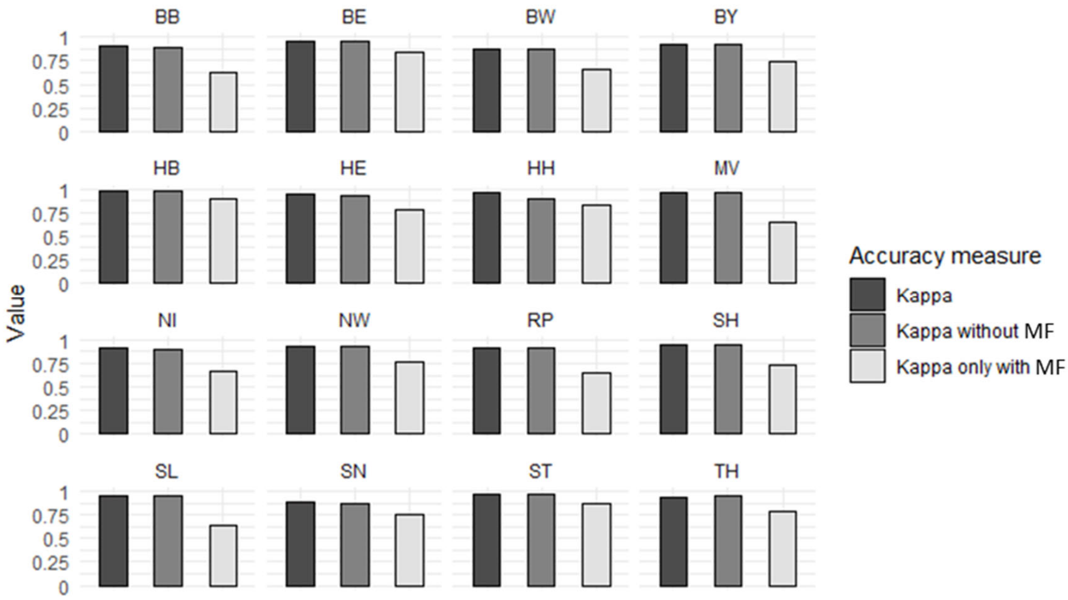


Fig. 8: Accuracy result for the different federal states using different sets of features (MF = morphological features)

The approach using Fully Convolutional networks results in an OA of 93% and a Kappa of 0.73 for the first semantic stage. If no semantic differentiation is conducted but only the detection of building footprint data is regarded a Kappa of 0.8 is achieved (see Fig. 9).

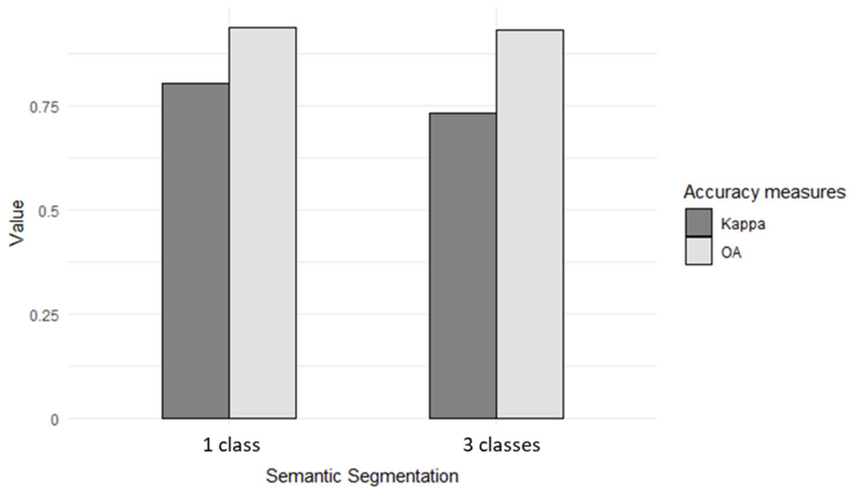


Fig. 9: Accuracy results for semantic segmentation of building footprints and the first semantic stage

Applying the Random Forest approach on the derived footprints with no semantic specification using the LoD1 data yielded low accuracies with a Kappa value of 0.32 and using the manually generated training and reference data yielded results with a Kappa value of 0.46 (see Fig. 10).

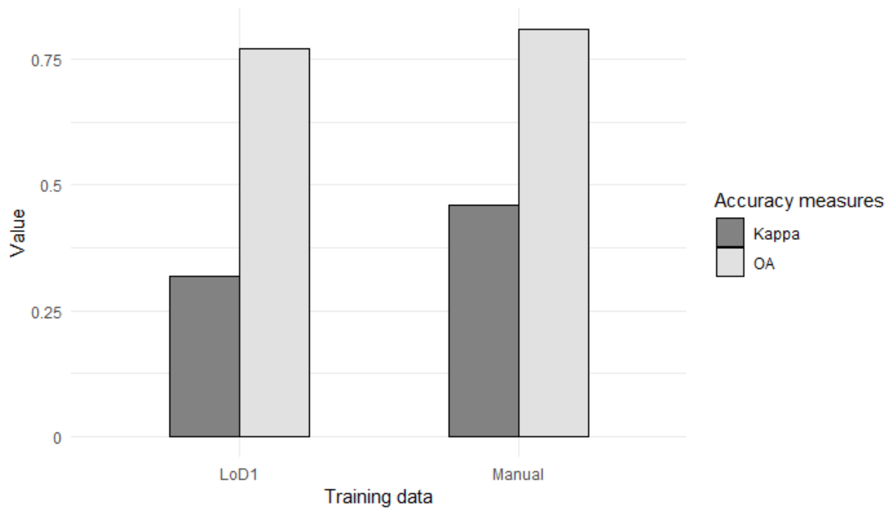


Fig. 10: Accuracy measures of Random Forest classification based on Deep Learning derived footprints using Random Forest

A visualisation of the results of the first and second semantic stage based on the different data basis can be taken from Fig. 11.

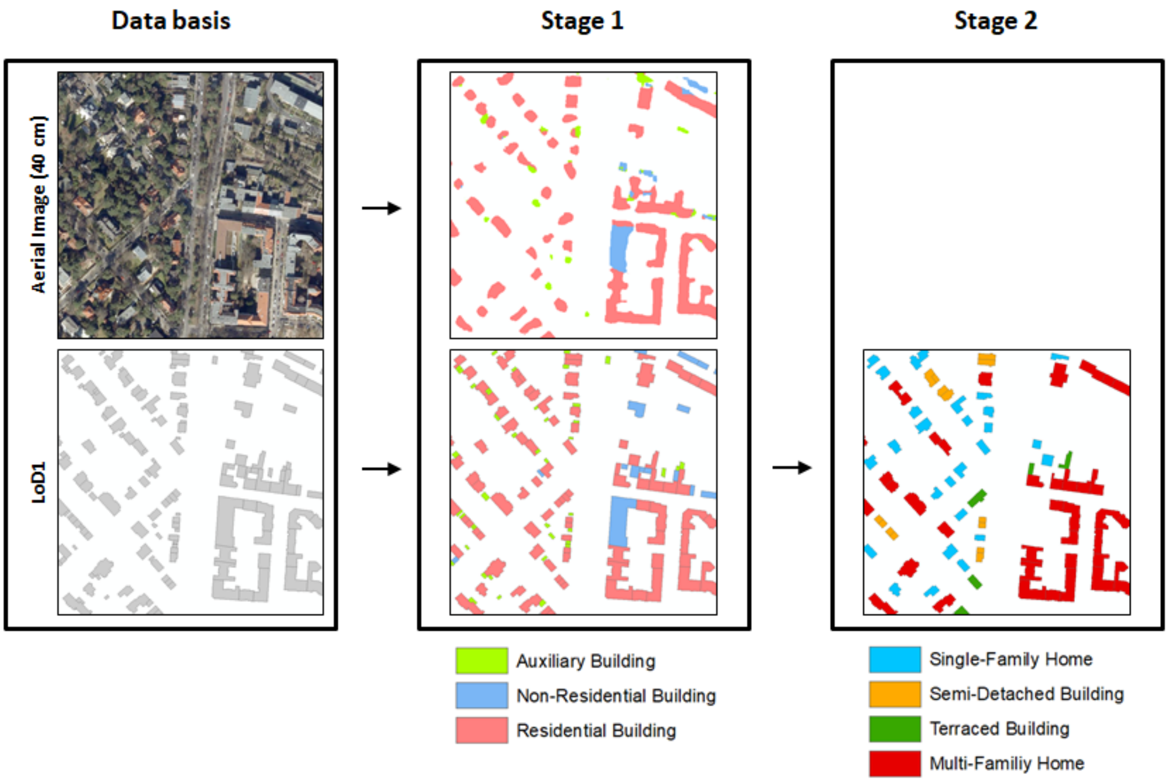


Fig. 11: Exemplary results for the two approaches

4 Discussion and Outlook

All accuracies of the RF are very high and thus, show that this approach can be applied for the derivation of building type information. Using features with topological and height information is important in this regard, as they are crucial in the process of distinguishing different building types. This is not only shown by the feature importance but also by the results of the set-up where different feature sets have been used. Even if morphological features do not seem to be as important as the ones mentioned before and no high accuracy can be gained when only morphological features are used, a stabilization of the classification result can be concluded (Kappa is lowered by ~0.2 without them). Hence, it is suggested to use these features as well. The higher Kappa value for the urban federal states, when only morphological features are used, are based on the fact that the forms of different building types in cities are more homogeneous than in the bigger rural states. The higher variation of importance for the indexed morphological features underlines this assumption. However, the exact reasons and implications for that need to be further researched.
When the results of the accuracy using census reference data is compared to the results of the manually generated reference data it can be deduced that the census data is indeed a viable source for training and reference data. The slight lowering in accuracy for the manual training data can be concluded on the one hand on the highly imbalanced amount of class samples, based on the random selection, and on the lower amounts of reference data.

The results of the first semantic stage for the Neural Network approach are promising, regarding the fact that building types are only differentiated based on spectral data. However, the variation of roof material and hence the broad spectral variability for the further differentiation of residential building types using Neural Networks resulted in comparatively inaccurate results. Further research is to be conducted in this context, especially in regard to the inclusion of height information for example. The lower classification accuracy using only the derived footprints for the RF approach from the Deep Learning approach can be traced back to the fact that differentiation based only on morphological aspects is quite difficult. The inclusion of height information in this case could increase the accuracy considerably (cf. WURM et al. 2016). Hence, this aspect needs to be further researched, especially since in some cases spectral data is cheaper to acquire than elevation data, making it a valuable data source for semantic labelling of building footprints for areas where no official footprint data is available.

5 References

BELGIU, M., TOMLJENOVIC, I., LAMPOLTSHAMMER, T.J., BLASCHKE, T. & HÖFLE, B., 2014: Ontology-based Classification of Building Types Detected from Airborne Laser Scanning Data. Remote Sensing **6**, 1347-1366.

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

HECHT, R., MEINEL, G. & BUCHROITHNER, M., 2015: Automatic Identification of Building Types Based on Topographic Databases – a Comparison of Different Data Sources. International Journal of Cartography **1**(1), 18-31.

INTERNATIONAL ENERGY AGENCY AND THE UNITED NATIONS ENVIRONMENT PROGRAMME, 2018: 2018 Global Status Report: Towards a Zero-Emission, Efficient and Resilient Buildings and Construction Sector.

LIAW, A. & WIENER, M., 2002: Classification and Regression by randomForest. R News **2**(3), 18-22.

LONG, J., SHELHAMER, E. & DARRELL, T., 2015: Fully Convolutional Networks for Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431-3440.

MEINEL, G., HECHT, R. & HEROLD, H., 2009: Analyzing Building Stock Using Topographic Maps and GIS. Building Research and Information **37**(5-6), 468-482.

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. ISPRS Journal of Photogrammetry and Remote Sensing **67**(1), 93-104.

SIMONYAN, K. & ZISSERMAN, A., 2014: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

STEEMERS, K. & YUN, G.Y., 2009: Household Energy Consumption: A Study of the Role Occupants. Building Research & Information **37**(5-6), 625-637.

WURM, M., SCHMITT, A. & TAUBENBÖCK, H., 2016: Building Types' Classification Using Shape-Based Features and Linear Discriminant Functions. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **9**(5), 1901-1912.

WURM, M., STARK, T., ZHU, X.X., WEIGAND, M. & TAUBENBÖCK, H., 2019: Semantic Segmentation of Slums in Satellite Images Using Transfer Learning on Fully Convolutional Networks. ISPRS Journal of Photogrammetry and Remote Sensing **150**, 59-69.

ZHANG, L., ZHANG, L. & DU, B., 2016: Deep Learning for Remote Sensing Data. A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine **4**(2), 22-40.

ZHU, X.X., TUIA, D., MOU, L., XIA, G.S., ZHANG, L., XU, F. & FRAUNDORFER, F., 2017: Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine **5**(4), 8-36.