



# Semi-supervised learning with constrained virtual support vector machines for classification of remote sensing image data

Christian Geiß<sup>a,b,\*</sup>, Patrick Aravena Pelizari<sup>a,c</sup>, Ozan Tunçbilek<sup>a</sup>, Hannes Taubenböck<sup>a,c</sup>

<sup>a</sup> German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Münchener Straße 20, 82234 Weßling, Germany

<sup>b</sup> University of Bonn, Department of Geography, Meckenheimer Allee 166, 53115 Bonn, Germany

<sup>c</sup> University of Würzburg, Institute of Geography and Geology, Chair of Remote Sensing, Oswald-Kuelpe-Weg 86, 97074 Würzburg, Germany

## ARTICLE INFO

### Keywords:

Image classification  
Virtual support vector machines  
Semi-supervised models  
Self-learning  
Active learning model heuristics

## ABSTRACT

We introduce two semi-supervised models for the classification of remote sensing image data. The models are built upon the framework of Virtual Support Vector Machines (VSVM). Generally, VSVM follow a two-step learning procedure: A Support Vector Machines (SVM) model is learned to determine and extract labeled samples that constitute the decision boundary with the maximum margin between thematic classes, i.e., the Support Vectors (SVs). The SVs govern the creation of so-called virtual samples. This is done by modifying, i.e., perturbing, the image features to which a decision boundary needs to be invariant. Subsequently, the classification model is learned for a second time by using the newly created virtual samples in addition to the SVs to eventually find a new optimal decision boundary. Here, we extend this concept by (i) integrating a constrained set of semi-labeled samples when establishing the final model. Thereby, the model constraint, i.e., the selection mechanism for including solely informative semi-labeled samples, is built upon a self-learning procedure composed of two active learning heuristics. Additionally, (ii) we consecutively deploy semi-labeled samples for the creation of semi-labeled virtual samples by modifying the image features of semi-labeled samples that have become semi-labeled SVs after an initial model run. We present experimental results from classifying two multispectral data sets with a sub-meter geometric resolution. The proposed semi-supervised VSVM models exhibit the most favorable performance compared to related SVM and VSVM-based approaches, as well as (semi-) supervised CNNs, in situations with a very limited amount of available prior knowledge, i.e., labeled samples.

## 1. Introduction

Techniques for robustly deriving thematic information from remote sensing image data is of high interest. The past paradigm shift from expert-based systems, which involved implementing dedicated if-then rules to extract thematic information, to machine learning-based methods, which derive rules automatically from empirical observations, remains relevant.

In recent years, especially deep learning methods gained increasing popularity (Li et al., 2019). The underlying *model structure* often exhibits high generalization capabilities when a substantial amount of prior knowledge is available. Models like fully convolutional neural networks (CNNs) (Long et al., 2015) enable the learning of a hierarchy of discriminative feature representations, often resulting in improved accuracy for pixel-level predictions, such as semantic segmentation, when an ample amount of training data is available. However, depending on

the domain under analysis, the collection of required prior knowledge for machine learning-based methods can remain costly. Consequently, it can be desirable to implement techniques that facilitate, for example, the efficient collection and compilation of prior knowledge such as cost-sensitive active learning (Persello et al., 2014; Geiß et al., 2018), the augmentation of the feature vector by using initial model outcomes to enhance estimates with the same level of deployed prior knowledge (Geiß and Taubenböck, 2015; Geiß et al., 2022a; Geiß et al., 2022b; Zhu et al., 2021), or the implementation of a *model structure* capable of achieving high generalization even when only a small amount of prior knowledge is available for model learning. Regarding the latter, Support Vector Machines (SVM) (Schölkopf and Smola, 2002) represent a framework that has shown excellent performance in situations with (i) a limited amount of prior knowledge, (ii) high dimensional feature vectors deployed for classification, and (iii) highly complex and intrinsically non-linear class patterns (Volpi et al., 2013).

\* Corresponding author at: German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Münchener Straße 20, 82234 Weßling, Germany.  
E-mail address: [christian.geiss@dlr.de](mailto:christian.geiss@dlr.de) (C. Geiß).

The SVM algorithm establishes a so-called hyperplane, i.e., decision boundary, in between the patterns of thematic classes. To account for non-linear classification problems, labeled samples can be projected to a space of additional dimensionality using a non-linear transformation function. In this space, the hyperplane is positioned to maximize the margin between the patterns of thematic classes (Schölkopf and Smola, 2002). It is important to note that the maximum margin between the thematic classes is equally determined by two marginal decision boundaries that enclose the labeled samples nearest to the hyperplane: the Support Vectors (SVs) (Borges, 1998). Only these samples are necessary to establish the model and its corresponding class boundaries. This algorithmic property is particularly well-suited for creating accurate models based on a limited amount of prior knowledge. Moreover, additional prior knowledge can be efficiently encoded into the model based on the SVs. Specifically, SVs can be strategically employed to make the classification model invariant. Thereby, invariance refers to model robustness regarding representation changes of the objects induced by variations in their shape and size, spatial composition, or signal-to-noise ratio, respectively (Camps-Valls et al., 2014). Considering changes in representation for determining optimal decision boundaries is referred to as model invariance (Izquierdo-Verdiguier et al., 2013). As a result, the learned decision boundaries of a model should enable high accuracy when classifying unlabeled samples, even in the presence of a substantial number of changed object representations (Geiß et al., 2019). Different possibilities how invariances can be encoded in an SVM model were proposed. Those foresee the engineering of kernel functions which lead to invariant SVM, the generation of artificially transformed samples, as well as hybrid methods which internalize the principles of both aforementioned approaches (DeCoste & Schölkopf, 2002). In this context, we adopt the idea of generating a set of samples that have been artificially transformed. Such so-called virtual samples are generated by modifying the image features to which the classification model needs to be invariant. However, virtual samples are solely generated from the training samples that have become an SV within the initial learning stage of the model. This is done to selectively enrich the training set which is deployed for a second learning stage of the model. This type of model belongs to the family of Virtual Support Vector Machines-based (VSVM) methods (DeCoste & Schölkopf, 2002).

Few studies in remote sensing image classification explored the integration of virtual samples to achieve SVM model invariance. For instance, Izquierdo-Verdiguier et al. (2013) focused on creating invariances related to object rotations and sizes using altered square-shaped image subsets. Nevertheless, creating feasible invariances requires careful pre-engineering by a specialist. In contrast to that, Geiß et al. (2019) employ a self-learning technique to discard non-informative virtual samples that were eventually formed during a potentially arbitrary process of creating invariances. Therefore, the procedure's goal is to select informative virtual samples that closely resemble their corresponding SVs. Otherwise, virtual samples may introduce divergence in the model, potentially reducing its generalization capabilities. The self-learning strategy serves as a pruning mechanism. If artificially transformed samples surpass empirically determined class-specific distances from their associated SVs or a specific margin distance, they are excluded when the model is retrained. This ensures that the virtual samples employed in the final model closely resemble their corresponding SVs and satisfy margin similarity requirements with respect to previous decision boundaries. As a result, the risk of causing model divergence is mitigated.

It can be noted, that the implemented self-learning strategy internalizes principles of active learning (Tuia et al., 2009). Model heuristics of active learning methods aim to identify few relevant unlabeled samples for prioritized labeling by an annotator (Samat et al., 2015). Relevance is for instance expressed by uncertainty. This rationale is implemented by selecting uncertain unlabeled samples which are located in close proximity to the border of the hyperplane (Demir et al., 2011). In the context of the self-learning strategy, the iterative

human-machine interaction of active learning techniques was transformed into a machine-machine interaction (Geiß et al., 2019).

Related mechanisms were followed also within semi-supervised learning techniques (Dópido et al., 2013; Lu et al., 2016; Chen et al., 2022; Shu et al., 2022). Semi-supervised methods aim to encode the structural information related to the unlabeled samples for a better representation of the thematic classes and fitting decision boundaries with a better generalization capability as compared to the deployment of labeled samples alone. In a typical semi-supervised learning procedure, unlabeled samples are iteratively labeled to create semi-labeled samples using an initially learned model. Subsequently, the semi-labeled samples are deployed in conjunction with the labeled ones for learning a final supervised model. However, improperly constrained semi-labeled samples can lead to model divergence (Li and Zhou, 2015). As a result, query functions in active learning and semi-supervised models often employ opposing criteria. Active learning focuses on identifying the most uncertain unlabeled samples nearest to decision boundaries, as subsequent labeling is considered reliable. In contrast, query functions in semi-supervised learning prioritize selecting the most reliable semi-labeled samples, which are often those farthest from the decision boundary (Persello and Bruzzone, 2014).

This is where the VSVM framework with a self-learning strategy provides a unique opportunity to deploy a set of constrained semi-labeled samples for encoding invariances. It allows us to target the augmentation of samples used for model learning by selectively choosing semi-labeled samples that resemble existing SVs and meet margin similarity requirements with respect to previous decision boundaries. Our goal in this work is to improve the generalization capabilities of VSVM models by incorporating a set of constrained semi-labeled samples. As such, the pool of samples from which the model is learned from is enriched by informative semi-labeled samples. Thereby, we naturally extend the VSVM framework in the context of semi-supervised learning, since we utilize SVM/VSVM-specific model heuristics for this task.

Generally, the discussed concepts are relevant in data settings where the ground sampling distance, i.e., the geometric resolution of the data, is much smaller than the objects that need to be extracted. Such a relation can emerge in various classification problems. However, particularly image data from remote sensing missions with a sub-meter geometric resolution including WorldView-I-IV or Pléiades feature such a relation. Furthermore, the proposed techniques, including the generation of invariances, are particularly relevant for remote sensing imagery with limited spectral resolution, such as multispectral imagery. This is especially important when there is a need to generate discriminative information due to a high level of intra-class variability and low inter-class variability.

We introduce two novel algorithms that enhance the VSVM framework by incorporating constrained semi-labeled samples and virtual semi-labeled samples to encode invariances during the learning process. Our contributions in this paper are as follows:

**Contribution 1 (Model 1):** Our first model extends the VSVM by incorporating a set of constrained semi-labeled samples to establish decision boundaries. We demonstrate its relevance by applying it to very high geometric resolution multispectral remote sensing imagery.

**Contribution 2 (Model 2):** Building upon the first model, our second model deploys semi-labeled samples to generate semi-labeled virtual samples. Specifically, we modify the image features of semi-labeled samples that have become semi-labeled SVs after the initial model runs. Constrained semi-labeled virtual samples are additionally integrated when establishing the final model.

This paper is structured as follows: we use Section 2 to describe the two semi-supervised VSVM models. The experimental setup is documented in Section 3 and the corresponding results are provided in Section 4. Concluding remarks are given in Section 5.

## 2. Proposed semi-supervised VSVM models

For our models, we built upon VSVM, which are detailed in [Section 2.1](#). Subsequently, we propose two semi-supervised extensions of the algorithm: (i) we also include semi-labeled samples in the model and prune non-informative both virtual samples and semi-labeled samples with a self-learning strategy from the model before a final relearning step ([Section 2.2](#)); (ii) additionally, we perturb image features of semi-labeled samples and additionally include informative virtual semi-labeled samples before a final relearning step ([Section 2.3](#)). [Section 2.4](#) documents the procedure to encode invariances from both labeled and semi-labeled data.

### 2.1. Virtual Support Vector Machine

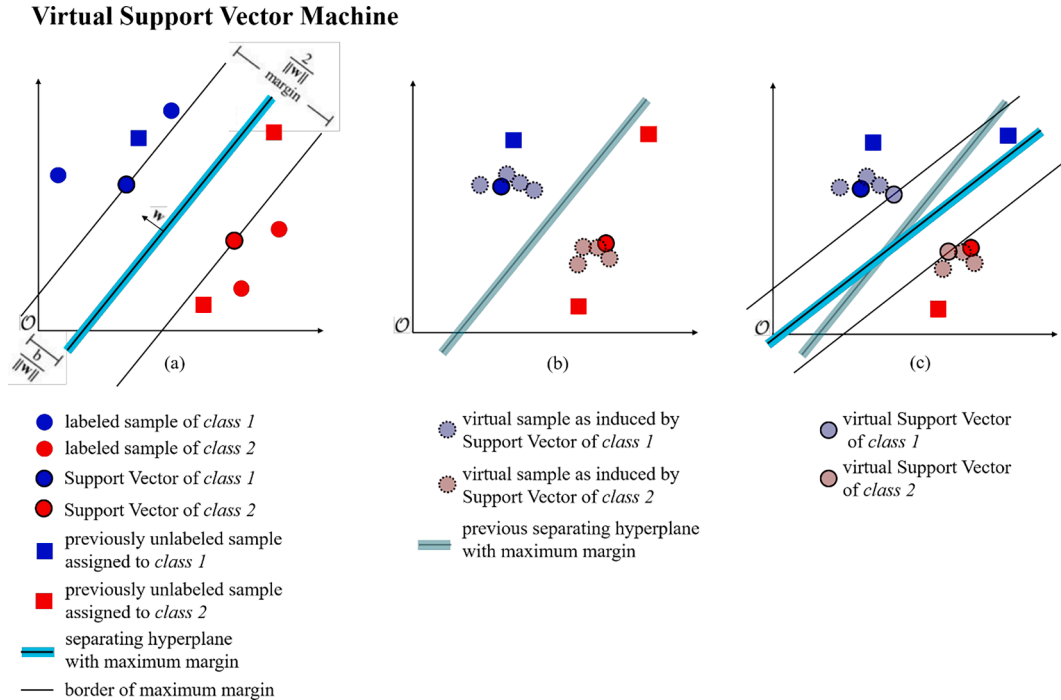
VSVM extent the well-known SVM algorithm. SVM fit a decision boundary in a way that enables a separation according to the maximum possible margin between different thematic classes ([Fig. 1a](#)) ([Schölkopf and Smola, 2002](#)).  $X = \{\mathbf{x}_i, y_i\}_{i=1}^l$  is the pool of labeled samples, with  $\mathbf{x}_i \in X$  and  $y_i = \{1, \dots, N\}$ . Thereby,  $X$  represents the  $d$ -dimensional input space  $\in \mathbb{R}^d$ . We find  $\tilde{U} = \{\mathbf{x}_i\}_{i=l+1}^{l+u} \in X$ , with  $u \gg l$  is the pool of unlabeled samples. In the VSVM approach, we begin by training an SVM model. A significant characteristic of SVM is that only a fraction of the available labeled samples makes up the actual decision boundary between classes, known as SVs. As a result, the trained SVM is used to identify labeled samples that serve as SVs. SVs determine the computation of invariances ([Section 2.4](#)) by exclusively modifying the image features of the identified SVs to which the model should be invariant. The resulting virtual samples ([Fig. 1b](#)) are combined with the SVs to retrain the model, potentially resulting in a new decision boundary ([Fig. 1c](#)). In existing scientific literature (e.g., [Izquierdo-Verdiguier et al., 2013](#)), virtual samples derived from SVs have been referred to as virtual Support Vectors (vSVs). We solely use the term vSVs if the generated virtual samples actually turn out to be samples closest to the decision boundary

after the relearning stage in order to follow an unambiguous terminology.

Further, VSVM hyperparameters need to be determined with a cross-validation strategy during the relearning stage. Otherwise virtual samples eventually exceed the number of SVs substantially. Consequently, the model can be dominantly fitted on virtual samples, while the actual prior knowledge, i.e., the SVs, has a marginal effect on model determination. Moreover, virtual samples closely resemble their corresponding SVs. Therefore, when constructing a training data pool that includes virtual samples, it may be challenging to achieve a rigorous separation and simulation of unseen data using cross-validation techniques. This phenomenon is commonly recognized as data leakage. To allow for learning a valid VSVM model, we implement the holdout method, i.e., strictly separating the data in train and test set ([Foody, 2009](#)). To account for both internal and external spatial autocorrelation, we strictly compile train and test set spatially disjoint to avoid overoptimistic model accuracy estimates ([Geiß et al., 2017a](#)).

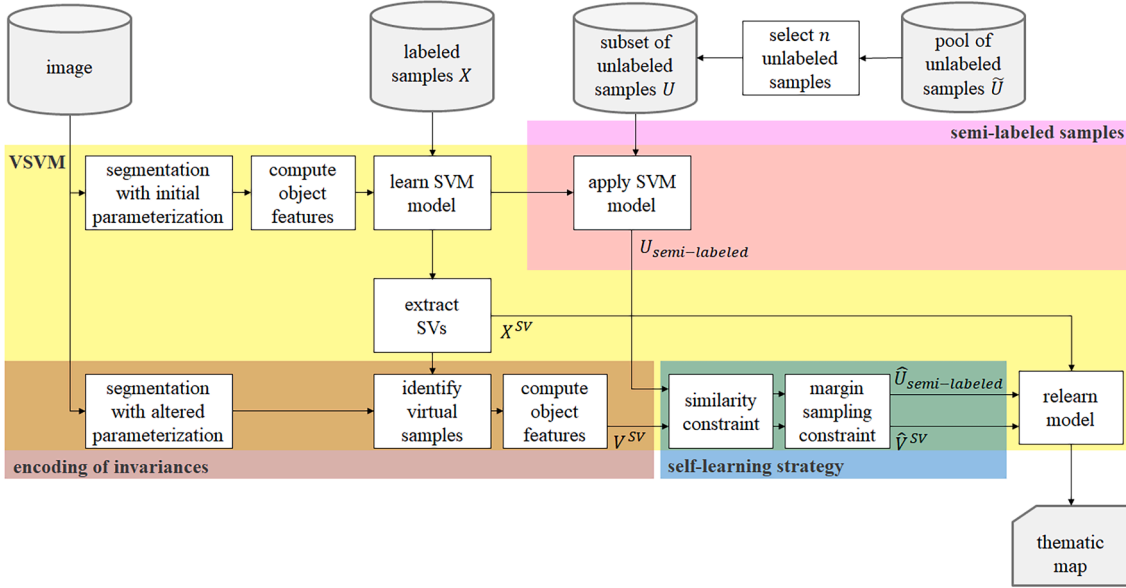
### 2.2. VSVM with self-learning constraints and semi-labeled samples

We extend VSVM by considering also unlabeled samples when establishing the model. [Fig. 2a](#) shows a corresponding scheme. Congruent with the VSVM approach, an SVM model is derived from  $X$  and deployed to extract labeled samples that represent SVs, i.e.,  $X^{SV}$ . Corresponding features are perturbed to create the corresponding pool of virtual samples  $V^{SV}$ . However, in parallel,  $n$  unlabeled samples  $U$  are randomly selected from the pool of unlabeled samples  $\tilde{U}$ , i.e.,  $U \subset \tilde{U}$ , and the SVM model is deployed to assign semi-labeled samples  $U_{semi-labeled}$ . To remove non-informative samples from  $V^{SV}$  and  $U_{semi-labeled}$ , we implement a self-learning mechanism. It is composed of a similarity and margin sampling constraint ([Fig. 3](#)). The former foresees the computation of the Euclidian distance in the feature space between the samples of  $V^{SV}$  and  $U_{semi-labeled}$  and the nearest SV ([Fig. 3a](#)) ([Lu et al., 2016](#); [Geiß et al., 2019](#)). We assume that samples contained in  $V^{SV}$  and  $U_{semi-labeled}$

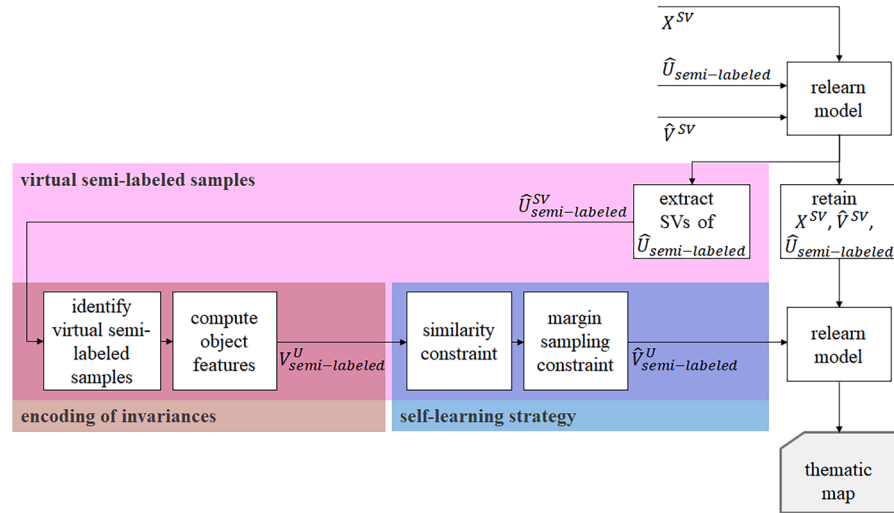


**Fig. 1.** Functionality of a VSVM: (a) the decision boundary, i.e., separating hyperplane, that enables a separation of the different thematic classes is fitted; the decision boundary is fitted in a way that enables a separation according to the maximum possible margin between the different thematic classes; (b) corresponding SVs are deployed to generate virtual samples; (c) the model is learned for a second time using SVs and virtual samples to possibly change the position of the decision boundary which maximizes the margin. Previously unlabeled samples are eventually reassigned to a thematic class w.r.t. the newly found decision boundary.

## (a) Virtual Support Vector Machine with self-learning constraints and semi-labeled samples



## (b) Virtual Support Vector Machine with self-learning constraints and virtual semi-labeled samples



**Fig. 2.** Block schemes of the proposed semi-supervised VSVM learning strategies. (a) VSVM with self-learning constraints and inclusion of constrained semi-labeled samples; (b) additional module where also virtual semi-labeled samples are generated and eventually pruned by the self-learning constraints before the relearning stage.

with a significant distance from an SV are non-informative and may lead to model divergence. Consequently, samples of  $V^{SV}$  and  $U_{semi-labeled}$  which exceed a certain class-specific distance threshold  $\delta$  w.r.t. to the SVs are pruned from the model. Additionally, a margin sampling method is implemented (Tuia et al., 2011; Geiß et al., 2017b; Geiß et al., 2018). This is done to remove samples contained in  $V^{SV}$  and  $U_{semi-labeled}$  that are positioned at a great distance from the hyperplane and are thus not likely to turn into a vSV or semi-labeled SV and add positively to the classification outcome. Only samples from  $V^{SV}$  and  $U_{semi-labeled}$  that are in close proximity of the hyperplane are retained. The tolerated margin distance is specified with the threshold  $l$  (Fig. 3b). Thereby, we use a one-against-one SVM for multiclass classification settings: if a sample's distance from the hyperplane is less than  $l$  for at least one of the class-specific hyperplanes, it satisfies the criteria. The optimal model selection for the constraint mechanism, which involves determining the class-specific distance to the SVs ( $\delta$ ) and distance to margin ( $l$ ), is

implemented as an additional minimization task. The optimal combination of these parameters is determined based on a classification accuracy metric. After pruning non-informative samples, we concatenate constrained semi-labeled samples  $\hat{U}_{semi-labeled}^{SV}$  and constrained virtual samples  $\hat{V}^{SV}$  with the SVs to form the new training set  $\tilde{X} = X^{SV} \cup \hat{V}^{SV} \cup \hat{U}_{semi-labeled}^{SV}$ . This combined dataset is used to retrain the model for a second time (Fig. 3c).

VSVM with self-learning constraints and semi-labeled samples

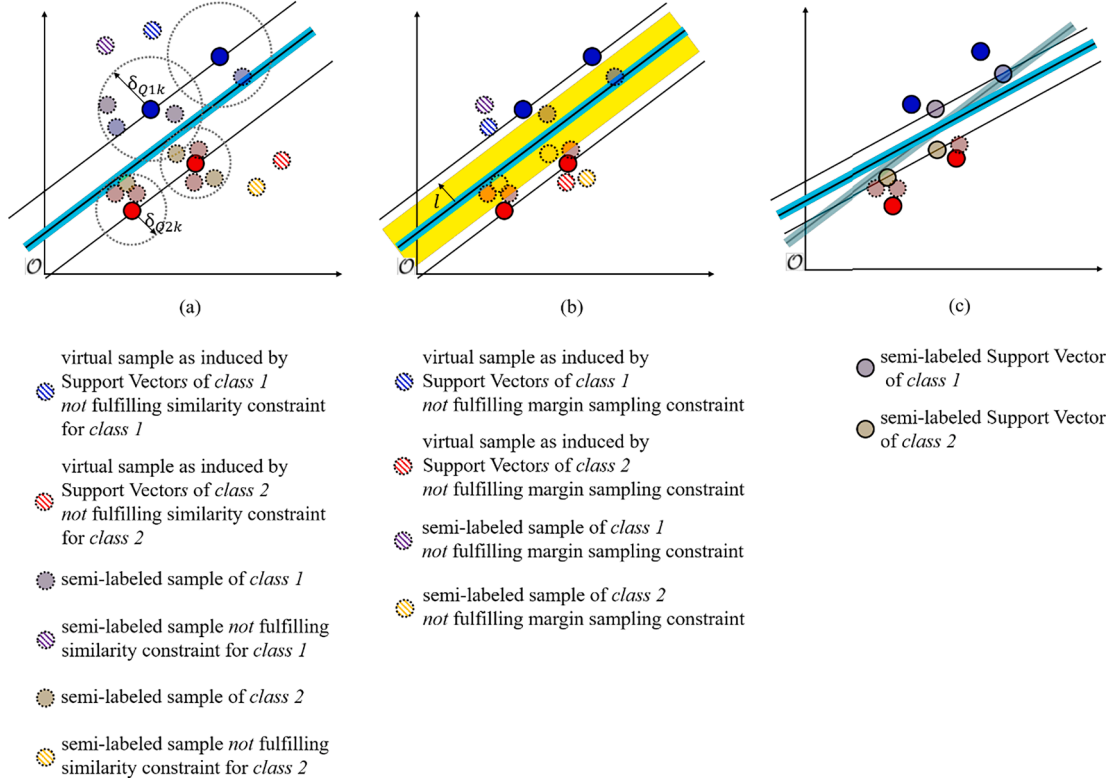
## Algorithm 1

**Inputs:**Labeled samples:  $X$ Pool of unlabeled samples:  $\tilde{U}$ **Output:**

(continued on next page)



### Virtual Support Vector Machine with self-learning constraints and semi-labeled samples



**Fig. 3.** Self-learning constraints where both virtual samples and semi-labeled samples are subsequently pruned from the model with a) the similarity constraint and b) the margin sampling constraint before c) relearning the model based on the remaining samples.

(continued)

A model learned with Support Vectors, constrained virtual samples  $\hat{V}^{SV}$ , and constrained semi-labeled samples  $\hat{U}_{semi-labeled}$

1. Learn an SVM using  $X$
2. Extract SVs from  $X$  and create the corresponding pool of SVs, i.e.,  $X^{SV}$

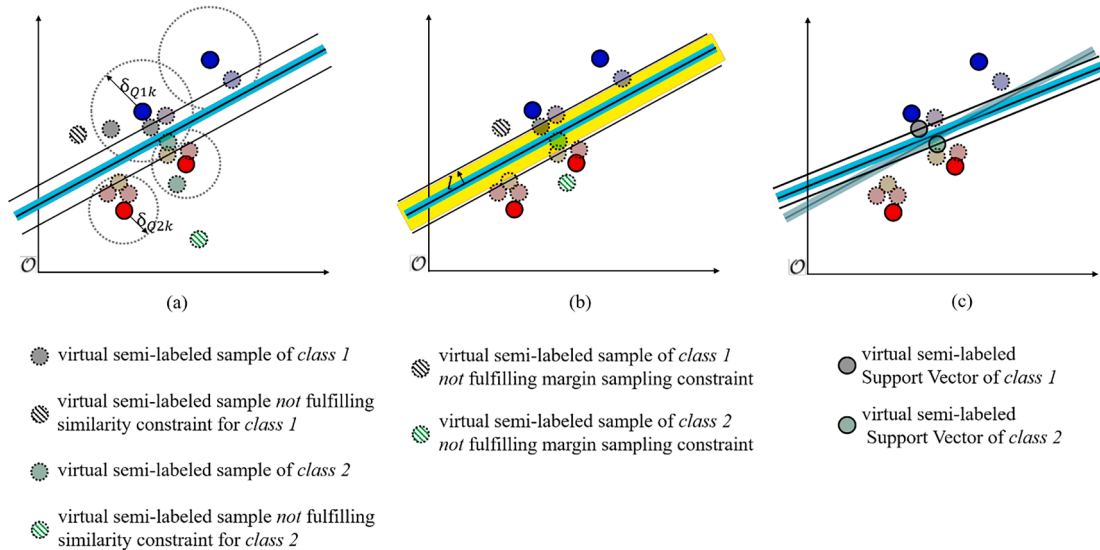
(continued on next column)

(continued)

3. Modify image features based on  $X^{SV}$  to create a pool of virtual samples  $V^{SV}$
4. Select  $n$  unlabeled samples from all unlabeled samples  $\tilde{U}$  to create a subset of unlabeled samples  $U$
5. Apply model to  $U$  and obtain semi-labeled samples  $U_{semi-labeled}$

(continued on next page)

### Virtual Support Vector Machine with self-learning constraints and virtual semi-labeled samples



**Fig. 4.** Extended self-learning strategy where virtual semi-labeled samples are pruned from the model with a) the similarity constraint and b) the margin sampling constraint before c) relearning the model based on the remaining samples.

(continued)

- 
6. Apply self-learning strategy on  $U_{\text{semi-labeled}}$  and  $V^{\text{SV}}$
  7. Establish subset of constrained semi-labeled samples  $\hat{U}_{\text{semi-labeled}}$  and constrained virtual samples  $\hat{V}^{\text{SV}}$
  8. Compile training set  $\tilde{X} = X^{\text{SV}} \cup \hat{V}^{\text{SV}} \cup \hat{U}_{\text{semi-labeled}}$
  9. Learn SVM model with  $\tilde{X}$
- 

### 2.3. VSVM with self-learning constraints and virtual semi-labeled samples

We further extend the previously proposed semi-supervised procedure by modifying the image features of semi-labeled samples that have become a semi-labeled SV after the second model run (Fig. 4). We extract SVs of  $\hat{U}_{\text{semi-labeled}}$  and add them to the pool  $\hat{U}_{\text{semi-labeled}}^{\text{SV}}$ . Analogous to Algorithm 1, we use  $\hat{U}_{\text{semi-labeled}}^{\text{SV}}$  to perturb features and encode invariances for specific semi-labeled samples that have become SVs, denoted as  $V_{\text{semi-labeled}}^{\text{U}}$ . We employ the self-learning strategy to prune samples and create a subset of constrained virtual semi-labeled samples, referred to as  $\hat{V}_{\text{semi-labeled}}^{\text{U}}$  (Fig. 4a and b). In the final step, we concatenate samples representing SVs in the training set, constrained virtual samples  $\hat{V}^{\text{SV}}$ , constrained semi-labeled samples  $\hat{U}_{\text{semi-labeled}}$ , and constrained virtual semi-labeled samples  $\hat{V}_{\text{semi-labeled}}^{\text{U}}$  to create the new training set  $\tilde{X} = X^{\text{SV}} \cup V^{\text{SV}} \cup \hat{U}_{\text{semi-labeled}} \cup \hat{V}_{\text{semi-labeled}}^{\text{U}}$ . This combined dataset is used for relearning the model (Fig. 4c).

VSVM with self-learning constraints and virtual semi-labeled samples

#### Algorithm 2

**Input** (=output of Algorithm 1):

A model learned with Support Vectors, constrained virtual samples  $\hat{V}^{\text{SV}}$ , and constrained semi-labeled samples  $\hat{U}_{\text{semi-labeled}}$

**Output:**

A model learned with Support Vectors, constrained virtual samples  $\hat{V}^{\text{SV}}$ , constrained semi-labeled samples  $\hat{U}_{\text{semi-labeled}}$ , and constrained virtual semi-labeled samples  $\hat{V}_{\text{semi-labeled}}^{\text{U}}$

1. Extract SVs of  $\hat{U}_{\text{semi-labeled}}$  and create the corresponding pool  $\hat{U}_{\text{semi-labeled}}^{\text{SV}}$
  2. Modify image features based on  $\hat{U}_{\text{semi-labeled}}^{\text{SV}}$  to create a pool of virtual semi-labeled samples  $V_{\text{semi-labeled}}^{\text{U}}$
  3. Apply self-learning strategy on pool of virtual semi-labeled samples  $V_{\text{semi-labeled}}^{\text{U}}$  to establish a subset of constrained virtual semi-labeled samples  $\hat{V}_{\text{semi-labeled}}^{\text{U}}$
  4. Compile training set  $\tilde{X} = X^{\text{SV}} \cup V^{\text{SV}} \cup \hat{U}_{\text{semi-labeled}} \cup \hat{V}_{\text{semi-labeled}}^{\text{U}}$
  5. Learn SVM model with  $\tilde{X}$
- 

### 2.4. Encoding of invariances

We encode invariances based on image segmentation techniques, which are an essential constituting aspect of object-based image analysis methods (Blaschke, 2010). The first step includes the modelling of the objects that need to be extracted from the (image) data with a segmentation algorithm. The resulting segments, i.e., super-pixels, are subsequently deployed to characterize the objects of interest. In challenging classification settings where only a very limited number of labeled samples are available, the training data frequently covers solely a minor fraction of all existing object variations in the image domain. Additionally, an optimal representation of the objects of an image domain by segments remains challenging, despite approaches to compute and select an optimal segmentation in an automated way (Geiß et al., 2016a). Our strategy to cope with this situation foresees the variation of the parameters of a segmentation algorithm to generate a variety of object representations (details on the deployed segmentation algorithm are provided in the experimental setup Section 3.2). The parameters are varied w.r.t. to both size (i.e., object scale) and shape properties of the modelled objects, respectively. The technical

implementation foresees the following processing steps: (i) a segmentation level with an initial parametrization of the segmentation algorithm is established; object features are calculated; an SVM model is established; and SVs are extracted (Fig. 2a); (ii) SVs are located and object representations from additional segmentation levels with varied parametrization are included in the model if they contain an SV; (iii) selected segments are deployed to compute features, which are integrated as virtual samples (cf. Fig. 1). Following this procedure, virtual sample equate according to the amount of SVs multiplied with the additional segmentation layers with varied parametrization. Likewise, virtual semi-labeled samples are generated by extracting the SVs of constrained semi-labeled samples, locating the SVs in the image domain, selecting corresponding additional segmentation levels with varied parametrization, and computing features which are depicted as virtual semi-labeled samples (Fig. 2b).

We encode two sorts of invariances for model enhancement: (object) scale and (object) shape, respectively. To render the model invariant w.r.t. size, i.e., scale, we establish a set of segmentation layers in a hierarchical way (Geiß et al., 2016b; Aravena Pelizari et al., 2018): the segments of a particular segmentation level must only be contained in one segment at the subsequent coarser segmentation level in order to guarantee an explicit hierarchy (cf. Geiß et al., 2019). Thereby, small scales allow for a valid representation of the smallest objects in the image, while large scales allow for representing the largest objects properly. To render the model invariant w.r.t. shape, we alter the parameters that constrain the shape-related characteristics of objects while maintaining the scale-related parameter unvaried. Analogous to the computational steps to establish invariance w.r.t. scale, a set of segmentations is computed for an exhaustive description of the image objects. Thereby, the self-learning mechanism is intended to remove non-informative virtual samples, as induced by the non-proper segment-based representation of objects, from the classification model in an automated fashion. Finally, generated virtual samples are added to the set of virtual samples  $V^{\text{SV}}$  and virtual semi-labeled samples  $V_{\text{semi-labeled}}^{\text{U}}$  (Fig. 5), respectively, which can be used for relearning the model.

## 3. Data and experimental setup

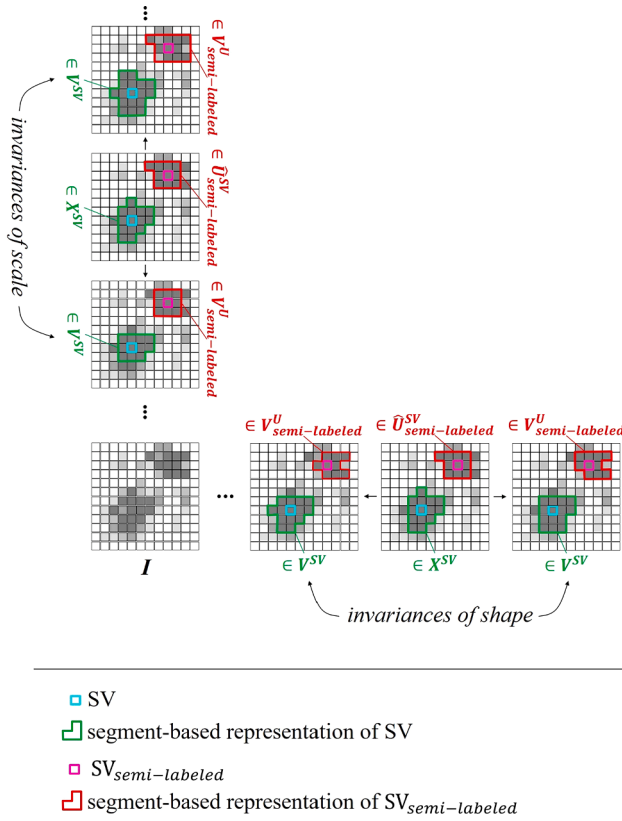
### 3.1. Data

We apply the algorithms for classifying two data sets of WorldView-II multispectral data with 0.5 m geometric resolution. The first data set covers a small part of the built-up area of Cologne in Germany, and was recorded on January 31, 2014 (Fig. 6a). The imagery was taken from an off-nadir position. Six thematic classes were defined, i.e., “bush/tree”, “meadow”, “roof”, “facade”, “shadow”, and “other impervious surface” (Fig. 6b). The six thematic classes were assigned using techniques of photointerpretation while including both additional aerial imagery and cadastral information. As mentioned in Section 2.1, we strictly separate the image data spatially according to training, test, and validation data (Fig. 6c).

The second data set represents a subset of 2000 × 2000 pixels showing the refugee camp of Hagadera in Kenya. It was acquired on March 01, 2012 (Fig. 6d). We distinguish five thematic classes, i.e., “built-up area”, “bush/tree”, “bare soil”, “fence/wall”, and “shadow” (Fig. 6e). The labels were determined with techniques of photointerpretation while integrating a thematic map from UNHCR (UNHCR, 2012). Fig. 6f documents the spatial separation of training, test, and validation data.

### 3.2. Experimental setup

To establish multiple segmentation layers for the generation of invariances, we followed a bottom-up region-growing segmentation procedure (Baatz & Schäpe, 2000). Thereby, a number of parameters need



**Fig. 5.** Illustration of the creation of both virtual samples (green) and virtual semi-labeled samples (red) regarding scale and shape properties for a labeled and semi-labeled sample, respectively. Initially, an image segmentation algorithm with varying parameters is applied to the image to model (varying green/red object outlines) and characterize (compute features) the objects of interest. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to be determined: To create an initial segmentation level, we parameterized the algorithm in a way that stresses the shape heterogeneity of generated segments. We followed this strategy since objects of the built environment feature distinct shape and size characteristics. The numerical value for the so-called scale parameter of the algorithm, which determines the extent of the modelled objects, corresponds to 20 for the Cologne data set and 25 for the Hagadera data set to establish a suitable tradeoff between under- and oversegmentation. When establishing scale invariance, we created nine extra segmentations with varied scale parameter for the first data set and seven extra segmentations for the second data set, respectively. To encode invariances of shape, we created for both images eight additional segmentation layers with varying parametrization of shape properties of modelled objects. A very detailed description of the optimization of all parameters can be found in Geiß et al. (2019).

An exhaustive number of features was computed to describe the modelled objects. The features comprise statistical metrics of central tendency (mean) and spread (standard deviation), regarding the bands of the multispectral imagery, and the *Normalized Differenced Vegetation Index*. Additionally, rotation-invariant texture metrics based on the grey-level co-occurrence matrix (GLCM) (Guo et al., 2021; Haralick et al., 1973) were computed. Regarding the latter, we deployed three measures, i.e., *mean*, *homogeneity*, and *dissimilarity*. Additionally, five shape-related features were computed, i.e., the so-called rectangular and elliptic fit, roundness, shape index, as well as compactness. Analogous to the segmentation procedure, we deployed the software eCognition (Trimble, 2014) to calculate the features based on provided or customized feature computation protocols. It can be noted that the

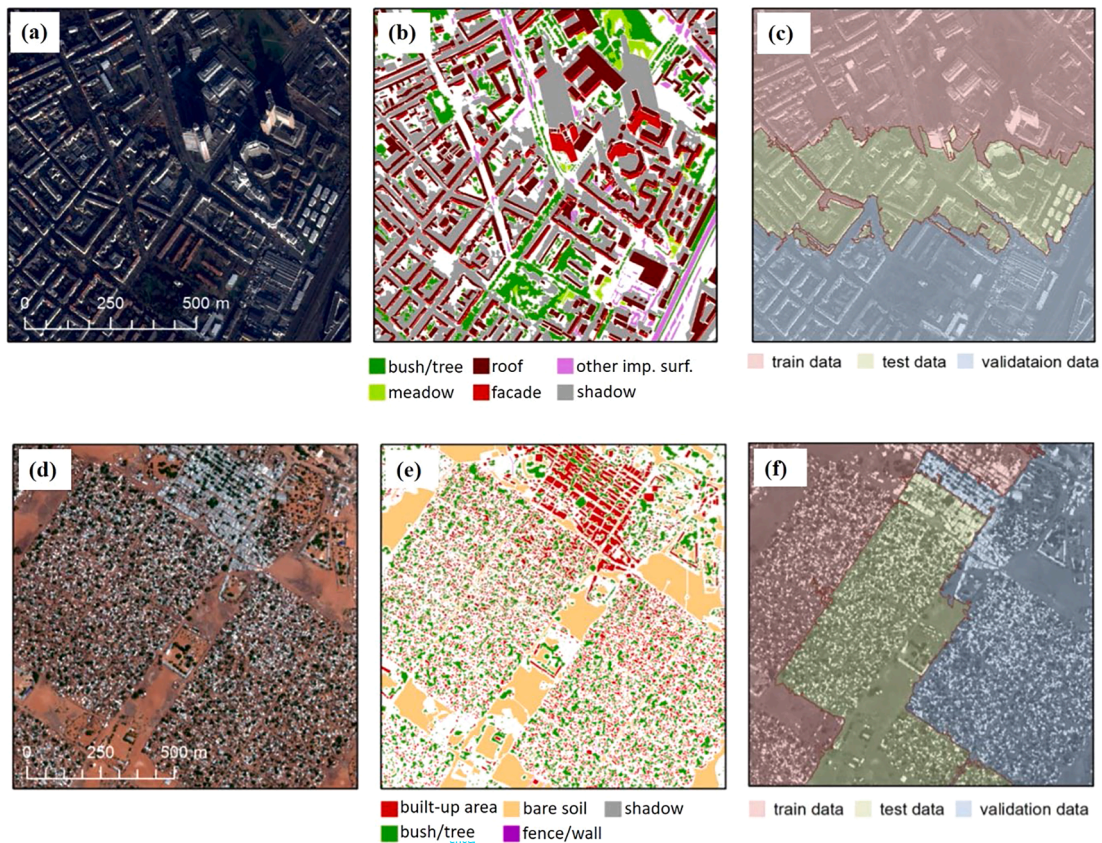
numerical feature values were normalized to a 0–1 interval based on the segmentation with the initial parameterization. Subsequently, feature values as induced by segmentations with varied parameterization were aligned accordingly.

We used Gaussian RBF kernels for the models. We determined  $C$  and  $\gamma$  for each model individually as follows:  $C = \{2^{-4}, 2^{-3}, \dots, 2^{12}\}$  and  $\gamma = \{2^{-5}, 2^{-4.5}, \dots, 2^3\}$ . We carried out experiments for both binary and multiclass classification, respectively. Regarding binary classification, we distinguish the class “bush/tree” from the remaining classes for data set I, and “built-up area” from the remaining classes for data set II. Naturally, we distinguish between all six and five of the thematic classes present in the Cologne and Hagadera data set, respectively, in terms of multiclass categorization. We created balanced training and test sets by randomly picking labeled samples from the training and test sets in a stratified way, using the same number of labeled samples per thematic class for learning and selecting a model. However, subsequently the amount of labeled samples per thematic class was varied to quantify the corresponding sensitivity of the model w.r.t. accuracy, i.e., we learn all models with a varying *number of labeled samples per class* = {10, 20, 30, 60, 90, 120, 160, 200}. Thereby, we treated the number of semi-labeled samples and virtual semi-labeled samples as a further hyperparameter that is required to be determined for all considered semi-supervised models in this work. Consequently, we deploy interchangeably up to 100 (20, 40, ..., 100) semi-labeled and virtual semi-labeled samples per thematic class additionally for model learning and select the model with the highest estimated generalization capability in terms of overall accuracy. Results for each model, learned with a specific number of labeled samples, are documented as averaged accuracies of 20 fully independent runs. However, independent from the SVM/V SVM-based algorithms, we implemented both CNNs (Geiß et al., 2022a) and semi-supervised CNNs to gain further insights on the competitiveness of the proposed models. Regarding the latter, we designed a semi-supervised CNN framework. Thereby, multinomial logistic regression is deployed as classifier, whose scores represent class-conditional probabilities given by the softmax function (Geiß et al., 2022a). The class-conditional probabilities, obtained after an initial model run, are deployed to determine reliable semi-labeled samples which are included in a second model run. Here it can be noted, that all considered methods in this paper have strictly seen the same amount of prior knowledge. This is in particular relevant, since especially deep learning-based methods are frequently pretrained based on additional prior knowledge and subsequently transferred to address small sample settings, for instance in the context of few-shot learning approaches (Wang et al., 2020). To account for the available computational resources for this study, we provide averaged accuracies as a function of five fully independent runs regarding the considered CNN-based techniques. Additionally, corresponding classification maps (with additional  $\kappa$  and F1 statistics) from single model runs with 20 labeled samples per thematic class are presented to particularly highlight visual differences in the classification maps for settings with a few labeled samples.

#### 4. Experimental results and discussion

The added value of the semi-supervised models is demonstrated by benchmarking them w.r.t. an SVM model which is built from features using the initial segmentation level, i.e., the model which is actually deployed for generating SVs (SVM-single-level). Moreover, an SVM is trained utilizing features computed from multiple segmentations (SVM-multi-level): further encoded object attributes are represented as extra features rather than virtual samples like in the VSVM approach (a likewise method is described in Bruzzone and Carlini, 2006). Additionally, we integrate accuracies of the VSVM when established with self-learning constraints (VSVM-SL) and without self-learning constraints (VSVM), respectively. Moreover, we also implemented a semi-supervised SVM, where also a self-learning strategy was adopted to





**Fig. 6.** Data set I – Cologne: (a) multispectral imagery, (b) thematic classes, (c) distinction of areas according to training, test, and validation data; Data set II – Hagadera: (d) multispectral imagery, (e) thematic classes, (f) distinction of areas according to training, test, and validation data.

solely include relevant semi-labeled samples (SVM-semi-labeled-SL). As additional benchmark, results from the supervised CNNs (CNN) and semi-supervised CNNs (CNN-semi-labeled) are also documented. In the following, the two newly proposed models are termed VSVM-SL-semi-labeled and VSVM-SL-virtual-semi-labeled, respectively. The considered classification problems represent situations where solely a few labeled training samples can be deployed for model learning.

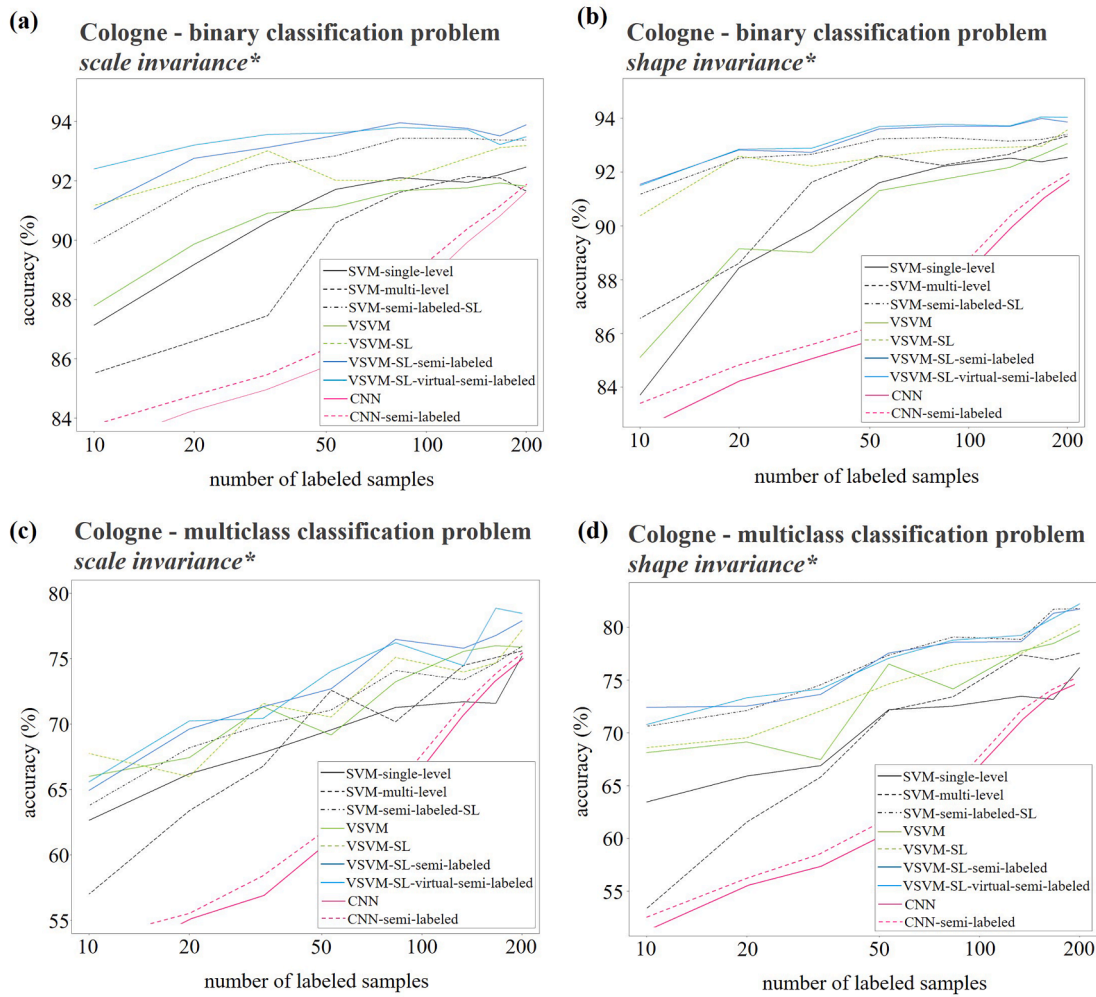
#### 4.1. Results from data set I (Cologne)

Fig. 7 documents overall accuracies as a function of the number of labeled samples for data set I. For the binary classification problem, both types of invariances reveal comparable accuracies (Fig. 7a-b). The two newly proposed semi-supervised VSVM models allow obtaining unambiguously the highest accuracies across the whole spectrum of available prior knowledge and establish already with a few labeled samples a plateau-like performance pattern. A notable competitor represents the constrained semi-supervised SVM model (SVM-semi-labeled-SL), with slightly decreased performance properties, which underlines the general usefulness of informative semi-labeled samples. The VSVM-SL approach also achieves high accuracy levels especially with a few labeled samples. Overall, this stresses the necessity to solely include constrained, i.e., informative, both semi-labeled or virtual samples, to avoid model divergence. Correspondingly, the accuracies of unconstrained VSVM, SVM-single-level, and SVM-multi-level reveal a gap in performance compared to the aforementioned approaches for settings with a few labeled samples. Lastly, both CNN-based approaches require a substantially larger number of labeled samples to accelerate the gain in accuracy and achieve a competitive accuracy level, whereby the CNN-semi-labeled model provides slight but consistent advantages regarding accuracy compared to the CNN.

The related classification maps obtained with 20 samples per class reflect these relationships (Fig. 8a): they indicate that the conventional SVM-based methods, i.e., SVM-single-level and SVM-multi-level, suffer from commission errors regarding the class “bush/tree”. In contrast, the more advanced methods are able to reduce errors of commission for the class “bush/trees”. This is also true for the CNN-based models, whereby the high level of commission errors regarding the class “bush/tree” produced by the CNN could be reduced by the CNN-semi-labeled approach. However, for both CNN-based methods a low level of spatial regularization can be observed. In contrast, in this binary classification setting, the two newly proposed models feature the most favorable tradeoff between errors of commission and omission, i.e., highest F1 values, and they provide spatially smooth classification maps.

A likewise accuracy pattern can be inferred from the multiclass classification problem with shape invariance (Fig. 7d). It is exigent that more labeled samples are required to obtain high accuracy levels in the multiclass environment than in the binary one. Thereby, the semi-supervised methods unambiguously allow for the highest accuracies. The strategy to encode additional object characteristics as features and not as virtual samples, i.e., SVM-multi-level, reveals increasing accuracies solely with an increasing number of labeled samples, i.e., with a decreasing level of overfitting. Thereby the SVM-single-level models are overtaken along the spectrum of available prior knowledge in an ideal-typical way by the SVM-multi-level models. This indicates that the virtual samples-based approaches avoid problems related to a curse of dimensionality, i.e., no additional dimensions are added to the feature space what can be disadvantageous in small sample settings. When encoding scale invariance (Fig. 7c), the two newly proposed semi-supervised VSVM models feature frequently the highest and most consistently increasing accuracies, which further underlines the beneficial performance properties induced by a constrained set of semi-





\*scale and shape invariance are solely intrinsic w.r.t. SVM/VSVM-based algorithms in this work

**Fig. 7.** Overall accuracy (%) reported as mean from twenty independent realizations as a function of the number of labeled samples per thematic class (x-axis); (a) binary classification problem with encoded scale invariance; (b) binary classification problem with encoded shape invariance; (c) multiclass classification problem with encoded scale invariance; (d) multiclass classification problem with encoded shape invariance.

labeled and virtual semi-labeled samples for model generation, respectively. Also here, both CNN-based models can solely reach a competitive accuracy level when learned from a substantial amount of prior knowledge. Corresponding classification maps achieved with 20 samples per class (Fig. 8b) further stress the properties of the SVM/VSVM-based semi-supervised models to enable spatially regularized and more precise maps in this very challenging multiclass classification setting with very few labeled samples.

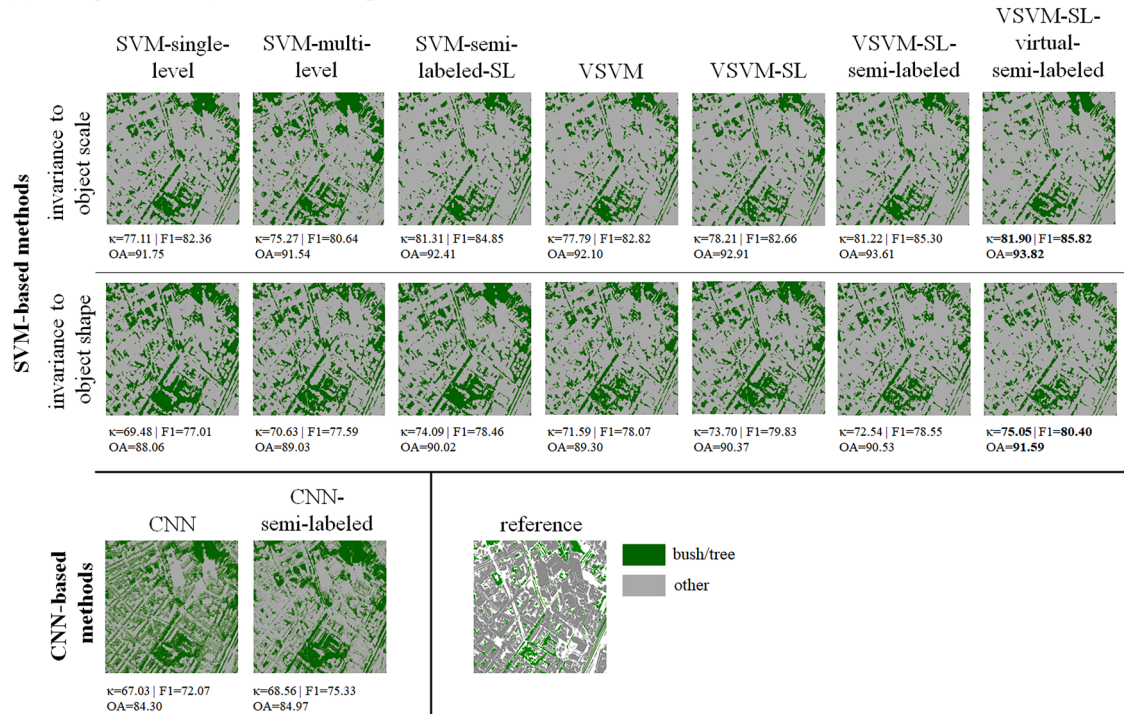
#### 4.2. Results from data set II (Hagadera)

We plotted overall accuracies w.r.t. the number of labeled samples for data set II in Fig. 9. In the binary classification context (Fig. 9a-b), the models utilizing constrained semi-labeled samples are consistently beneficial. Notably, regarding scale invariance, the model which also includes virtual semi-labeled samples (VSVM-SL-virtual-semi-labeled) induces additional performance gains with very little prior knowledge, i. e., 10–20 labeled samples per thematic class. Also here, the constrained VSVM model (VSVM-SL) allows for competitive accuracies and the unconstrained models, i. e., VSVM, SVM-single-level, and SVM-multi-level, feature the least favorable accuracy patterns of the SVM/VSVM-based methods. Analogous to data set I, the (semi-)supervised CNN algorithms require a substantially larger number of labeled samples for

reaching a competitive accuracy level. However, while the SVM/VSVM-based methods reach a plateau fairly soon with only moderate improvements given additional prior knowledge, the (semi-)supervised CNN can consistently improve as the amount of prior knowledge grows. Thereby, the CNN-semi-labeled approach provides slight but consistent advantages regarding accuracy compared to the CNN. Also, the associated classification maps mirror these numbers (Fig. 10a). SVM-based maps feature a substantial error of commission regarding “built-up” areas. Instead, the (semi-supervised) constrained virtual sample-based models allow for considerably few errors of commission for “built-up” areas while jointly enabling spatially well-regularized yet fine-grained maps. Contrarily, given such as restricted amount of prior knowledge, the CNN-based maps suffer from a high level of omission of “built-up” areas.

Multiclass classification setting results are documented in Fig. 9c-d. The models which include constrained semi-labeled samples, i. e., SVM-semi-labeled-SL and VSVM-SL-semi-labeled, enable the highest accuracies. Thereby, our newly proposed VSVM-SL-semi-labeled model enables consistently the highest accuracy with a few labeled samples. When models are learned with 50 or more labeled samples per class, all SVM/VSVM-based techniques converge to a plateau of maximum accuracy, similar to the binary classification problem. Thereby, conventional SVM-based methods, i. e., SVM-single-level and SVM-multi-level,

## (a) Cologne – binary classification problem



## (b) Cologne – multiclass classification problem

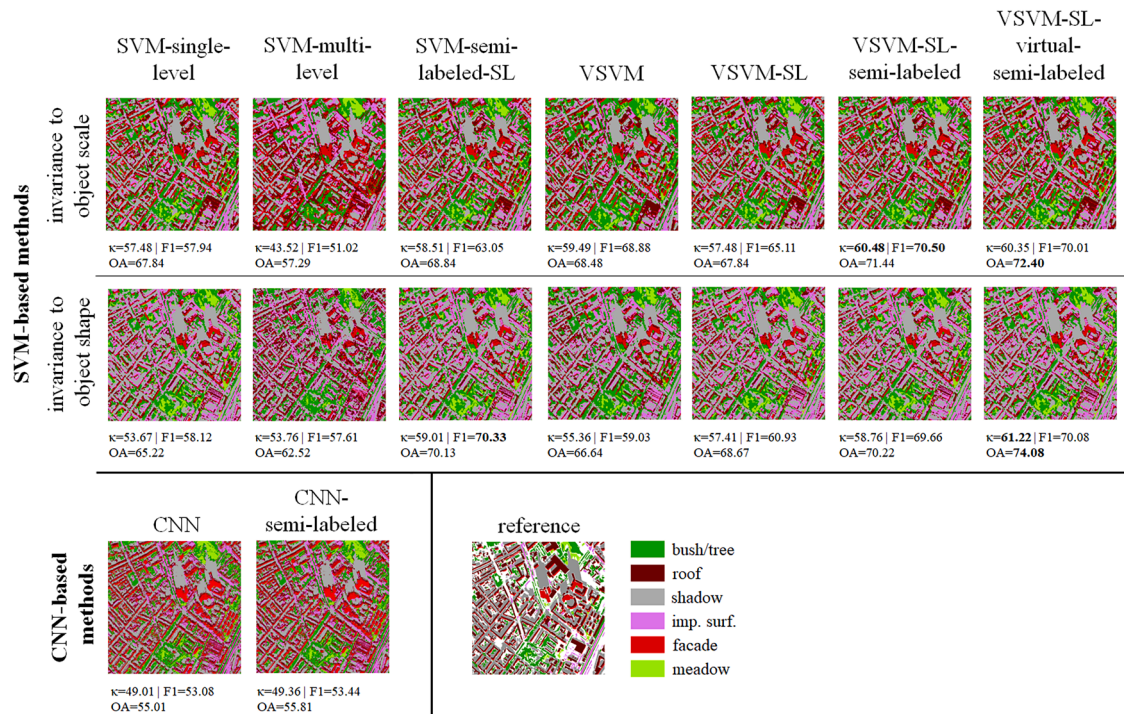
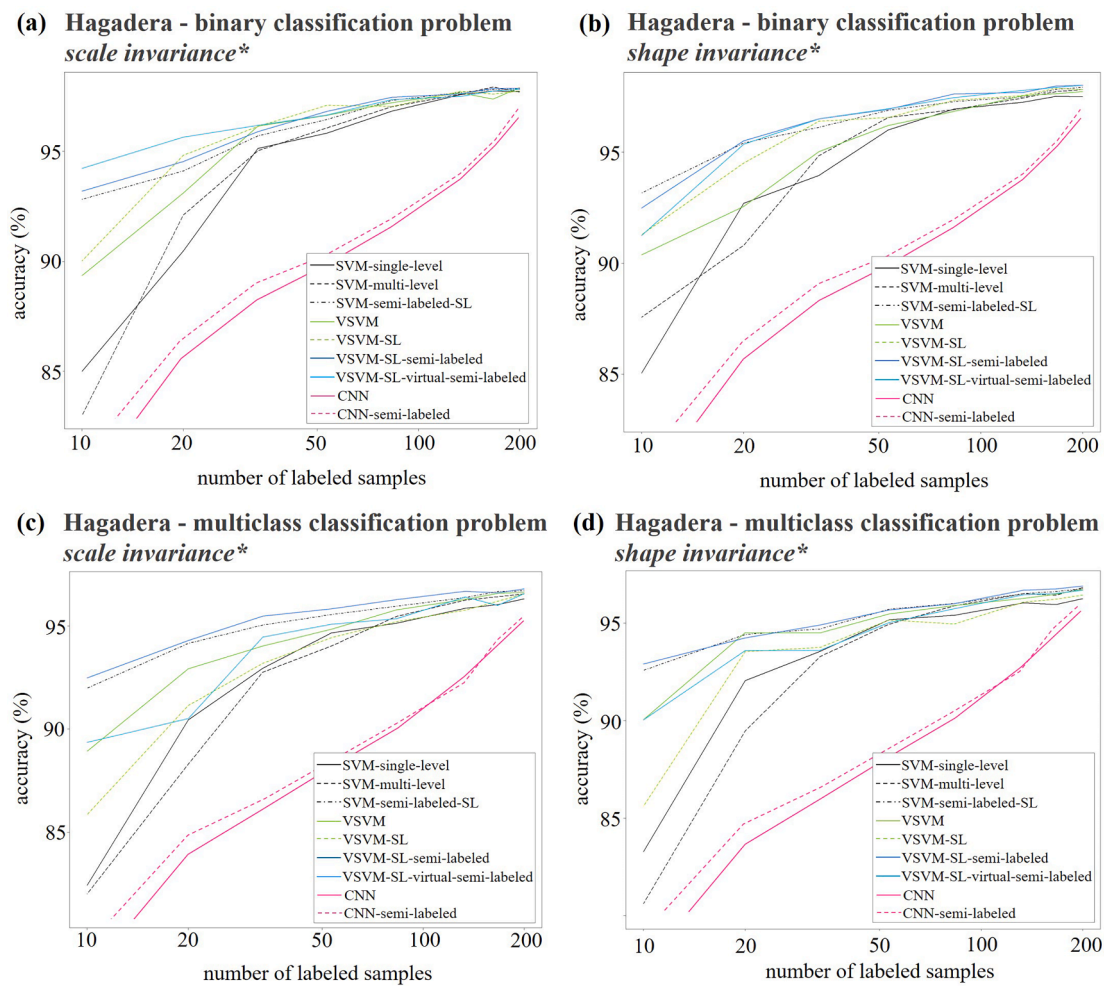


Fig. 8. Classification maps obtained with the various models for the binary (a) and multiclass (b) classification problem, respectively.

reveal least favorable performance properties. Also analogous to the binary classification problem, the (semi-)supervised CNN algorithms require a larger number of labeled samples and are not competitive in very small samples settings. For the scale invariance setting, a substantial commission error regarding the class “fence/wall” can be observed (Fig. 10b), especially when relying on the SVM-multi-level approach. However, a substantially better tradeoff in this multiclass classification setting can be achieved by the other SVM/VSVM-based

methods, especially when applying the semi-supervised models. The obtained maps from the scale invariance setting suggest that all methods except for the two newly proposed models suffer from a commission error w.r.t. the class “built-up”. As such, VSVM-SL-semi-labeled and VSVM-SL-virtual-semi-labeled simultaneously internalize a reduction of commission errors w.r.t. the class “built-up” and a reduction of omission errors regarding the class “fence/wall” compared to the baseline SVM, respectively, which allows for more accurate and fine-grained mapping





\*scale and shape invariance are solely intrinsic w.r.t. SVM/VSVM-based algorithms in this work

**Fig. 9.** Overall accuracy (%) reported as mean from twenty independent realizations as a function of the number of labeled samples per thematic class (x-axis); (a) binary classification problem with encoded scale invariance; (b) binary classification problem with encoded shape invariance; (c) multiclass classification problem with encoded scale invariance; (d) multiclass classification problem with encoded shape invariance.

results. To sum up, the outcomes for data set II also underline the benefits of the virtual samples-based semi-supervised learning techniques in terms of model accuracy.

## 5. Conclusion and outlook

We introduced two semi-supervised learning algorithms which are based on the framework of VSVM. The first model extends VSVM by integrating additionally constrained semi-labeled samples. Thereby, the model constraintment, i.e., the selection mechanism for including solely informative semi-labeled samples, builds upon a self-learning procedure composed of two active learning heuristics. The second model consecutively deploys semi-labeled samples for generation of semi-labeled virtual samples by modifying the features of semi-labeled samples that have become a semi-labeled SV after initial model runs. The proposed techniques were deployed to classify two multispectral data sets with a sub-meter geometric resolution. Classification results highlight the efficiency of the suggested techniques, which provide better accuracy properties in settings with very few labeled samples compared to related benchmark methods including SVM (single-level and multi-level), and VSVM with self-learning and without self-learning constraints, as well as (semi-)supervised CNNs, respectively.

Future work can naturally adapt the proposed algorithms in the

context of a collaborative learning approach, i.e., combining active learning with semi-supervised learning (e.g., Munoz-Mari et al., 2012; Pan et al., 2018), since the self-learning procedure already internalizes dedicated active learning heuristics for SVM. Beyond, it would be interesting to substitute the presented super-pixel-based invariance generation process with a representation learning approach (Bengio et al., 2013), presumably based on models that aim to generalize well in small sample scenarios, such as contrastive learning (Jaiswal et al., 2021).

## CRediT authorship contribution statement

**Christian Geiß:** Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Patrick Aravena Pelizari:** Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Ozan Tunçbilek:** Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Hannes Taubenböck:** Conceptualization, Methodology, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

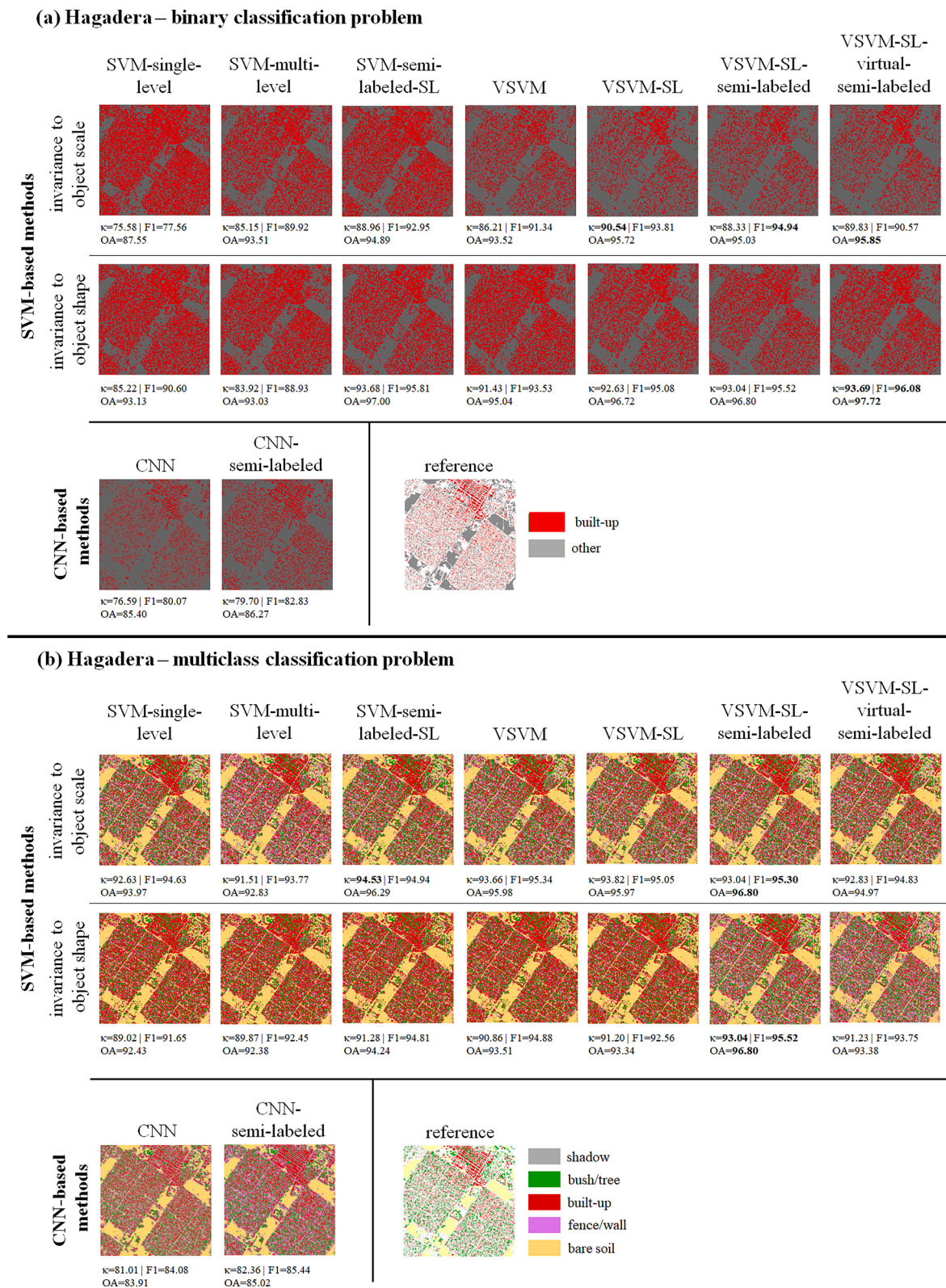


Fig. 10. Classification maps obtained with the various models for the binary (a) and multiclass (b) classification problem, respectively.

the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgements

This work is part of the research project RIESGOS 2.0 (03G0905A-B), which is funded by the German Federal Ministry of Education and Research (BMBF). We thank Simonas Garsva for initial test runs during the seminar “remote sensing” @ Technical University of Munich in



2019/2020.

## References

- Aravena Pelizari, P., Spröhnle, K., Geiß, C., Schoepfer, E., Plank, S., Taubenböck, H., 2018. Multi-sensor feature fusion for very high spatial resolution built-up area extraction in temporary settlements. *Remote Sens. Environ.* 209, 793–807. <https://doi.org/10.1016/j.rse.2018.02.025>.
- Baatz, M., Schäpe, A., 2000. Multiresolution Segmentation—An Optimization Approach for High Quality Multi-Scale Image Segmentation, ser. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), *Angewandte Geographische Informations-Verarbeitung XII*. Herbert Wichmann Verlag, Karlsruhe, Germany, pp. 12–23.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8), 1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>.
- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65, 2–16. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>.
- Bruzzzone, L., Carlini, L., 2006. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Trans. Geosci. Remote Sens.* 44 (9), 2587–2600. <https://doi.org/10.1109/TGRS.2006.875360>.
- Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2, 121–167.
- Camps-Valls, G., Tuia, D., Bruzzzone, L., Benediktsson, J.A., 2014. Advances in Hyperspectral Image Classification: Earth monitoring with statistical learning methods. *IEEE Signal Process. Mag.* 31 (1), 45–54. <https://doi.org/10.1109/MSP.2013.2279179>.
- Chen, J., Sun, B., Wang, L., Fang, B., Chang, Y., Li, Y., Zhang, J., Lyu, X., Chen, G., 2022. Semi-supervised semantic segmentation framework with pseudo supervisions for land-use/land-cover mapping in coastal areas. *Int. J. Appl. Earth Obs. Geoinf.* 112, 102881 <https://doi.org/10.1016/j.jag.2022.102881>.
- DeCoste, D., Schölkopf, B., 2002. Training invariant support vector machines. *Mach. Learn.* 46 (1–3), 161–190.
- Demir, B., Persello, C., Bruzzzone, L., 2011. Batch-Mode Active-Learning Methods for the Interactive Classification of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* 49 (3), 1014–1031. <https://doi.org/10.1109/TGRS.2010.2072929>.
- Dópido, I., Li, J., Marpu, P.R., Plaza, A., Bioucas Dias, J.M., Benediktsson, J.A., 2013. Semisupervised Self-Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 51 (7), 4032–4044. <https://doi.org/10.1109/TGRS.2012.2228275>.
- Foody, G. M., 2009. On Training and Evaluation of SVM for Remote Sensing Applications. Camps-Valls, G., Bruzzzone, L. (Eds.), *Kernel Methods for Remote Sensing Data Analysis*. John Wiley & Sons, Ltd, Chichester, UK, 85–109, 2009. 10.1002/9780470748992.ch4.
- Geiß, C., Jilge, M., Lakes, T., Taubenböck, H., 2016a. Estimation of Seismic Vulnerability Levels of Urban Structures with Multisensor Remote Sensing. *IEEE J-STARS* 9 (5), 1913–1936. <https://doi.org/10.1109/JSTARS.2015.2442584>.
- Geiß, C., Klotz, M., Schmitt, A., Taubenböck, H., 2016b. Object-based Morphological Profiles for Classification of Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* 54 (10), 5952–5963. <https://doi.org/10.1109/TGRS.2016.2576978>.
- Geiß, C., Aravena Pelizari, P., Schrade, H., Brenning, A., Taubenböck, H., 2017a. On the Effect of Spatially Non-disjoint Training and Test Samples on Estimated Model Generalization Capabilities in Supervised Classification with Spatial Features. *IEEE Geosci. Remote Sens. Lett.* 14 (11), 2008–2012. <https://doi.org/10.1109/LGRS.2017.2747222>.
- Geiß, C., Aravena Pelizari, P., Blickensdörfer, L., Taubenböck, H., 2019. Virtual Support Vector Machines with Self-Learning Strategy for Classification of Multispectral Remote Sensing Imagery. *ISPRS J. Photogramm. Remote Sens.* 151, 42–58. <https://doi.org/10.1016/j.isprsjprs.2019.03.001>.
- Geiß, C., Brzoska, E., Aravena Pelizari, P., Lautenbach, S., Taubenböck, H., 2022b. Multi-target Regressor Chains with Repetitive Permutation Scheme for Characterization of Built Environments with Remote Sensing. *Int. J. Appl. Earth Obs. Geoinf.* 106, 102657 <https://doi.org/10.1016/j.jag.2021.102657>.
- Geiß, C., Taubenböck, H., 2015. Object-based postclassification relearning. *IEEE Geosci. Remote Sens. Lett.* 12 (11), 2336–2340. <https://doi.org/10.1109/LGRS.2015.2477436>.
- Geiß, C., Thoma, M., Pittore, M., Wieland, M., Dech, S., Taubenböck, H., 2017b. Multitask Active Learning for Characterization of Built Environments with Multisensor Earth Observation Data. *IEEE J-STARS* 10 (12), 5583–5597. <https://doi.org/10.1109/JSTARS.2017.2748339>.
- Geiß, C., Thoma, M., Taubenböck, H., 2018. Cost-sensitive Multitask Active Learning for Characterization of Urban Environments with Remote Sensing. *IEEE Geosci. Remote Sens. Lett.* 15 (6), 922–926. <https://doi.org/10.1109/LGRS.2018.2813436>.
- Geiß, C., Zhu, Y., Qiu, C., Mou, L., Zhu, X.X., Taubenböck, H., 2022a. Deep Relearning in the Geospatial Domain for Semantic Remote Sensing Image Segmentation. *IEEE Geosci. Remote Sens. Lett.* 19, 8002705. <https://doi.org/10.1109/LGRS.2020.3031339>.
- Guo, Y., Fu, Y.H., Chen, S., Bryant, C.R., Li, X., Senthilnath, J., Sun, H., Wang, S., Wu, Z., de Beurs, K., 2021. Integrating spectral and textual information for identifying the tasseling date of summer maize using UAV based RGB images. *Int. J. Appl. Earth Obs. Geoinf.* 102, 102435 <https://doi.org/10.1016/j.jag.2021.102435>.
- Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* SMC-3:610–621. 10.1109/TSMC.1973.4309314.
- Izquierdo-Verdiguier, E., Laparra, V., Gómez-Chova, L., Camps-Valls, G., 2013. Encoding Invariances in Remote Sensing Image Classification with SVM. *IEEE Geosci. Remote Sens. Lett.* 10 (5), 981–985. <https://doi.org/10.1109/LGRS.2012.2227297>.
- Jaiswal, A., Ramesh Babu, A., Zaki Zadeh, M., Banerjee, D., Makedon, F., 2021. A Survey on Contrastive Self-Supervised Learning. *Technologies* 9, 2. <https://doi.org/10.3390/technologies9010002>.
- Li, J., Huang, X., Gong, J., 2019. Deep neural network for remote-sensing image interpretation: Status and perspectives. *Nat. Sci. Rev.* 6 (6), 1082–1086. <https://doi.org/10.1093/nsr/nwz058>.
- Li, Y.-F., Zhou, Z.-H., 2015. Towards Making Unlabeled Data Never Hurt. *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (1), 175–188. <https://doi.org/10.1109/TPAMI.2014.2299812>.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 3431–3440. 10.48550/arXiv.1411.4038.
- Lu, X., Zhang, J., Li, T., Zhang, Y., 2016. A Novel Synergetic Classification Approach for Hyperspectral and Panchromatic Images Based on Self-Learning. *IEEE Trans. Geosci. Remote Sens.* 54 (8), 4917–4928. <https://doi.org/10.1109/TGRS.2016.2553047>.
- Munoz-Mari, J., Tuia, D., Camps-Valls, G., 2012. Semisupervised classification of remote sensing images with active queries. *IEEE Trans. Geosci. Remote Sens.* 50 (10), 3751–3763. <https://doi.org/10.1109/TGRS.2012.2185504>.
- Pan, C., Li, J., Wang, Y., Gao, X., 2018. Collaborative learning for hyperspectral image classification. *Neurocomputing* 275, 2512–2524. <https://doi.org/10.1016/j.neucom.2017.11.035>.
- Persello, C., Bruzzzone, L., 2014. Active and Semisupervised Learning for the Classification of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* 52 (11), 6937–6956. <https://doi.org/10.1109/TGRS.2014.2305805>.
- Persello, C., Boularias, A., Dalponte, M., Gobakken, T., Naesset, E., Schölkopf, B., 2014. Cost-Sensitive Active Learning With Lookahead: Optimizing Field Surveys for Remote Sensing Data Classification. *IEEE Trans. Geosci. Remote Sens.* 52 (10), 6652–6664. <https://doi.org/10.1109/TGRS.2014.2300189>.
- Samat, A., Gamba, P., Du, P., Luo, J., 2015. Active extreme learning machines for quad-polarimetric SAR imagery classification. *Int. J. Appl. Earth Obs. Geoinf.* 35, 305–319. <https://doi.org/10.1016/j.jag.2014.09.019>.
- Schölkopf, B., Smola, A., 2002. *Learning With Kernels—Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA: MIT Press.
- Shu, Q., Pan, J., Zhang, Z., Wang, M., 2022. MTCNet: Multitask consistency network with single temporal supervision for semi-supervised building change detection. *Int. J. Appl. Earth Obs. Geoinf.* 115, 103110 <https://doi.org/10.1016/j.jag.2022.103110>.
- Trimble, 2014. eCognition developer 9.0 reference book. München: Germany Trimble Documentation.
- Tuia, D., Ratle, F., Pacifici, F., Kanevski, M.F., Emery, W.J., 2009. Active Learning Methods for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* 47 (7), 2218–2232. <https://doi.org/10.1109/TGRS.2008.2010404>.
- Tuia, D., Copa, L., Kanevski, M., Munoz-Mari, J., 2011. A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. *IEEE J. Sel. Top. Signal Process.* 5 (3), 606–617. <https://doi.org/10.1109/JSTSP.2011.2139193>.
- UNHCR, Hagadera Refugee Camp Overview as of January 2012. Geographic Information Systems and Mapping Unit, UNHCR Regional Support Hub, Nairobi, 2012. Available online: <https://data2.unhcr.org/en/documents/download/31535>.
- Volpi, M., Tuia, D., Bovolo, F., Kanevski, M., Bruzzzone, L., 2013. Supervised change detection in VHR images using contextual information and support vector machines. *Int. J. Appl. Earth Obs. Geoinf.* 20, 77–85. <https://doi.org/10.1016/j.jag.2011.10.013>.
- Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M., 2020. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv.* 53 (3), 1–34. <https://doi.org/10.1145/3386252>.
- Zhu, Y., Geiß, C., So, E., Jin, Y., 2021. Multi-temporal Relearning with Convolutional LSTM Models for Land Use Classification. *IEEE J-STARS* 14, 3251–3265. <https://doi.org/10.1109/JSTARS.2021.3055784>.