MULTI-TASK ACTIVE LEARNING FOR CHARACTERIZATION OF BUILT ENVIRONMENTS WITH MULTI-SENSOR EARTH OBSERVATION DATA

Christian Geiß[1], *Member*, *IEEE*, Matthias Thoma[2], Massimiliano Pittore[3], Marc Wieland[4], Stefan Dech[1], *Member*, *IEEE*, and Hannes Taubenböck[1]

[1] German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), 82234 Wessling-Oberpfaffenhofen, Germany; christian.geiss@dlr.de, stefan.dech@dlr.de, hannes.taubenboeck@dlr.de
[2] University of Applied Sciences Weihenstephan-Triesdorf, 85354 Freising, Germany; matthias.thoma@hswt.de
[3] GFZ German Research Centre for Geosciences, 14467 Potsdam, Germany; pittore@gfz-potsdam.de
[4] University of Oxford, OX1 3PS Oxford, United Kingdom; marc.wieland@zoo.ox.ac.uk

Abstract— In this paper, we propose a multi-task active learning (AL) framework for an efficient characterization of buildings using features from multi-sensor earth observation data. Conventional AL methods establish query functions based on a preliminary trained learning machine to guide the selection of additional prior knowledge (i.e., labeled samples) for model improvement with respect to a single target variable. In contrast to that, here, we follow three multi-task AL meta-protocols to select unlabeled samples from a learning set which can be considered relevant with respect to multiple target variables. In particular, multi-task AL methods based on multi-variable criterion, alternating selection, rank combination, as well as hybrid approaches, which internalize multiple principles from the different meta-protocols, are introduced. Thereby, the alternating selection strategies implement a so-called one-sided selection (i.e., single-task AL selection for a reference target variable with simultaneous labeling of the residual target variables) with a changing leading variable in an iterative selection process. The multi-variable criterion-based methods and rank combination approaches aim to select unlabeled samples based on combined single-task selection decisions. Experimental results are obtained from two application scenarios for the city of Cologne, Germany. Thereby, the target variables to be predicted comprise building material type, building occupancy, urban typology, building type, and roof type. Comparative model accuracy evaluations underline the capability of the introduced methods to provide superior solutions with respect to one-sided selection and random sampling strategies.

Index Terms— Multi-task Active Learning, Support Vector Machines, Very High Resolution Imagery, LiDAR, Building Material Type, Building Occupancy, Urban Typology, Building Type, Roof Type

I. INTRODUCTION

A large diversity of methods were developed, which turn Earth observation (EO) data into land cover information allowing numerous applications related to characterization and assessment of urban environments. Those applications comprise for example identification of settlement areas for the quantification of urban expansion [1], [2], or the characterization of urban morphology [3], [4]; however, they also comprise energy-related analysis [5] [6], and assessments of natural hazard vulnerability and risk [7]-[9], among others.

Especially data with a very high spatial resolution (VHR) from optical sensors such as WorldView-I–III or GeoEye, and digital surface models (DSM) from LiDAR allow for a detailed characterization of objects of the built environment. In this manner, recent studies focused on the assignment of various categorical labels to buildings using EO: Geiß *et al.* [10], [11] combine spaceborne multispectral imagery and DSM data from airborne radar measurements acquired by pair antennas. They estimate seismic vulnerability levels of buildings (using a scoring method and categorization according to the European Macroseismic Scale 1998 [12]), and aim at the mapping of the buildings' structural types including construction materials and type of lateral-load-resisting system (e.g., 'confined masonry', or 'reinforced concrete') for earthquake loss estimation. Belgiu *et al.* [13] use airborne LiDAR for discrimination of buildings types according to the categories 'apartment', 'residential', and 'industrial'. Du *et al.* [14] distinguish different apartment types, as well as residential, administrative, commercial, and industrial buildings, among others, based on features derived from VHR imagery. Wurm *et al.* [15] fuse elevation data from photogrammetric DSMs, which were derived from airborne stereoscopic images and real estate cadaster for discrimination of semi/detached houses, block development, perimeter block development, terraced houses, and halls.

From a methodological point of view, the aforementioned studies all relied on principles of empirical inference. That is, deriving a rule (e.g., a decision function) from limited but properly encoded prior knowledge (i.e., labeled samples) to assign a label to the residual instances under analyses (i.e., unlabeled samples). Thereby, algorithmic modeling approaches [16] such as Random Forest (RF) [17] and Support Vector Machines (SVM) [18], which are able to cope with high dimensional feature vectors in a beneficial way for achieving a high predictive accuracy, were dominantly employed. Such models learn from empirical data and are useful when an explicit modeling based on, e.g., mechanical models is too complex. At the same time, such approaches need a sufficient amount of prior knowledge in order to

achieve a suitable generalization capability on unseen data [19], [20]. However, compilation of such prior knowledge in terms of a sufficient training set by, for example, labor-intensive *in situ* ground surveys is often costly to obtain.

Recently, active learning (AL) strategies [21] were introduced to the remote sensing community [22]-[24] to alleviate the burden associated with the compilation of a training set. Thereby, the aim of AL methods is to compile effective training sets with few relevant labeled samples for subsequent model learning. Thus, from an application-oriented point of view, AL strategies can be considered in particular relevant if the process of training set compilation is very costly. In order to do so, AL methods deploy predefined heuristics to rank unlabeled instances in the domain under analysis that can be considered the most valuable for improvement of estimation accuracy of a preliminary trained learning machine. Then, those unlabeled samples are labeled in a prioritized manner by an expert (i.e., oracle). Latest approaches also include the spatial domain for this task [25]-[27], and consider labeling costs arising from ground surveys [28], [29].

In parallel, modern *in situ* capture tools were developed, which enable a systematic and comprehensive collection of data. For instance, mobile mapping systems such as omnidirectional cameras mounted on vehicles allow for automated compilation of georeferenced omnidirectional imagery according to predefined routing constrains [30], [31]. This kind of imagery enables measuring important building features such as building height, number of floors, or vertical structural irregularities. Moreover, the analysis of the captured image sequences can be performed manually subsequent to data collection through visual interpretation by, for instance, structural engineers. In this manner, experts can assign a set of tags (i.e., categorical labels) to the analyzed building specifying individual structural features which cannot be detected automatically, or they can label the building as a whole according to a finite set of typologies which are relevant for seismic or flood hazard-related analysis, among others. This kind of screening procedure is called remote rapid visual screening (RRVS) and has the advantage that the operators can inspect a large number of structures in short time, also exploiting available ancillary information and providing an efficient first-level assessment of expected vulnerability. Moreover, decoupling data acquisition and data analysis allows for decentralized and location-independent analysis (i.e., also allowing a crowd-sourcing approach to image interpretation).

In this paper, we uniquely deploy AL strategies to estimate building characteristics based on features from multi-sensor earth observation data, where labeled training and testing samples are compiled from a RRVS procedure. More

significantly, we render the AL problem as a multi-task optimization procedure. Thereby, a model is not optimized for a single target variable – as in previous studies – but aims for learning an efficient solution for multiple target variables simultaneously. In our application context, categorical labels for multiple target variables must be assigned to a building simultaneously in order to characterize different building properties in a RRVS AL task in an efficient way. Such a multi-task AL strategy can be considered in particular relevant if more than one target variable must be labeled and the labeling process itself is very costly. To the best of our knowledge, such a setting was not considered in the field of remote sensing previously. This can be related to the fact that in most common remote sensing applications such as land use/land cover (LULC) mapping only one target variable must be estimated.

From a methodological point of view, we built upon a batch-mode AL strategy with uncertainty and diversity criterion. Subsequent to that, we implement different multi-task AL procedures based on *alternating selection* of the target variables within an iterative selection procedure, as well as *multi-variable criterion*-based methods and *rank combination* approaches to select the most valuable samples with respect to all target variables. Selected samples are subsequently deployed to learn independent models for predicting each target variable. We evaluate the relevance of the multi-task AL strategies in the context of two challenging application scenarios for which RRVS data are available. There, we rely on VHR multispectral imagery and DSM data for characterization of buildings. Overall, the categorical labels refer to target variables which describe *material type*, *occupancy*, *urban typology*, *building type*, and *roof type*.

The remainder of the paper is organized as follows. Section II documents the deployed and developed methods in terms of feature calculation and selection, as well as single and multi-task AL strategies. Section III is used to describe study area, data sets and parameterization of methods. Results of the actual experiments are reported in section IV. We give concluding remarks and draw future perspectives in section V.

II. METHODS

In the following we calculate and select features from multi-sensor earth observation data for characterization of buildings (Sec. IIA). Subsequently, we follow a batch-mode AL method by assessing unlabeled samples based on uncertainty and diversity (Sec. IIB). Based on this method, we establish a multi-task AL framework by following different meta-protocols to query those unlabeled samples which are most informative for multiple target variables simultaneously (Sec. IIC).

A. Feature Calculation and Selection

a) *IB – individual building* b) *AB – aggregated building* c) *BB – building block*

Fig. 1. Three different spatial levels for feature calculation; Features are computed for a) individual building geometries, b) aggregated building geometries, and c) building blocks.

An exhaustive number of features is calculated on three spatial levels using building geometries from the OpenStreetMap (OSM) project, VHR multispectral imagery and height information from a normalized DSM (nDSM), for an exhaustive and discriminative characterization of buildings. The spatial levels comprise individual buildings (*IB*) (Fig. 1a), aggregated buildings (*AB*) (Fig. 1b), and building blocks (*BB*) (Fig. 1c). The footprints of individual buildings represent the actual mapping geometries, whereas aggregated building geometries, and building blocks are intended to describe the spatial context which the individual buildings are embedded in. Thereby, aggregated building geometries and building blocks can be interpreted as spatial features, which account for the neighborhoods of the elements under analysis. Consequently, an individual building inherits the feature values from the corresponding aggregated building geometry and building block. The building blocks were derived from an affiliated street network. This is intended to allow for reflecting the urban morphology in an appropriate spatial way [32], which is constituted by distinct areas that are generally irregularly shaped. Simultaneously, the difficulty of having to determine *a priori* e.g., an optimal window/kernel size for artificial spatial units, such as quadratic objects, is avoided [33], [11].

The complete list of calculated features is provided in Table I. The first group of features computed for both *IB* and *AB* is related to the two-dimensional extent. In addition, measures are calculated on *BB* level, which describe the extent of different LULC classes, among others. To this purpose, a LULC classification is derived from VHR imagery for surfaces not covered by the building geometries in a supervised manner using an SVM model on a multilevel object-based image analysis procedure [34]. Estimated model generalization capabilities suggest a reliable determination of relevant residual LULC classes (precision and recall rates beyond 85%), which comprise "impervious surface", "bare soil", "meadows", "bushes/trees", and "water". Besides, the shape of both *IB* and *AB* is approximated with measures that deploy a comparison with two-dimensional geometrical forms such as square, rectangle, or ellipse, among others [15], [35]-[39]. Additionally, measures such as inverted floor area ratio, which establishes a relation of floor area covered by a building

and estimated total area of all floors of a building [40], are considered.

Based on the nDSM data, a set of measures is computed, which describe the three-dimensional extent considering both maximum and median values for the spatial levels of analysis. Median values were found empirically to be less sensitive to height variations and provide consequently a more reliable height approximation compared to e.g., mean values [41]. Those features are complemented by descriptors of three-dimensional shape [36].

To account for spatial context and configuration of *IB* and *AB*, orientation and geographical distance measures are introduced, among others. Those are complemented by features describing spatial alignment and relative composition of LULC elements [40], [42].

Spectral features derived from VHR multispectral imagery serve primarily as descriptors of roof surface material and arrangement [43]. Mean and standard deviation values of the different image bands are calculated in addition to band ratios and index features, which are intended to emphasize spectral dissimilarities [10], [11], [44]. Finally, rotation-invariant texture measures based on the grey-level co-occurrence matrix (GLCM) [45] are derived, since it was shown that such features can provide supplementary information if the spectral resolution is limited and the ground sampling distance is much smaller than the objects of interest, what is frequently the case in VHR imagery. Thereby, we selected three measures from the set of 14 originally proposed GLCM measures, since some are strongly correlated with each other [46]. Dominant share of features were computed in the software environment eCognition [36] using already implemented or customized protocols.

TABLE I

LIST OF FEATURES FROM DIFFERENT CATEGORIES WITH NUMBER OF FEATURES ACCORDING TO SPATIAL SCALES (IB = INDIVIDUAL BUILDING, AB = AGGREGATED BUILDING, BB = BUILDING BLOCK).

Features	Number of features		
	IB	*AB*	*BB*

Category	Features			
Geometry – Extent 2D	Area, length, width, length / width, length of main line, width of main line, length of main line / width of main line, perimeter	8	8	
	Inverted floor area ratio	1		
	Area of BB, average building area based on IB within BB, sum of building areas within BB, sum of bushes/trees areas within BB, sum of meadow areas within BB, sum of impervious and bare soil areas within BB, average length / width based on IB within BB, average length / width based on AB within BB, average length of main line / width of main line based on IB within BB, average length of main line / width of main line based on AB within BB			10
Geometry – Shape 2D	Number of building vertices, shape index, fractal dimension, density, border index, compactness, rectangular fit, radius of largest enclosed ellipse, radius of smallest enclosing ellipse, elliptic fit, roundness, normalized perimeter index, normalized proximity index, normalized spin index, areal asymmetry	15	15	
Geometry – Extent 3D	Building height based on maximum height, building height based on median height	2		
	Building height based on maximum height within AB, area-weighted building height based on maximum heights of IB within AB, building height based on median height within AB, area-weighted building height based on median heights of IB within AB		4	
	Highest IB based on maximum height within BB, area-weighted building height based on maximum heights of IB within BB, median of building height based on median heights of IB within BB, area-weighted building height based on median heights of IB within BB			4

Category	Features			
	Building volume(s) based on maximum height(s), building volume(s) based on median height(s)	2	2	2
Geometry – Shape 3D	Shape index 3D based on maximum height, shape index 3D based on median height, asymmetry 3D based on maximum height, asymmetry 3D based on median height	4	4	
	Shape index 3D based on area-weighted maximum heights of IB, shape index 3D based on area-weighted median heights of IB, asymmetry 3D based on area-weighted maximum heights of IB, asymmetry 3D based on area-weighted median heights of IB			4
Spatial Context / Configuration	Orientation of building according to major axis of minimum bounding rectangle, distance to nearest building, distance to city center (Cologne Cathedral)	3	3	
	Standard deviation of orientations of IB, Standard deviation of orientations of AB, median of distances to nearest building based on IB, median of distances to nearest building based on AB, number of IB within BB, number of AB within BB, median of inverted floor area ratio based on IB, building aggregation measure, normalized building aggregation measure, vegetation fraction, impervious surface area, urban density, vegetation volume to built-up volume, urban vegetation index, site occupancy index			15

Spectral – 1st order	Mean(blue), mean(green), mean(red), mean(NIR), StDev(blue), StDev(green), StDev(red), StDev(NIR), mean(blue) / mean(green), mean(blue) / mean(red), mean(blue) / mean(NIR), mean(green) / mean(red), mean(green) / mean(NIR), mean(red) / mean(NIR), mean(green) / [mean(green)+ mean(red)+ mean(NIR)], mean(red) / [mean(green)+ mean(red)+ mean(NIR)], mean(NIR) / [mean(green)+ mean(red)+ mean(NIR)], brightness, normalized differenced vegetation index, soil-adjusted vegetation index	20	20	
Spectral – 2nd order	$GLCM_{inv.}$(Angular 2nd moment), $GLCM_{inv.}$(entropy), $GLCM_{inv.}$(homogeneity)	3	3	
$\sum$		58	63	31

Overall, each *IB* is represented by a 152-dimensional stacked feature vector before feature selection.

In general, it can be noted that high-dimensional feature vectors often exhibit redundancy, show inter-correlations between features and can induce the ''Hughes phenomenon'' (which states that for a limited amount of samples the predictive power decreases as the dimensionality of the feature vector increases) [47]. In addition to that, multivariate classification methods can be prone to over-fitting. Filtering out the least promising features and thus reducing the dimensionality of the feature vector can attenuate the aforementioned issues. We establish a feature selection protocol prior to and independent of the actual AL task to impose *ceteris paribus*-near conditions for comparison of the different AL methods later on (i.e., ensure that accuracy levels of individual AL methods are not negatively influenced by the dimensionality of feature vectors and comparability of different AL methods is not further affected by an intrinsic feature selection process).

To this purpose, we follow a multistep procedure for the selection of most informative features for the individual target variables using principles as introduced by the authors of [48]. *First*, all features are normalized to a 0-1 interval. Feature subsets are compiled based on a correlation matrix, which relies on Pearson product-moment correlation coefficient. Thereby highly correlated features are excluded by evaluating different correlation coefficient-related thresholds τ. *Subsequently*, for each of those sets, a ranking of the individual features with respect to a target variable is established. It is derived from the receiver operator characteristic curves of the features, as estimated from cross-validation based on a training set, when learning an SVM model from a feature set. The ranking of the features according to their relevance allows selecting the z most important ones. *Finally*, a set of features is determined for the respective target variables using a combination of τ and z, which allows for the highest predictive accuracy. Thereby, large numbers of features are penalized, and robust models, as quantified by a low variance of predictive accuracy according to different model runs, are favored. An exhaustive number of correlation coefficient-related thresholds $\tau = \{1, 0.99, 0.975, 0.95, 0.925, 0.9\}$ and numbers of features $z = \{20, 25, 30, 35, 40, 50, 70\}$ were established from empirical prior analysis and subsequently evaluated based on 50 independent trials. The final feature vectors comprise 25 dimensions for the variable *material type*, 40 dimensions for *occupancy*, and 20 dimensions for the variables *urban typology*, *building type*, as well as *roof type*.

B. Single-task Batch-Mode Active Learning with Uncertainty and Diversity Criterion

We consider a training set composed initially of n labeled samples $L = \{\mathbf{x}_i, y_i\}_{i=1}^{n}$, where $\mathbf{x}_i$ represents a vector of stacked features (in our case features from multiple earth observation data characterizing the building footprints) and y_i is a discrete label defined among Q classes (here different classes according to, for instance, a building's *material type*). We also consider a learning set composed of u unlabeled samples $U = \{\mathbf{x}_j\}_{j=n+1}^{n+u}$, in real-world scenarios frequently with $u \gg n$, in our case the residual buildings of the area under analysis B (i.e., $U = B - L$). To increase the training set L with samples chosen from the learning set U to be labeled manually by an expert, an active learning algorithm aims at selecting them properly in order to jointly minimize the number of unlabeled samples to be labeled (i.e., number of interactions with the expert), and maximize the accuracy of the subsequent classification model.

We build upon a batch-mode active learning technique (i.e., selecting more than one sample during each iteration) [21], which aims to internalize the prioritized selection of both uncertain and diverse unlabeled samples (Fig. 2). Therefore, a batch $X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_h\}$ of h samples is selected by evaluating multiclass-level uncertainty (MCLU) [49] and angle-based diversity (ABD) [50] within an SVM framework.

MCLU is used to evaluate the uncertainty of samples. In contrast to other heuristics such as Query by Committee or Query by Bagging [24], which also assess uncertainty of samples, MCLU requires solely a single model for evaluation. Therefore, a confidence value $c(\mathbf{x})$ for each unlabeled sample $\mathbf{x} \in U$ is computed from its functional distance $f_i(\mathbf{x}), i = 1, 2, ..., d$, to d decision functions of binary SVM in a one-against-one multiclass architecture (however, it can also be computed in a one-against-all architecture). For computation of $c(\mathbf{x})$ a difference function is used, which incorporates the uncertainty between the two most likely classes based on the

first and second largest distance to the SVM hyperplane:

$$r_{1max} = \underset{i=1,2,...,d}{\arg\max} \{f_i(\mathbf{x})\}$$
$$r_{2max} = \underset{j=1,2,...,d, j \neq r_{1max}}{\arg\max} \{f_j(\mathbf{x})\}$$
$$c(\mathbf{x}) = f_{r_{1max}}(\mathbf{x}) - f_{r_{2max}}(\mathbf{x}). \tag{1}$$

The sample in the model which minimizes $c(\mathbf{x})$ is the sample closest to the decision boundary (i.e., the most uncertain sample) between class r_{1max} and class r_{2max} (i.e., the most probable classes). However, especially in batch-based AL procedures, a single selection heuristic can lead to the prioritized selection of outliers and redundant samples, and searches in noisy or irrelevant parts of the feature space [21].

Consequently, a second selection criterion is integrated in addition. ABD is based on the cosine angle distance of hyperplanes which are induced by the samples. This technique is rooted in theoretical considerations on optimal version space separation [50]. Based on the kernel function $K(.,.)$ used to describe the inner product in the considered feature space, the cosinus of the (undirected) angles between two hyperplanes H_i and H_j, which correspond to the unlabeled samples $\mathbf{x}_i$ and $\mathbf{x}_j$ (with normal vectors $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$), is calculated as:
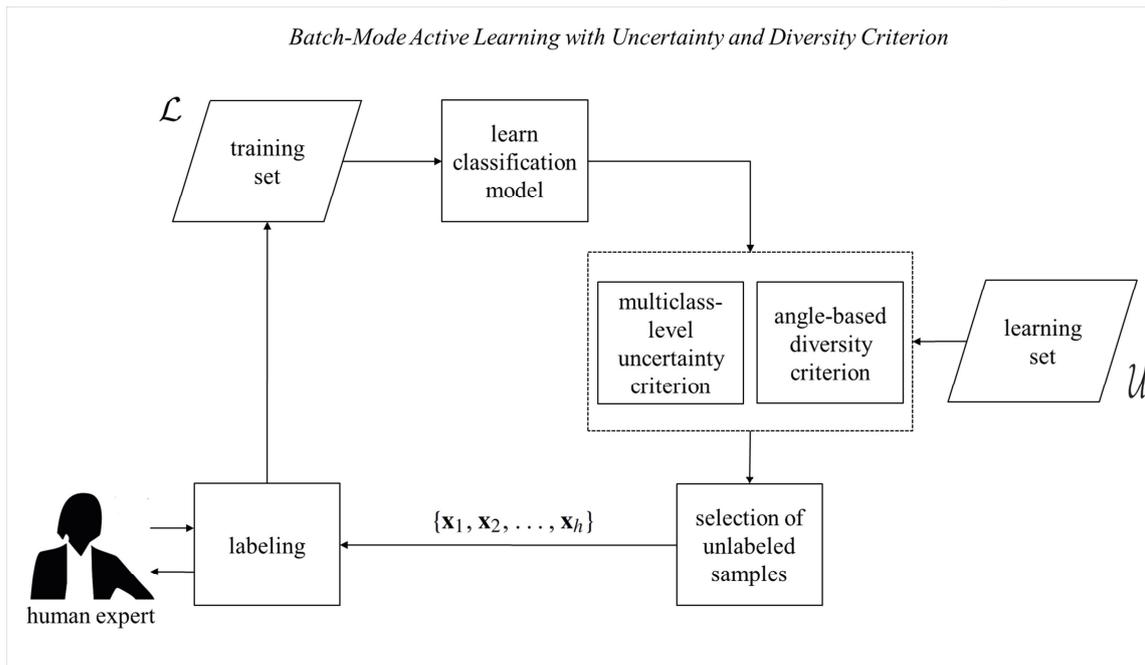
$$\left|\cos(\angle(H_i, H_j))\right| = \frac{|\langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle|}{\|\phi(\mathbf{x}_i)\| \|\phi(\mathbf{x}_j)\|}$$
$$= \frac{K(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{K(\mathbf{x}_i, \mathbf{x}_i)K(\mathbf{x}_j, \mathbf{x}_j)}} \tag{2}$$

and the angle between H_i and H_j:

$$\angle(H_i, H_j) = \cos^{-1}\left(\frac{K(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{K(\mathbf{x}_i, \mathbf{x}_i)K(\mathbf{x}_j, \mathbf{x}_j)}}\right). \tag{3}$$

The angle of two unlabeled samples is small if they are similar and large if they are differing.

For a combined selection of samples according to the

Batch-Mode Active Learning with Uncertainty and Diversity Criterion

Fig. 2. Flowchart of a single iteration of the single-task active learning method adopted.

presented heuristics, we first select the most uncertain m samples from U. Thereby, the sample with lowest $c(\mathbf{x})$ value is added initially to the batch. The subsequent sample $\mathbf{x}_t$ of the batch is selected from the remaining m candidates based on a combined metric:

$$\mathbf{x}_t = \arg\min_{i=1,\dots,m} \left\{ \lambda |c(\mathbf{x}_i)| + (1-\lambda) \cdot \left[\max_{x_j \in X} \frac{K(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{K(\mathbf{x}_i, \mathbf{x}_i) K(\mathbf{x}_j, \mathbf{x}_j)}} \right] \right\} \quad (4)$$

where λ is a weighting parameter to allow for a tradeoff between uncertainty and diversity. The term multiclass-level uncertainty - angle-based diversity (**MCLU-ABD**) is used in the remainder of the paper when referring to this AL technique.

C. Multi-task Batch-Mode Active Learning

In contrast to single-task AL, multi-task strategies aim to query those unlabeled samples which are most informative for all involved target variables. A naïve strategy to expand a single-task scenario to a multi-task setting is one-sided selection (i.e., perform single-task AL selection for a reference target variable and label the residual target variables simultaneously) [51]. Such a strategy internalizes an intrinsic selection for the reference variable (i.e., leading variable) and extrinsic selection for the residual variables. Consequently, a training set which was compiled based on one-sided selection is likely to have a good performance on the reference variable but is not guaranteed to be superior compared to random selection with respect to the residual variables, especially if the different annotation tasks are highly diverse.

In contrast to that, three meta-protocols are followed in this paper to specifically address the minimization of labeling costs with respect to all target variables simultaneously, namely selection of unlabeled samples based on *multi-variable criterion*, *alternating selection* and *rank combination*.

1) Multi-variable criterion: The governing strategy of multi-variable criterion multi-task AL refers to the computation of a rationally scaled metric involving multiple variables. Here, uncertainty is computed for each variable first and the individual uncertainties are combined based on a i) minimum function or ii) average function (it can be noted that the minimum numeric value of the MCLU criterion corresponds to maximum uncertainty, and the average function reveals the same results as a summation function, if every variable is weighted equally) (Fig. 3):

i) Compute $c(\mathbf{x})$ for each variable and combine the numeric values with a minimum function; add the most uncertain sample to X; select m candidates with maximum absolute uncertainty for computation of the ABD criterion based on all features; add the most uncertain and diverse samples from m candidates to the residual positions of X (**MCLU-ABD *MVC-max***);

ii) Compute $c(\mathbf{x})$ for each variable and combine the numeric values with an average function; add the most uncertain sample to X; select m candidates with maximum mean uncertainty for computation of the ABD criterion based on all features; add the most uncertain and diverse samples from m candidates to the residual positions of X (**MCLU-ABD *MVC-mean***).
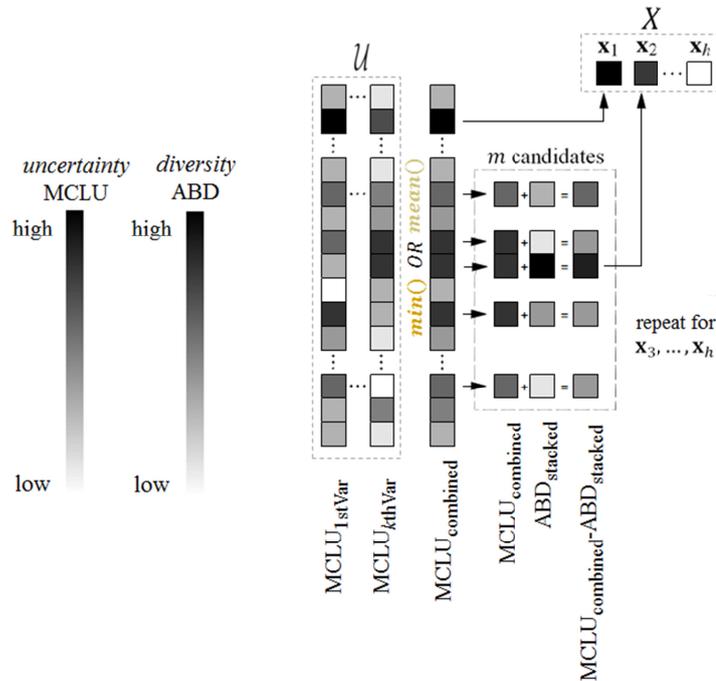
Fig. 3. Scheme of multi-variable criterion multi-task AL for compilation of the batch.

2) Alternating selection: The idea of alternating selection is to implement one-sided selection with a changing leading variable in an iterative selection process, to establish a straight-forward tradeoff between different optimized single-task selection decisions [51]. Within one model iteration, a batch of uncertain and diverse unlabeled samples is labeled. As such, samples are selected and labeled intrinsically for the leading variable and extrinsically for the residual variables. The query function changes the leading variable depending on the learning task at hand. As such, the process iterates through an ordered list of variables. It is possible to put more weight on individual variables by allowing them to lead the selection more frequently (i.e., appear more often on the ordered list) [21]. However, here, we establish equally weighted selections what corresponds to a round-robin strategy. In the experiments we consider two variants of this MCLU-ABD *one-sided alternating selection – equally weighted* technique (Fig. 4a): the first variant deploys only the features of the leading variable for computing hyperplanes in kernel space regarding the ABD criterion (**MCLU-ABD *AS***), whereas the second variant relies on a *stacked feature vector* based on all variables for evaluation of the *ABD criterion* (**MCLU-ABD *AS-stacked***).

A hybrid approach with two variants, which combines the constituting principles of alternating selection and multi-variable criterion multi-task AL is visualized in Fig. 4b:

i) Compute $c(\mathbf{x})$ for leading variable, select the sample with maximum absolute uncertainty, and add it to X; compute $c(\mathbf{x})$ for each variable and combine the numeric values with a minimum function; select m candidates with maximum absolute uncertainty for computation of the ABD criterion based on all features; add the most uncertain and diverse samples from m candidates to the residual positions of X (**MCLU-ABD *AS-MVC-max***);

ii) Compute $c(\mathbf{x})$ for leading variable, select the sample with maximum absolute uncertainty, and add it to X; compute $c(\mathbf{x})$ for each variable and combine the numeric values with an average function; select m candidates with maximum mean uncertainty for computation of the ABD criterion based on all features; add the most uncertain and diverse samples from m candidates to the residual positions of X (**MCLU-ABD *AS-MVC-mean***).

3) Rank combination: The idea of rank combination is to select samples for prioritized labeling based on combined single-task selection decisions. To this purpose, an ordinally scaled rank metric is computed for each unlabeled sample with respect to each considered learning task [51]. Here, we establish a pure and a hybrid variant of this general idea (Fig. 5). This is done to establish a strategy which selects informative samples with respect to *all* learning algorithms, and a strategy which internalizes also aspects of a selection of

8
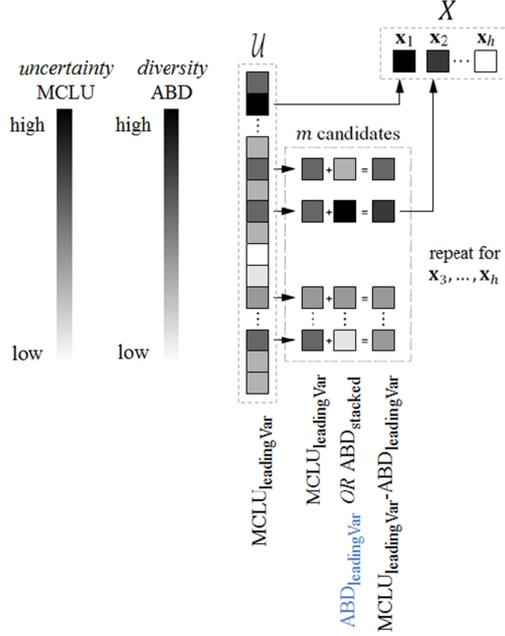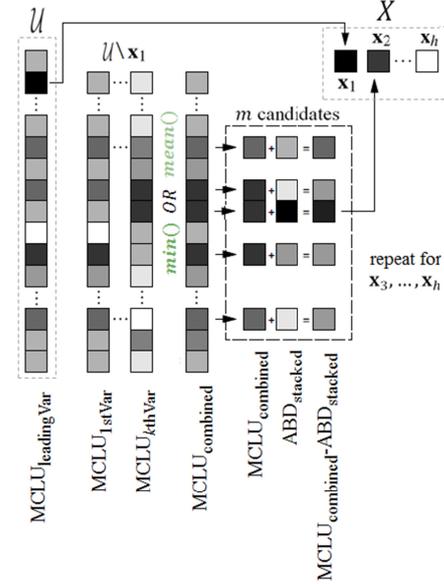
Fig. 4. a) Scheme of alternating selection-based multi-task AL for compilation of the batch; b) Scheme of hybrid alternating selection-based and multi-variable criterion multi-task AL strategy for compilation of the batch.

samples which are highly informative for one task but possibly less informative for another task, respectively.
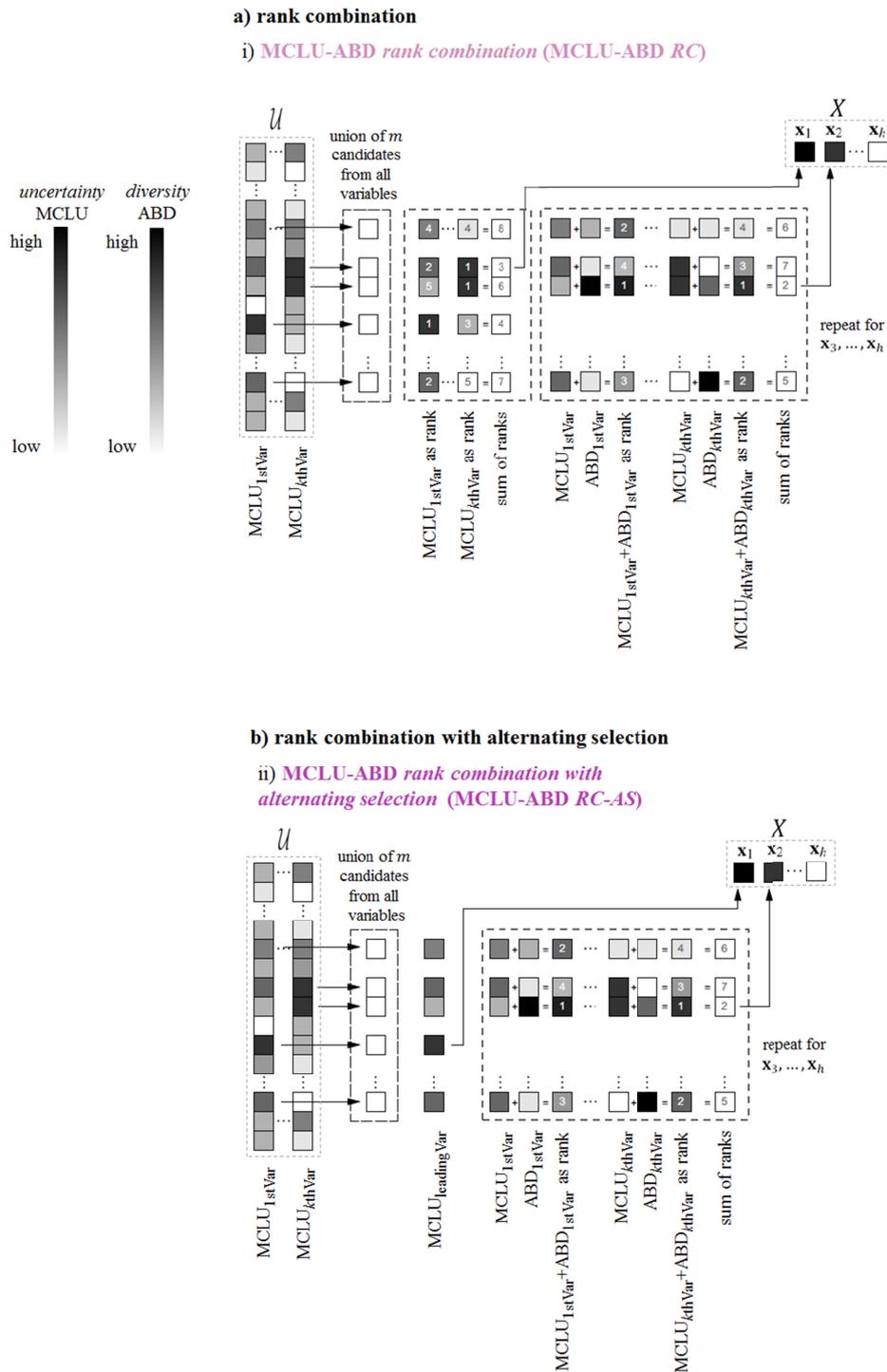
i) Compute $c(\mathbf{x})$ for each variable and select a pool of m candidates with maximum absolute uncertainty with respect to each variable; sum the ranking positions of the samples in the pool of m candidates according to absolute uncertainty of each variable and add the sample with highest rank to X; sum ranking positions of the samples in the pool of m candidates according to absolute uncertainty and diversity of each variable and add the samples with highest ranks to the residual positions of X (**MCLU-ABD *RC***).

ii) Compute $c(\mathbf{x})$ for each variable and select a pool of m candidates with maximum absolute uncertainty with respect to each variable; select the sample with maximum absolute uncertainty according to the leading variable and add it to the batch; sum ranking positions of the samples in the pool of m candidates according to absolute uncertainty and diversity of each variable and add the samples with highest ranks to the residual positions of X (**MCLU-ABD *RC-AS***).

## III. EXPERIMENTAL SETUP

A. Study Area and Data Sets

Two experiments were carried out with data from the city of Cologne, Germany. The city has approximately 1.05 million inhabitants, who live in 136 000 residential buildings [52]. The city is threatened slightly by earthquakes and heavily by floods [53] - a situation which triggers the need for up-to-date information on relevant building characteristics. Building geometries were compiled from cadastral data sources, which were also made publicly available via OSM. Building blocks for further feature calculation were generated from the main street network of the OSM data set for Cologne. VHR multispectral imagery was acquired by the WorldView-II sensor. The pansharpened imagery features a spatial resolution of 50 cm. The employed 4-band multispectral (blue, green, red, NIR) imagery was acquired on January 31, 2014. In addition, elevation information is extracted from LiDAR data with a spatial resolution of one meter.

a) rank combination

i) **MCLU-ABD** *rank combination* **(MCLU-ABD *RC*)**

b) rank combination with alternating selection

ii) **MCLU-ABD** *rank combination with alternating selection* **(MCLU-ABD *RC-AS*)**

Fig. 5. a) Scheme of rank combination-based multi-task AL for compilation of the batch; b) Scheme of hybrid rank combination-based and alternating selection multi-task AL strategy for compilation of the batch.

1) For the first RRVS experiment, we integrate imagery which was collected with an omnidirectional imaging system (Fig. 6a). The maximum survey area was bound to the inner-city districts of Cologne due to time- and cost-restraints for the field campaign. However, the sampled area covers all relevant so-called urban structure types (i.e., distinctive and homogeneous assemblages of LULC elements) as determined by prior analysis [54]. Routing of the system was based on available topographically corrected OSM data following the procedure described in Pittore *et al.* [31] (Fig. 6b). Collected data are fed to a RRVS system (Fig. 6c), where they are presented to the operator in a structured way. In addition, labeling constraints can be implemented according to predefined taxonomic specifications. Here, we implement

10
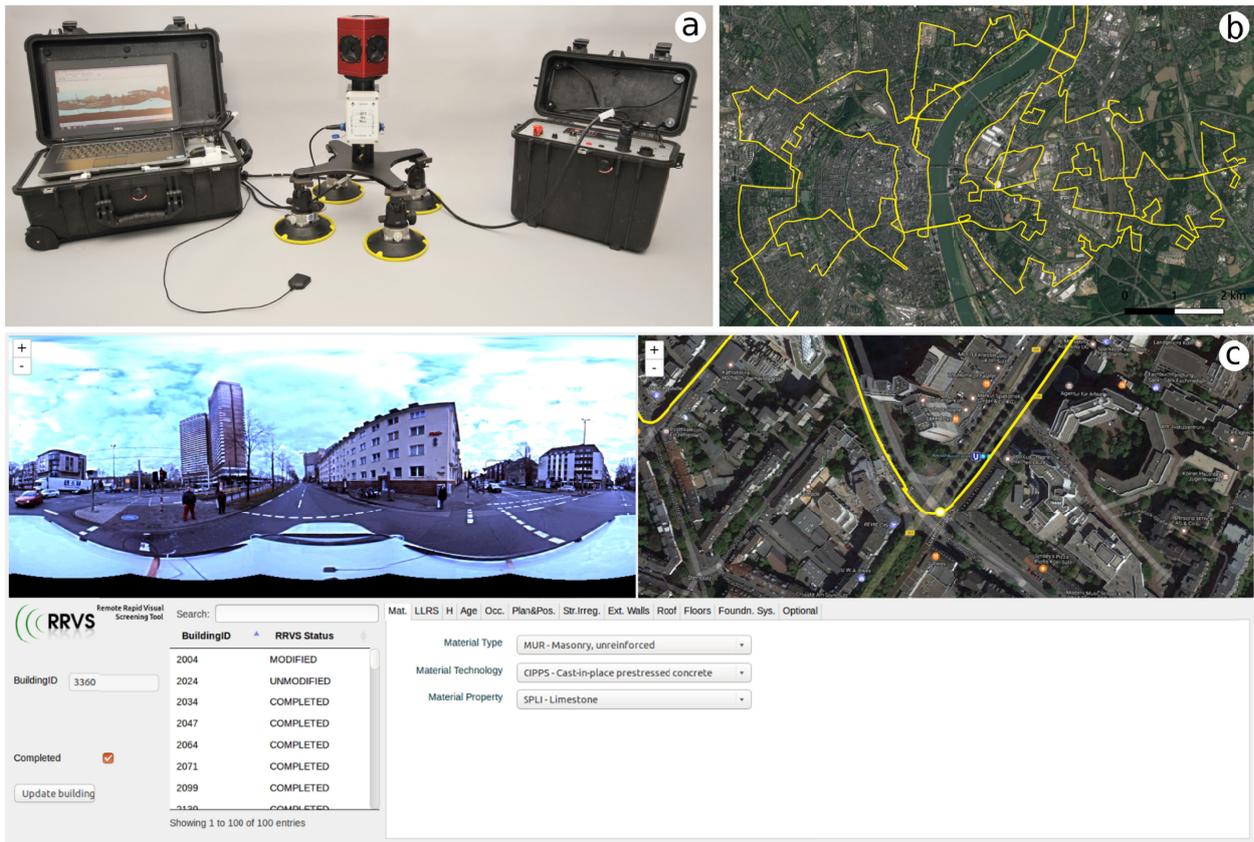
Fig. 6. a) Mobile mapping system for rapid remote visual screening procedure. The system uses a Ladybug 3 omnidirectional camera from Point Grey Research Ltd and a custom data capturing and storage device with navigation unit; b) GPS camera path for the inner city of Cologne; c) Web-based Remote Rapid Visual Screening (RRVS) system with omnidirectional image viewer, map interface and data entry form.

target variables and affiliated classes which correspond to the version 2 of the structural taxonomy proposed by the Global Earthquake Model [55]. The target variables are *material type*, *occupancy*, and *urban typology*. As such, experiment I corresponds to a three-task annotation scenario. The variables with affiliated classes are documented in Table II. For the systematic numeric evaluation of the different methods, labeled samples were split into disjoint training and test set in a stratified manner. The first is used for model learning and optimization of hyperparameters using cross-validation, whereas the latter is employed to estimate the generalization capabilities of models on completely unseen data. Thereby, it can be noted that not all unlabeled samples can be populated completely by the operator for the variables *material type* and *occupancy*. As such, all samples in the test set provide labels for all three variables. In contrast, the training set contains buildings where labels for only one, two or all three variables are available. While this is a realistic scenario, we additionally show learning curves where these samples are excluded for a specific target variable when they could not be labeled in regard to this variable – this enhances the comparability of the different methods, especially to the single-task learning methods.

TABLE II

TARGET VARIABLES AND POPULATION OF THE CLASSES FOR TRAINING/LEARNING AND DISJOINT TEST SET OF EXPERIMENT I

material type	train	test	Σ
concrete	98	32	130
masonry	323	107	430
Σ	**421**	**139**	**560**

occupancy	train	test	Σ
residential	280	95	375
commercial and public	19	6	25
mixed use	46	17	63
other occupancy type	60	21	81
Σ	**405**	**139**	**544**

urban typology	train	test	Σ

semi/detached houses	156	31	187
terraced houses	168	39	207
block development	163	48	211
mixed development	178	21	199
Σ	**665**	**139**	**804**

2) For the second experiment, we simulate an RRVS procedure by integrating a data set which carries building characteristics for the complete building inventory of the city of Cologne. The data set is based on cadastral sources and was specifically compiled in the context of a flood-related decision support system for the city of Cologne. From that data set we select the target variables *building type*, and *roof type*. As such, experiment II corresponds to a two-task annotation scenario. The variables with affiliated classes and numbers of labeled samples are documented in Table III. Concordant to the setup of experiment I, labeled samples were split into disjoint training and test sets. In addition, in this case not all buildings are populated completely with respect to both variables. Consequently, models are also evaluated in two ways, whereby partially unlabeled samples are either included or excluded when drawing the learning curves. From the complete pool of labeled training samples, a maximum of 800 samples is drawn for model learning and the complete pool of labeled test samples is used for accuracy evaluation.

TABLE III

TARGET VARIABLES AND POPULATION OF THE CLASSES FOR TRAINING/LEARNING AND DISJOINT TEST SET OF EXPERIMENT II

| **building type**	*train*	*test*	Σ
detached house	12 549	3 463	16 012
detached building block	5 166	1 513	6 679
perimeter block development	10 275	2 871	13 146
terraced house	37 958	11 228	49 186
garage	39 646	4 132	43 778
Σ	**105 594**	**23 207**	**128 801**
roof type	*train*	*test*	Σ
flat roof	49 938	6 005	55 943
single pitch roof	9 100	1 325	10 425
gable roof	75 204	15 877	91 081
Σ	**134 242**	**23 207**	**157 449**

B. *Experimental Setup*

For a feasible tradeoff between computational efforts and model accuracy a batch size of $h = 5$ for all AL strategies was adopted. The weighting parameter λ for MCLU-ABD was set to 0.6. This value showed favorable performance properties in previous studies [49]. For the SVM, we deployed Gaussian RBF kernels, which take the form $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\|\mathbf{x} - \mathbf{x}_i\|^2 / 2\sigma^2)$. Learning the most appropriate SVM in conjunction with an RBF kernel requires the definition of
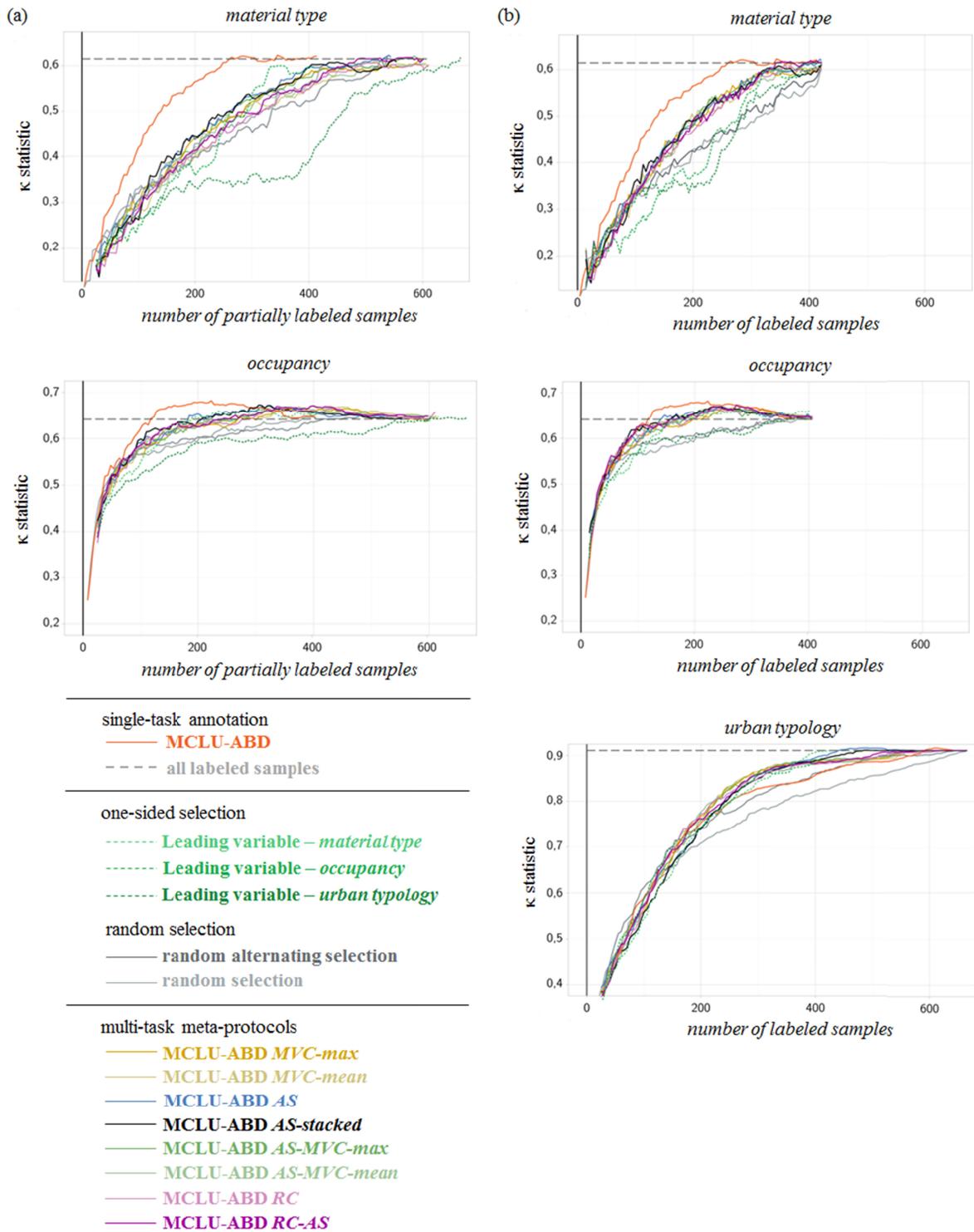
Fig. 7. Results for single-task AL, all labeled samples, extrinsic labeling strategies, and different multi-task AL methods; a) accuracy estimations where partially unlabeled samples are included in the learning process, and b) accuracy estimations where partially unlabeled samples are excluded from the learning process with respect to variables *material type* and *occupancy*.

the cost parameter C and the kernel width parameter γ. Tuning of C and γ was addressed by a grid search strategy based on fivefold cross-validation. Thereby, model accuracies are evaluated in terms of estimated κ on the average of three

independent trials, where an exhaustive optimization of hyperparameters is carried out $\Phi \in \{C, \gamma\}: C = \{2^{-5}, 2^{-3}, \ldots, 2^{15}\}, \gamma = \{2^{-15}, 2^{-13}, \ldots, 2^3\}$. At the beginning of an AL process only few labeled samples are
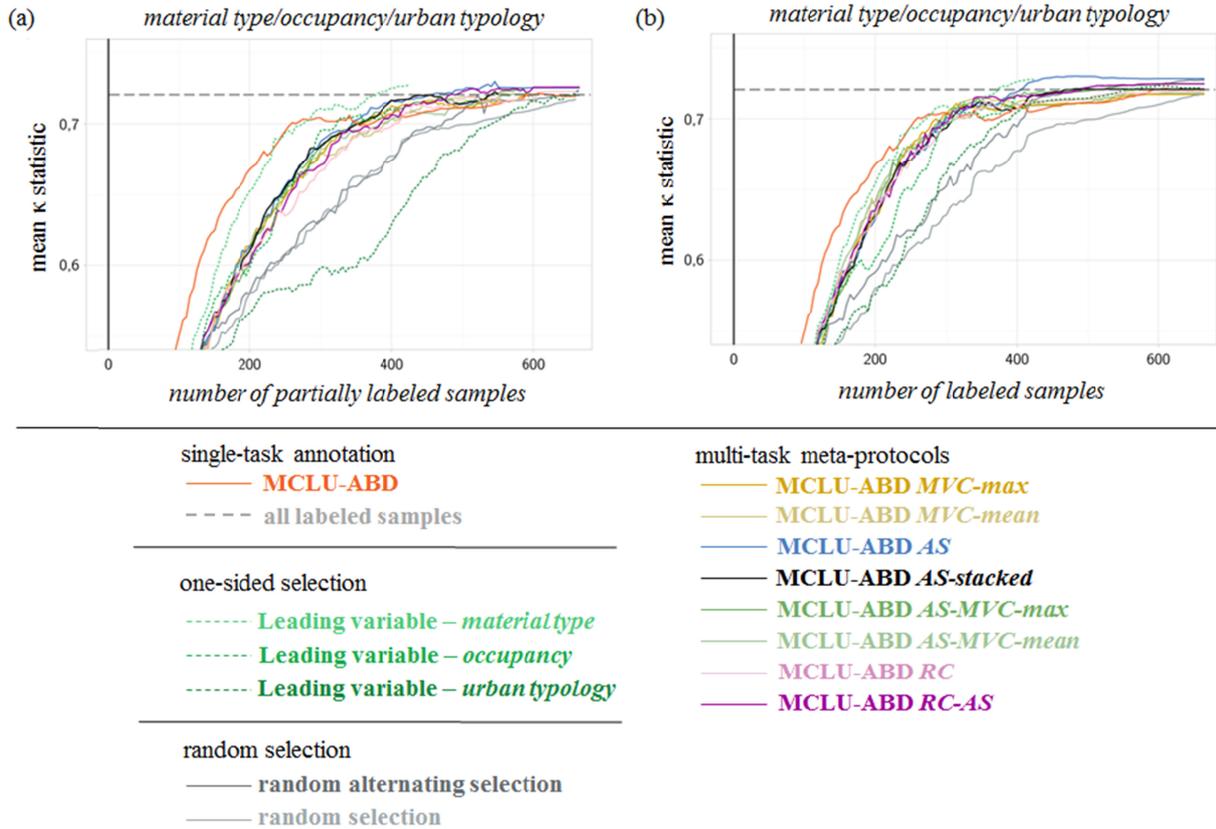
14

Fig. 8. Mean κ statistics of the individual results for single-task AL, all labeled samples, extrinsic labeling strategies, and different multi-task AL methods with respect to the three target variables; a) mean accuracy estimations where partially unlabeled samples are included in the learning process, and b) mean accuracy estimations where partially unlabeled samples are excluded from the learning process.

used for learning the model. However, in the subsequent process the number of labeled samples increases and thereby also the computational efforts for hyperparameter optimization. This can disturb the interactive process and slow down the actual labeling procedure, especially when multiple models are learned simultaneously as in this multi-task AL framework. At the same time, a training set of comprehensive size demands less adjustments since a stable model can already be established. Consequently, we implement a strategy which optimizes hyperparameters very frequently at the beginning of a learning process and less frequently with an increasing number of labeled samples in the model. Presented estimated generalization capabilities of learned models are computed from mean κ statistic values of 50 independent trials with initial training sets made up of two samples per class.

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

In general, a suitable multi-task protocol should enable faster achievement of high accuracies compared to random selection, and show accuracies as close as possible to single-task AL optimization. As a potentially stronger baseline compared to random selection, we consider one-sided sample selection. To this purpose, single-task AL selection for a reference target variable is followed with simultaneous labeling of the residual target variables.

As mentioned in the experimental setup section, not all unlabeled samples can be labeled completely for all variables

by the operator. This induces stretched accuracy functions compared to single-task AL models when evaluating model performance as a function of number of samples considered for learning, since more samples need to be queried to actually encode additional information in the model. Therefore, we provide experimental results for two settings, where the partially unlabeled samples are either included or excluded in the learning curves. That is also why we present two curves for random selection later on. One uses the limited pool of unlabeled samples that actually can be labeled for the regarded target variable (comparable to the single-task AL setting; termed "random selection"). The second curve is based the complete pool of unlabeled samples for all target variables in the same manner as the alternating selection methods (termed "random alternating selection"). Finally, we learned an SVM model for every variable based on a large number of available samples to indicate an upper bound of accuracy.

A. Experiment I: RRVS with Data from Omnidirectional Imaging System

Accuracy estimates in terms of κ statistic [56] of the different methods for the first RRVS experiment are visualized in Fig. 7. Fig. 7a shows results for variables where partially unlabeled samples were included, which is only relevant for the variables *material type* and *occupancy*. Fig. 7b provides additional results where unlabeled samples were excluded for those two variables. First it can be noted that general accuracy
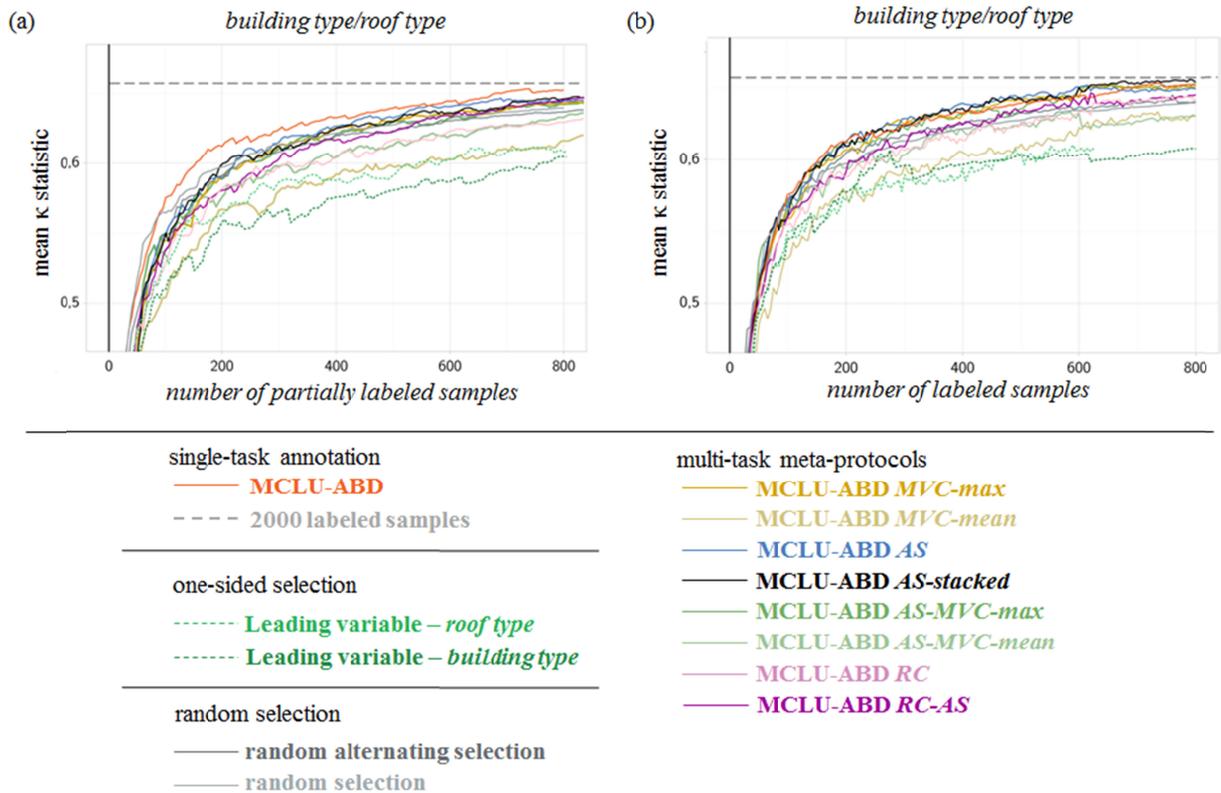
15
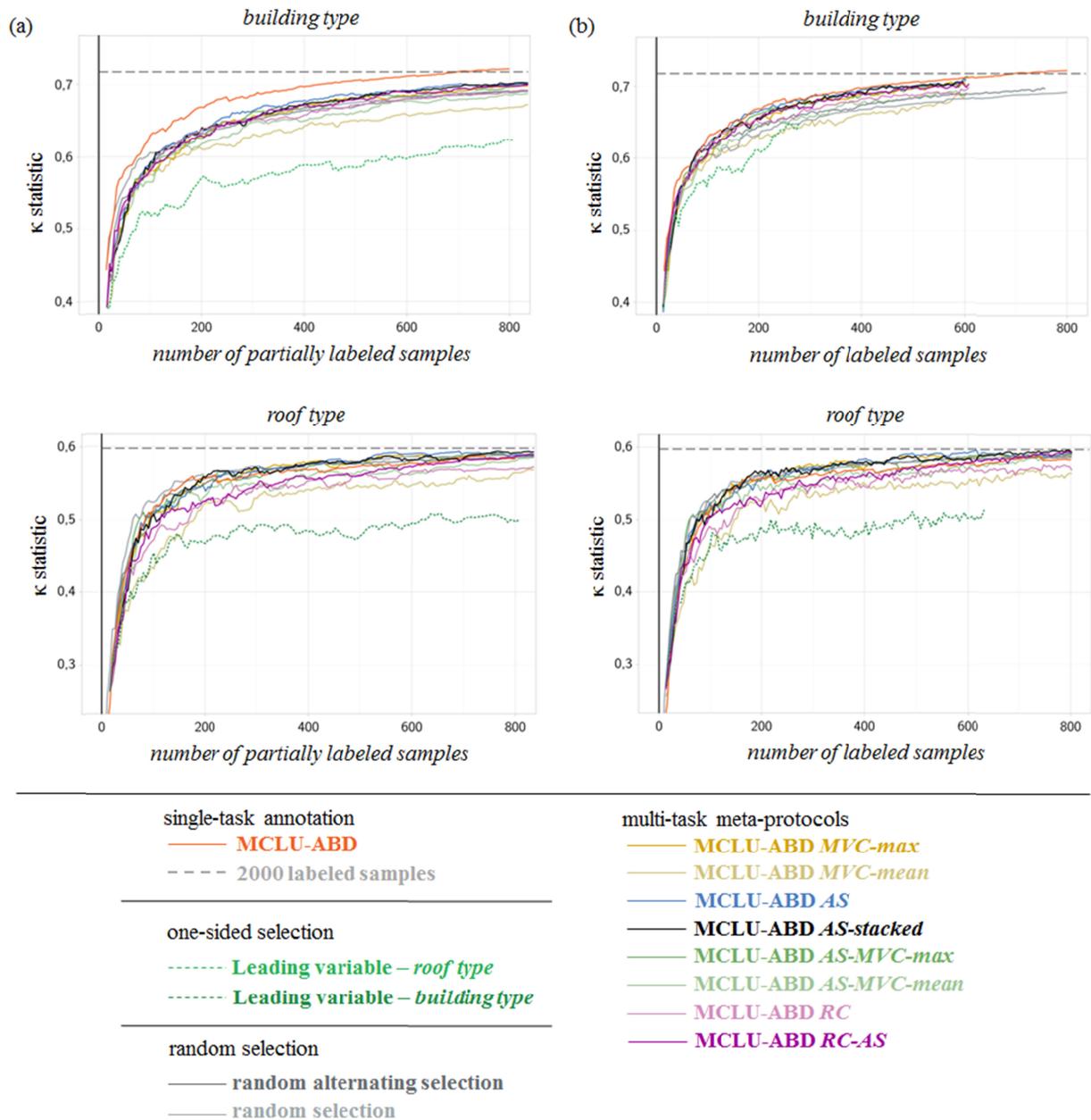
Fig. 9. Results for single-task AL, all labeled samples, extrinsic labeling strategies, and different multi-task AL methods; a) accuracy estimations where partially unlabeled samples are included in the learning process, and b) accuracy estimations where partially unlabeled samples are excluded from the learning process.

levels of the variables differ: the best models for the variables *material type* and *occupancy* show substantial agreements (i.e., $\kappa > 0.6$), whereas models for *urban typology* achieve excellent agreements (i.e., $\kappa > 0.8$). This can be related to the fact that the inference scheme for the latter variable can rather directly exploit the physical description of the urban morphology by remote sensing with respect to the fairly quantitative class properties. Instead, the inference schemes for the residual variables must assign more qualitative labels based on rather indirect relations.

Especially from the results with excluded partially unlabeled samples, it can be seen that the accuracy levels of random selections are consistently exceeded by the single-task AL method and more notably also by the multi-task AL approaches. As expected, MCLU-ABD allows most favorable accuracies for the majority of models for *material type* and *occupancy*. It can be observed that maximum accuracies are obtained with a reduced set of labeled samples, which gives an indication that some misleading samples are included in the pool. As desired, those are selected late for model learning.

There is hardly a clear pattern regarding the most favorable multi-task AL strategy in this setting. Methods which are based on alternating selection (especially MCLU-ABD *AS* and MCLU-ABD *AS-stacked*) show favorable accuracies for the variables *material type* and *occupancy*, however, they feature simultaneously hardly a superior behavior with respect

16

Fig. 10. Mean κ statistics of the individual results for single-task AL, all labeled samples, extrinsic labeling strategies, and different multi-task AL methods with respect to the two target variables; a) mean accuracy estimations where partially unlabeled samples are included in the learning process, and b) mean accuracy estimations where partially unlabeled samples are excluded from the learning process.

to *urban typology* when compared to results from the multi-variable criterion-based and rank combination methods. At the same time, an extrinsic labeling principle can yield favorable properties, if the different annotation tasks are not highly dissimilar and if an individual annotation task is not exceedingly challenging in general. This is indicated in the results of the variable *urban typology*, which shows highest classification accuracies of all three tasks, where multi-task strategies were able to partly outperform the single-task method, and also one-sided sample selections yielded favorable performance properties. However, one-sided sample selections showed similar or partly worse performance properties when compared to random selection strategies for the variables *material type* and *occupancy*. This underlines the general advantage of the multi-task meta-protocols, which feature superior solutions consistently for all three variables.

As a further means, we computed the mean κ statistic based on the results for individual variables (Fig. 8). It can be noted that single task AL allows for reaching a plateau on a high accuracy level rapidly, whereas random sampling strategies perform generally worst. Interestingly, one-sided sample selection strategies feature a diverse accuracy pattern. In this example, extrinsic labeling in conjunction with the leading variable *material type* shows almost comparable accuracies as single-task AL. In contrast, very poor accuracies are obtained if variable *urban typology* governs the selection of samples. Only slightly better accuracies are achieved if

variable *occupancy* is leading the selection process. Thus, the accuracies of one-sided sample selection are highly dependent on which variable is leading the selection. In this example, accuracy levels of one-sided sample selection are directly related to the difficulty of the classification problem: the less challenging it is, the more likely it can be properly addressed based on extrinsic labeling. Correspondingly, a very challenging variable, which leads the selection, can enable comparatively high accuracies. In contrast to one-sided sample selection, accuracies of the multi-task meta-protocols all show a favorable accuracy pattern. Thereby, the variability in accuracy of individual methods is very low, what unambiguously underlines the validity of the implemented concepts.

B. Experiment II: Simulation of RRVS based on data from flood-related decision support system

Accuracy estimates in terms of κ statistic of the different methods for the second RRVS experiment are visualized in Fig. 9. Fig. 9a shows results for the two variables where partially unlabeled samples were included, and Fig. 9b provides results where unlabeled samples were excluded. Generally, it is observable that predictions with respect to the variable *roof type* are more challenging compared to *building type*. The first hardly achieves a substantial level of

17

agreement, whereas the latter can consistently reach a substantial level with few labeled samples.

In this example, single-task MCLU-ABD could hardly achieve the accuracy levels of random selection for the variable *roof type*, whereas the alternating selection strategies MCLU-ABD *AS*, and **MCLU-ABD** *AS-stacked*, as well as strategies relying on maximum absolute uncertainty (i.e., MCLU-ABD *MVC-max*, MCLU-ABD *AS-MVC-max*) reach comparable levels and achieve simultaneously higher accuracies for the variable *building type*. The increase of accuracy with respect to the single-task approach can be related to the partially extrinsic labeling strategy and affiliated exhaustive exploration of the feature space. In this manner, it can be noted that uncertainty sampling assumes that the current hypothesis (i.e., decision function) is close to an optimum and searches for samples which are near to already established boundaries. In contrast, the partly extrinsic sample selection strategy allows here for an initial, more exploratory sampling of the feature space before working on intensive refinement of near-optimum class boundaries. However, purely extrinsic labeling as internalized by the one-sided selection strategies show lowest accuracies for both learning tasks.

In addition, it can be noted that the rank combination methods (i.e., MCLU-ABD *RC*) perform inferior for the variable *roof type* and also approaches, which rely on maximum mean uncertainty (i.e., MCLU-ABD *MVC-mean*, MCLU-ABD *AS-MVC-mean*) feature lower levels of accuracies for both variables. Notably, those methods strictly implement a strategy which selects informative samples with respect to *all* learning algorithms. Our results give a strong indication that it is a very risky strategy to systematically neglect samples which are highly informative for one task but uninformative for another. As such, the approaches which are not based on maximum mean uncertainty but on maximum absolute uncertainty (i.e., MCLU-ABD *MVC-max*, MCLU-ABD *AS-MVC-max*) as well as the strategies which select samples explicitly with respect to a single, leading variable (i.e., alternating selection) can alleviate this problem and show favorable performance properties.

Lastly, mean κ statistics based on the results for both variables were computed and are visualized in Fig. 10. Obtained results unambiguously mirror the previous findings: Multi-task AL strategies which select samples that are very relevant for one task but not necessarily for another task generally allow for higher accuracies compared to strategies that select samples which are simultaneously relevant for all tasks. One-sided selections perform consistently worst and cannot be considered as viable options here.

V. CONCLUSIONS AND FUTURE PERSPECTIVES

In this paper, we have introduced a multi-task AL framework for efficient compilation of training sets for characterization of buildings according to multiple categorical variables. To this purpose, we followed three multi-task AL meta-protocols based on multi-variable criterion, alternating selection, and rank combination and proposed a variety of novel methodological variants. The proposed methods were applied to the problem of optimizing labeling efforts by a human expert in the context of classification of building characteristics using VHR multispectral imagery and LiDAR data. The experimental results underline the usefulness of the proposed query functions. They can enable faster convergence with significantly increased classification accuracy of learned models compared to one-sided selection and random sampling strategies. Thereby, multi-task AL strategies that do not strictly neglect unlabeled samples which are highly relevant for one task but not for the other, such as alternating selection and methods based on maximum absolute uncertainty, proofed useful to consistently ensure favorable performance properties.

In the future, we aim to extend the current work and render the method as a cost-sensitive [28, 29] multi-task AL approach for guided *in situ* data collection with respect to multiple target variables. This is intended to further enable systematic and efficient characterization of built environments for applications such as natural hazard vulnerability and risk assessments. Thereby, the concept of Multiple Kernel Learning [57] can also be exploited to specifically address multimodal and heterogeneous data settings. In addition to that, multioutput models (e.g., [58], [59]) that can predict simultaneously several target variables and also feature AL capabilities appear as a consistent extension with respect to the presented framework.

REFERENCES

[1] H. Taubenböck *et al.*, "Monitoring urbanization in mega cities from space," *Remote Sens. Environ.*, vol. 117, pp. 162–176, Feb. 2012.
[2] H. Taubenböck *et al.*, "New dimensions of urban landscapes: The spatiotemporal evolution from a polynuclei area to a mega-region based on remote sensing data," *Appl. Geogr.*, vol. 47, pp. 137–153, Feb. 2014.
[3] J. Graesser *et al.*, "Image based characterization of formal and informal neighborhoods in an urban landscape," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1164–1176, Aug. 2012.
[4] H. Taubenböck *et al.*, "Delineation of central business districts in mega city regions using remotely sensed data," *Remote Sens. Environ.*, vol. 136, pp. 386–401, Sep. 2013.
[5] C. Geiß, H. Taubenböck, M. Wurm, T. Esch, M. Nast, C. Schillings, and T. Blaschke, "Remote sensing-based characterization of settlement structures for assessing local potential of district heat," *Remote Sens.*, vol. 3, no. 7, pp. 1447-1471, 2011.
[6] I. Andric, N. Gomes, A. Pina, P. Ferrao, J. Fournier, B. Lacarriere, O. Le Corre, "Modeling the long-term effect of climate change on building heat demand: Case study on a district level," *Energy and Buildings*, vol. 126, pp. 77-93, Aug. 2016.

18

[7] C. Geiß, and H. Taubenböck, "Remote sensing contributing to assess earthquake risk: From a literature review towards a roadmap," *Nat. Hazards*, vol. 68, no. 1, pp. 7–48, Aug. 2013.

[8] M. Pittore, and M. Wieland, "Towards a rapid probabilistic seismic vulnerability assessment using satellite and ground-based remote sensing," *Nat. Hazards*, vol. 68, no. 1, pp. 115–145, Aug. 2013.

[9] M. Harb, F. Dell'Acqua, "Remote Sensing in Multirisk Assessment," IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 1, 2017.

[10] C. Geiß *et al.*, "Assessment of seismic building vulnerability from space," *Earthquake Spectra*, vol. 30, no. 4, pp. 1553–1584, Nov. 2014.

[11] C. Geiß, P. Aravena Pelizari, M. Marconcini, W. Sengara, M. Edwards, T. Lakes, and H. Taubenböck, "Estimation of seismic buildings structural types using multi-sensor remote sensing and machine learning techniques," *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 175-188, 2015.

[12] G. Grünthal, R. Musson, J. Schwarz, and M. Stucchi, "European Macroseismic Scale. Cahiers de Centre Européen de Géodynamique et de Seismologie," vol 15, Luxembourg, 1998.

[13] M. Belgiu, I. Tomljenovic, T. Lampoltshammer, T. Blaschke, and B. Höfle, "Ontology-based classification of building types detected from airborne laser scanning data," *Remote Sens.*, vol. 6, no. 2, pp. 1347–1366, Feb. 2014.

[14] S. Du, F. Zhang, and X. Zhang, "Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 107–119, 2015.

[15] M. Wurm, A. Schmitt, and H. Taubenböck, "Building Types' Classification Using Shape-Based Features and Linear Discriminant Functions," *IEEE JSTARS*, vol. 9, no 5, pp. 1901-1912, May 2016.

[16] L. Breiman, "Statistical Modeling: The Two Cultures," *Statistical Science*, vol. 16, no. 3, pp. 199-231, 2001.

[17] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[18] B. Schoelkopf, and A. Smola, "Learning with Kernels," MIT Press, Cambridge MA, 2002.

[19] M. Wieland, and M. Pittore, "Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images," *Remote Sens.*, vol. 6, no. 4, pp. 2912–2939, 2014.

[20] C. Geiß, M. Jilge, T. Lakes, and H. Taubenböck, "Estimation of Seismic Vulnerability Levels of Urban Structures With Multisensor Remote Sensing," *IEEE JSTARS*, vol. 9, no 5, pp. 1913-1936, May 2016.

[21] B. Settles, Active Learning, Morgan & Claypool, 2012.

[22] P. Mitra, B. U. Shankar, and S. K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognit. Lett.*, vol. 25, no. 9, pp. 1067–1074, Jul. 2004.

[23] D. Tuia, F. Ratle, F. Pacifici, M.F. Kanevski, and W.J. Emery, "Active Learning Methods for Remote Sensing Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218-2232, 2009 .

[24] D. Tuia, L. Copa, M. Kanevski, and J. Munoz-Mari, "A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 606-617, 2011.

[25] A. Stumpf, N. Lachiche, J.P. Malet, N. Kerle, and A. Puissant, "Active Learning in the Spatial Domain for Remote Sensing Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2492 – 2507, 2014.

[26] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W.J. Emery, "SVM Active Learning Approach for Image Classification Using Spatial Information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2217-2233, 2014.

[27] S. Rougier, A. Puissant, A. Stumpf, N. Lachiche, "Comparison of sampling strategies for object-based classification of urban vegetation from Very High Resolution satellite images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 51, pp. 60-73.

[28] B. Demir, L. Minello, and L. Bruzzone "Definition of Effective Training Sets for Supervised Classification of Remote Sensing Images by a Novel Cost-Sensitive Active Learning Method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp.1272-1284, 2014.

[29] C. Persello, A. Boularias, M. Dalponte, T. Gobakken, E. Næsset, and B. Schölkopf "Cost-Sensitive Active Learning with Lookahead: Optimizing Field Surveys for Remote Sensing Data Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6652- 6664, 2014.

[30] M. Wieland, M. Pittore, S. Parolai, J. Zschau, B. Moldobekov, and U. Begaliev, "Estimating building inventory for rapid seismic vulnerability assessment: Towards an integrated approach based on multi-source imaging," *Soil Dynamics and Earthquake Engineering*, vol. 36, pp.70-83, 2012.

[31] M. Pittore, M. Wieland, M. Errize, C. Kariptas, and I. Güngör, "Improving post-earthquake insurance claim management: a novel approach to prioritized geospatial data collection," *ISPRS International Journal of Geoinformation*, 4, 2401-2427, 2015.

[32] H. Taubenböck, I. Standfuß, M. Klotz, and M. Wurm, "The physical density of the city – Deconstruction of the delusive density measure with evidence from European megacities," *ISPRS Internatl. Journal of Geo-Information*, vol. 5, no. 11, pp. 1-24, 2016.

[33] M. Herold, X. Liu, and K. Clarke, "Spatial metrics and image texture for mapping urban land use," *Photogrammetric Engineering and Remote Sensing*, vol. 69, pp. 991-1001, 2003.

[34] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," IEEE Trans. Geosci. Remote Sens., vol. 44, no. 9, pp. 2587–2600, Sep. 2006.

[35] S. Steiniger, T. Lange, D. Burghardt, and R. Weibel, "An Approach for the Classification of Urban Building Structures Based on Discriminant Analysis Techniques," *Transactions in GIS*, vol. 12, no. 1, pp. 31–59, 2008.

[36] Trimble, "eCognition developer 9.0 reference book," München: Germany Trimble Documentation, 2014.

[37] Z. Sun, H. Fang, M. Deng, A. Chen, P. Yue, and L. Di, "Regular shape similarity index: A novel index for accurate extraction of regular objects from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3737–3748, 2015.

[38] K. McGarigal, S.A. Cushman, and E. Ene, "FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps," Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at: http://www.umass.edu/landeco/research/fragstats/fragstats.html 2012.

[39] S. Angel, J. Parent, and D.L. Civco, "Ten compactness properties of circles: measuring shape in geography," Canadian Geographer / Le Géographe canadien, vol. 54, no. 4, pp. 441–461, 2010.

[40] C. Berger, M. Voltersen, R. Eckardt, J. Eberle, T. Heyer, and N. Salepci *et al.*, "Multi-modal and multi-temporal data fusion: Outcome of the 2012 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1324–1340, Jun. 2013.

[41] M. Wurm, H. Taubenböck, M. Schardt, T. Esch, S. Dech, "Object-based image information fusion using multisensor earth observation data over urban areas," Int. J. Image Data Fusion, vol. 2, pp. 121-147, 2011.

[42] P. Tompalski, and P. Wezyk, "LiDAR and VHRS data for assessing living quality in cities - an approach based on 3D spatial indices," ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIX-B6 (2012), pp. 173–176, 2012.

[43] M. Mueller, K. Segl, U. Heiden, and H. Kaufmann, "Potential of high-resolution satellite data in the context of vulnerability of buildings," *Nat. Hazards*, vol. 38, pp. 247-258, 2006.

[44] T. Leichtle, C. Geiß, M. Wurm, T. Lakes, and H. Taubenböck, "Unsupervised change detection in VHR remote sensing imagery – an object-based clustering approach in a dynamic urban environment," *International Journal of Applied Earth Observation and Geoinformation*, vol. 54, pp. 15-27, 2017.

[45] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Transactions on Systems, Man, and Cybernetics SMC-3:610–621, 1973.

[46] F. Pacifici, M. Chini, and W. Emery, "A neural network approach using multi-scale textural metrics from very high resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276-1292, 2009.

[47] G.F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inform. Theory*, vol. 14, no. 1, pp. 55–63, 1968.

[48] I. Guyon, and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.* vol. 3, pp. 1157–1182, 2003.

[49] B. Demir, C. Persello, and L. Bruzzone, "Batch-Mode Active-Learning Methods for the Interactive Classification of Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, 2011.

[50] K. Brinker, "Incorporating diversity in active learning with support vector machines," in Proc. Int. Conf. Mach. Learn., pp. 59-66, 2003.

[51] R. Reichart, K. Tomanek, and U. Hahn, "Multi-task active learning for linguistic annotations," In: ACL, 2008.

[52] Sta. Köln. Amt für Stadtentwicklung und Statistik (Hrsg.): Statistisches Jahrbuch Köln 2015. Version: 2015. http://www.stadt-koeln.de/ politik-und-verwaltung/statistik/jahrbuecher/. 2015. – Forschungsbericht. – ISSN 0933–632X

[53] G. Gruenthal, A.H. Thieken, J. Schwarz, K.S. Radtke, A. Smolka, and B. Merz, "Comparative Risk Assessments for the City of Cologne – Storms, Floods, Earthquakes," *Nat. Hazards*, vol. 38, pp. 21–44, 2006.

[54] M. Wieland, "Exposure estimation for rapid seismic vulnerability assessment: an integrated approach based on multi-source imaging," *Dissertation*, Technical University of Berlin, Germany, 2013.

[55] S. Brzev, C. Scawthorn, A.W. Charleson, L. Allen, M. Greene, K. Jaiswal, V. Silva, "GEM Building Taxonomy Version 2.0," GEM Technical Report 2013-02 V1.0.0, 188 pp., GEM Foundation, Pavia, Italy, doi: 10.13117/GEM.EXP-MOD.TR2013.02, 2013.

[56] G.M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, pp. 185–201, 2002.

[57] M. Gönen, E. Alpaydın, "Multiple kernel learning algorithms," *J. Mach. Learn. Res.*, vol. 12, pp. 2211–2268, 2011.

[58] K. Crammer, Y. Singer, "On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines," *Journal of Machine Learning Research*, vol. 2, pp. 265-292, 2001.

[59] D. Tuia, J. Verrelst, L. Alonso, F. Perez-Cruz and G. Camps-Valls, "Multioutput Support Vector Regression for Remote Sensing Biophysical Parameter Estimation," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 804-808, July 2011.