



## Article

# Active-Learning Approaches for Landslide Mapping Using Support Vector Machines

Zhihao Wang \* and Alexander Brenning

Department of Geography, Friedrich Schiller University Jena, Loebdergraben 32, 07743 Jena, Germany; alexander.brenning@uni-jena.de

\* Correspondence: zhihao.wang@uni-jena.de

**Abstract:** Ex post landslide mapping for emergency response and ex ante landslide susceptibility modelling for hazard mitigation are two important application scenarios that require the development of accurate, yet cost-effective spatial landslide models. However, the manual labelling of instances for training machine learning models is time-consuming given the data requirements of flexible data-driven algorithms and the small percentage of area covered by landslides. Active learning aims to reduce labelling costs by selecting more informative instances. In this study, two common active-learning strategies, uncertainty sampling and query by committee, are combined with the support vector machine (SVM), a state-of-the-art machine-learning technique, in a landslide mapping case study in order to assess their possible benefits compared to simple random sampling of training locations. By selecting more “informative” instances, the SVMs with active learning based on uncertainty sampling outperformed both random sampling and query-by-committee strategies when considering mean AUROC (area under the receiver operating characteristic curve) as performance measure. Uncertainty sampling also produced more stable performances with a smaller AUROC standard deviation across repetitions. In conclusion, under limited data conditions, uncertainty sampling reduces the amount of expert time needed by selecting more informative instances for SVM training. We therefore recommend incorporating active learning with uncertainty sampling into interactive landslide modelling workflows, especially in emergency response settings, but also in landslide susceptibility modelling.

**Keywords:** active learning; landslide modelling; support vector machine; machine learning



**Citation:** Wang, Z.; Brenning, A. Active-Learning Approaches for Landslide Mapping Using Support Vector Machines. *Remote Sens.* **2021**, *13*, 2588. <https://doi.org/10.3390/rs13132588>

Academic Editors: Stephen Grebby and Stuart Marsh

Received: 3 May 2021

Accepted: 29 June 2021

Published: 1 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Despite significant progress in landslide hazard assessment and mitigation, these hazards still present a major challenge for policymakers to reduce monetary losses and casualties. The occurrence probability of landslides, which broadly include a large variety of downslope movement processes on hillslopes under the effects of gravity [1], varies greatly in space and time as a result of complex patterns of predisposing factors and temporal variation in triggering factors. Considering the ongoing global trends of urbanization, deforestation, and climate change, landslide science faces the growing challenge of having to update landslide hazard assessments and provide rapid post-disaster information in the event of regional triggering events such as rainstorms and earthquakes [2–4]. For example, an earthquake in Tomakomai, Japan triggered about 10,000 landslides causing 36 deaths [3], and in 2018, landslides triggered by seasonal heavy precipitation caused approximately 105 deaths and USD 212 million in losses in China [5]. In Italy, as in many other regions worldwide, landslides are mostly triggered by intense or prolonged rainfall [6]. These hazards often cause long-term economic loss, population displacement, and negative effects on the natural environment.

Landslide mapping refers to the manual or automated detection and delineation of actual landslides that are appreciable in remote-sensing imagery or based on their topographic footprint [7–9]. Additionally, the growing availability of light detection and

ranging (LiDAR) derived high-resolution digital terrain models (HRDTM) allows us to detect landslides where passive optical sensors are limited (e.g., within the forest) [10,11]. This classification task is related to landslide susceptibility mapping, which focuses on estimating the probability of future landslide occurrences based on predisposing factors: usually topographic, geological, and land use/land cover conditions. Evidently, factors that control susceptibility can also provide valuable information for landslide mapping [12], which also requires post-event remote-sensing data (e.g., optical or LiDAR). Conversely, landslide inventories created by means of landslide mapping are a necessary input for landslide susceptibility mapping using supervised classification models. Together, landslide mapping and susceptibility modelling play a critical role in providing information that is necessary for decision making in emergency situations and for reducing risk in the development of spatial planning strategies.

Machine-learning techniques are increasingly being adopted in landslide modelling, as they have the potential to better adapt to complex nonlinear Earth surface processes and their interactions with land use than parametric statistical techniques such as logistic regression. Examples for black-box machine-learning models include the SVM, artificial neural networks, and random forests, whereas the generalized additive model (GAM) as an intermediate-complexity model is popular due to its nonlinear but more interpretable structure [13–16]. However, these data-driven supervised learning algorithms need a large number (e.g., thousands) of observations of landslide presence/absence, which are usually derived from manually digitized landslide inventories. Creating landslide inventories and more generally the manual labelling or annotation of these instances is a very time-consuming task, which increases the cost of landslide modelling studies and leads to delays in post-disaster situation awareness.

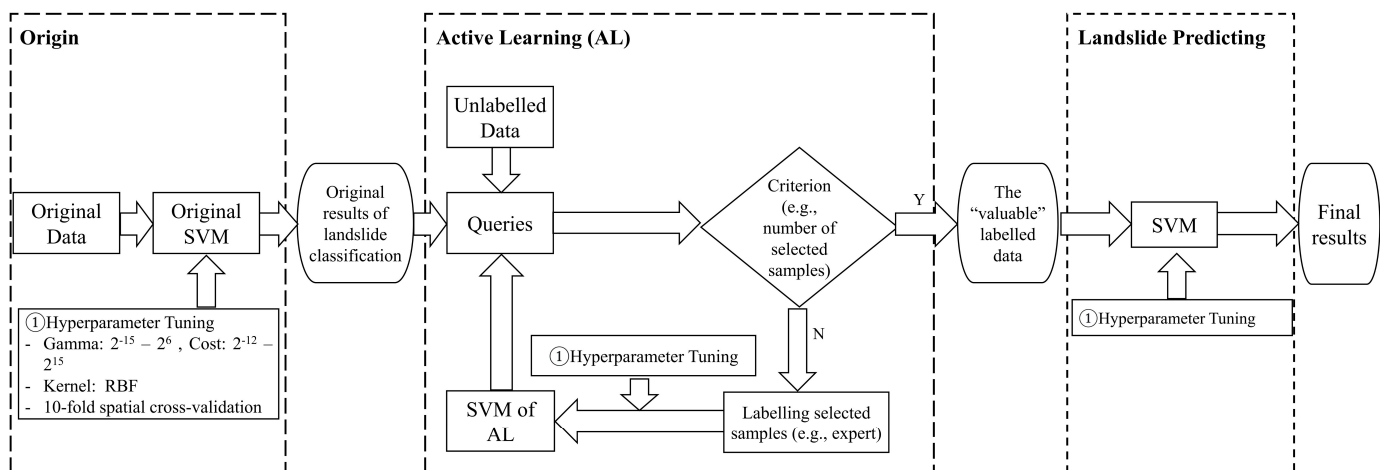
Active learning (AL) is a framework that promises to reduce this burden by selecting “informative” instances for the user to label [17,18]. In each of these queries to the user, an additional small batch of unlabelled instances (e.g., grid cells) is selected based on an informativeness measure and then presented to an “oracle” (i.e., a human annotator) for labelling. Active learning aims to achieve better accuracies using as few labelled instances as possible, thereby minimizing the cost of obtaining labelled data. Active learning has increasingly been adopted in remote-sensing classification [19–21] but has rarely been adopted in the context of landslide mapping [22].

The SVM has become increasingly popular in the context of landslide modelling along with other nonlinear techniques such as the generalized additive model [15,23,24]. Compared to the less flexible GAM, the SVM is capable of modelling nonlinear interactions among predictors while avoiding overfitting through regularization. Therefore, active learning based on the SVM for predicting landslides was adopted in this paper.

Hence, the main objective of this study is to assess the potential of different active-learning strategies for landslide mapping based on limited amounts of labelled data. We consider two popular active-learning query strategies in combination with the SVM in a case study from the Andes of southern Ecuador [25].

## 2. Materials and Methods

In this study, we use active-learning strategies [26] to sample “interesting” (i.e., informative) locations for training SVM models for landslide detection, and we compare this approach to a simple random sampling strategy (Figure 1). In active learning, a small training data set is initially retrieved to obtain a preliminary SVM fit. This model’s classifications then allow us to identify relevant additional instances that have the greatest potential to improve the model fit. These instances are then labelled, and the SVM is retrained with the additional training data. This step is repeated to investigate changes in model performance with increasing instance size. Hyperparameters are tuned in each individual step, and model performances are estimated in identical test sets to ensure comparability. Details of this procedure are explained in the following sections, and Settles (2010) [26] provides a detailed overview of active-learning strategies.



**Figure 1.** Overview of the workflow of active learning for landslide detection based on the SVM.

### 2.1. Active-Learning and Traditional Learning Strategies Used

Active learning is a subfield of machine learning that is also referred to as query learning or optimal experimental design in the statistical literature. It is a very broad field encompassing many different approaches [26–30]. In general, the purpose of active learning is to require as few labelled instances as possible while achieving a level of high accuracy. After starting with a small initial training set, small batches of additional “informative” instances are presented to an expert for labelling.

There are three different settings of active learning, namely membership query synthesis, stream-based selective sampling, and pool-based sampling [26,27]. For our study area, we have a large collection of unlabelled data. Due to uncertainty in the random generation process of instances using membership query synthesis [31] and the high cost of labelling selected instances one by one using stream-based selective sampling, pool-based sampling was adopted in this study. Pool-based sampling can be further divided into two main categories: uncertainty sampling and query by committee [19].

We denote the unlabelled data set as  $x$  and the classes of the unlabelled data set as  $y$ .  $P_\theta$  is the posterior probability as estimated by the current model  $\theta$  in a given active-learning step.

#### 2.1.1. Uncertainty Sampling

Uncertainty sampling chooses the instances that are predicted with the lowest confidence, i.e., that are associated with the greatest uncertainty in the current model [32]. It may be the most common and simplest active-learning approach. We briefly present three popular uncertainty-sampling strategies, but we choose only margin sampling due to the mathematical equivalence of these approaches in two-class situations.

##### (1) Least Confidence

The least-confidence strategy for a sequence of models queries the instances for which the current model has the least confidence as the predicted classes are equally likely [33]. Therefore, the most “informative” instances are selected by

$$\hat{y} = \operatorname{argmax}_y P_\theta(y|x), \quad (1)$$

$$x_{LC}^* = \operatorname{argmax}_x (1 - P_\theta(\hat{y}|x)), \quad (2)$$

where  $\hat{y}$  is the most likely label, i.e., the class label with the highest posterior probability under the current model  $\theta$ .  $x_{LC}^*$  represents the instance that the current model  $\theta$  is most likely to mislabel.

##### (2) Margin Sampling (MS)

Margin sampling (MS) was proposed to additionally take advantage of information regarding the posterior probabilities of all of the labels, not only the most likely one [34]. Here, instances with the smallest margin between the posterior of the first and the second most likely labels are selected. Since these are more ambiguous, the model has difficulty in differentiating between the two most likely class labels. Hence, knowing the true label would help the classifier to discriminate them more effectively. The instances are selected using

$$x_{MS}^* = \operatorname{argmin}_x (P_{\theta}(\hat{y}_1|x) - P_{\theta}(\hat{y}_2|x)), \quad (3)$$

where  $\hat{y}_1$  and  $\hat{y}_2$  are the first and the second most likely labels, respectively.

### (3) Entropy measure

One of the more general pool-based sampling strategies is based on the entropy measure [35]. This strategy uses entropy, which is an information-theoretic measure of uncertainty of a random variable. It aims at using information from all of the remaining classes to detect the most informative instances. Intuitively, the entropy measure strategy should perform better than the least-confidence and MS strategies, especially for very large label sets. Instances are selected using

$$E(y, x) = - \sum_{i=1}^k P_{\theta}(y_i|x) \log P_{\theta}(y_i|x), \quad (4)$$

$$x_{EP}^* = \operatorname{argmax}_x E(y, x), \quad (5)$$

where  $E(y, x)$  is the entropy value of class  $y$  for instance  $x$ . Instances with the highest entropy value, which imply more uncertainty in the distribution, are selected as  $x_{EP}^*$ .

Each of these uncertainty sampling strategies have their own application scenarios. In binary classification, however, all three are equivalent in selecting instances with the posterior class probabilities closest to 0.5 [26]. We implemented uncertainty sampling using the equation given for margin sampling, and we therefore refer to it as margin sampling in the rest of the paper.

#### 2.1.2. Query by Committee

Query by committee (QBC) is another more theoretically motivated active learning algorithm that selects informative unlabelled instances based on different models (a committee) trained on the current labelled training set [26,36]. Based on the posterior probabilities predicted by the different committee members, the unlabelled instances with the maximum disagreement are selected. Two important measures of disagreement among the models are the Kullback–Leibler (KL) divergence [37,38] and vote entropy [39]. Because KL divergence applies several independent justifications and calculates the average difference between the label distributions of any committee, it is considered the better approach to selecting informative instances [30]. Hence, KL divergence was adopted in this study. Let  $C = \{\theta(1), \dots, \theta(C)\}$  denotes the set of models forming the committee. Then

$$P_C(y_i|x) = \frac{1}{|C|} \sum_{c=1}^C P_{\theta(c)}(y_i|x), \quad (6)$$

corresponds to the committee's average posterior probability of class label  $y_i$ , and

$$D(P_{\theta(c)}||P_C) = \sum_i P_{\theta(c)}(y_i|x) \log(P_{\theta(c)}(y_i|x)/P_C(y_i|x)), \quad (7)$$

denotes the KL divergence, which we try to maximize on average over all committee members:

$$x_{KL}^* = \operatorname{argmax}_x \frac{1}{C} \sum_{c=1}^C D(P_{\theta(c)}||P_C), \quad (8)$$

This strategy focuses on the instances  $x_{KL}^*$  with the larger average difference with respect to the label distributions of any one committee.

Various strategies can be applied to creating a committee, such as bootstrap resampling of the training data [40,41]. We decided to set up a committee of SVM classifiers trained using different hyperparameter values since the behaviour of the SVM strongly varies with its cost and bandwidth parameters,  $C$  and  $\gamma$  [42]. In each active-learning round, we randomly sampled 250 pairs of hyperparameter values ( $\log_2 C$  between  $-12$  and  $+15$  and  $\log_2 \gamma$  between  $-12$  and  $+6$ ). The best-performing 25 hyperparameter settings were then selected to form a committee for that round.

### 2.1.3. Random Sampling as a Baseline

In addition to the active-learning strategies described above, simple random sampling (RS) was used for comparison. It randomly selects instances from the unlabelled data with equal probability and does not try to assess the utility of the data for landslide mapping.

## 2.2. Landslide Classification Model

In this study we used a support vector machine (SVM) model, which is a flexible supervised machine-learning technique [43]. It has previously shown competitive performances in landslide modelling [12,15,44,45]. This technique is particularly appealing in active learning because its flexibility can be tuned extremely well, allowing it to transition from a strongly penalized simple model to a more complex one as the sample size grows larger. The SVM can be applied in both one-class and two-class cases. Yao, et.al. (2008) compared one-class and two-class SVM on landslide analysis and concluded that two-class SVM could have better prediction efficiency than one-class SVM [46]. Therefore, in this study, we applied two-class SVM as the active-learning classifier.

Because the flexibility of the SVM is controlled by its hyperparameters  $\gamma$  (bandwidth) and  $C$  (cost), a  $k$ -fold cross-validation was used in each active-learning iteration to optimize them [47]. In this cross-validation, the training set is split into  $k$  equally sized partitions, one of which is retained for testing the model, and the remaining  $k - 1$  partitions are used as training data. This process is repeated  $k$  times, and every partition is used once as the validation data. Performance estimates are averaged over the  $k$  partitions to obtain a cross-validation estimate of the performance measure. We used  $k = 10$ , which is a commonly used setting. Given the spatial nature of our data, we used  $k$ -means clustering of the sample coordinates to generate spatial cross-validation partitions [48]. The choice of the SVM kernel function is less critical; therefore, the radial basis function kernel was adopted.

The area under the receiver operating characteristic (ROC) curve [49,50] (AUROC) [51] was used as a measure of predictive performance. Its range is between 0.5 (no predictive skill) and 1 (perfect separation).

### 2.3. Repetition and Performance Estimation

The workflow described above was repeated 150 times in order to eliminate the influence of random variability on our results. In active learning as well as random sampling, we draw an additional batch 25 points in each iteration or epoch after an initial random sample of 210 points in the first step. This initial data contains 10 landslide points and 200 non-landslide points, roughly representing the spatial landslide density within the study area. With the chosen batch size, on average, one additional landslide point will be drawn from the study area even in random sampling, given the 4% landslide density. We performed 50 iterations in order to observe the convergence of results for large instance sizes, although such large sample sizes (1460 labelled instances) are not of practical relevance.

The entire study area (87,223 non-landslide and 2569 landslide grid cells) as the target area for landslide mapping was used for analysing and comparing AUROC performances obtained by SVMs with active- (MS, QBC) and passive-learning strategies (RS). We ex-

tracted 55,887 non-landslide and 1663 landslide grid cells from the overall data set to serve as a pool of candidate grid cells from which we sampled the training data set.

In order to gain insight into the importance of predictors at different active-learning stages, we further calculated the permutation-based variable importance as a simple overall measure of predictive importance [52]. This was applied in a spatial cross-validation framework, i.e., by making predictions on spatially disjoint cross-validation test sets [53]. Accumulated local effects (ALE) [54] plots were further generated to visualize the shape of the relationships between important predictor variables and SVM landslide classifications.

All statistical analyses were conducted using the open-source statistical software R (version 3.6.3) [55] and its contributed packages “sperrorest” for spatial cross-validation [56], “e1071” for SVM modelling [57], “ROCR” for AUROC estimation [58], and “iml” for model interpretation [59]. The R package “RSAGA” [60] and the open-source GIS SAGA [61] were used for geodata processing.

#### 2.4. Study Area and Data

The study area of our case study is located in the Andes of Southern Ecuador in the Reserva Biológica San Francisco (RBSF). The RBSF is located between the provincial capitals of Loja and Zamora (3°58′30″ S, 79°4′25″ W) [62]. The slopes are steep (1st quartile of slope angle: 28.8°, median: 36.5°) and covered with evergreen lower and upper mountain rainforest [63]. The annual precipitation in the study area ranges from 2000 mm in the lower parts to more than 6000 mm between 2900 and 3100 m a.s.l. with nearly daily rainfall [64]. This area is characterized by a high frequency of landslide occurrences, which underlines the potential utility of active-learning techniques for generating event-based landslide inventories on demand with as little labelling as possible. In this study area, landslides are important ecosystem disturbances that trigger local vegetation successions and thus contribute to habitat complexity in this unique hotspot of plant biodiversity [65,66]. Landslide processes were previously studied in more detail with a focus on geomorphic process rates and the possible effects of human land use [25,67]. In this case study, we focus on the “natural” part of the RBSF study area of Muenchow et al. (2012). The dataset includes 178 landslides with a mean landslide size of 793 m<sup>2</sup>. We refer the reader to Muenchow et al. (2012) for further geomorphological detail and an analysis of landslide susceptibility [25].

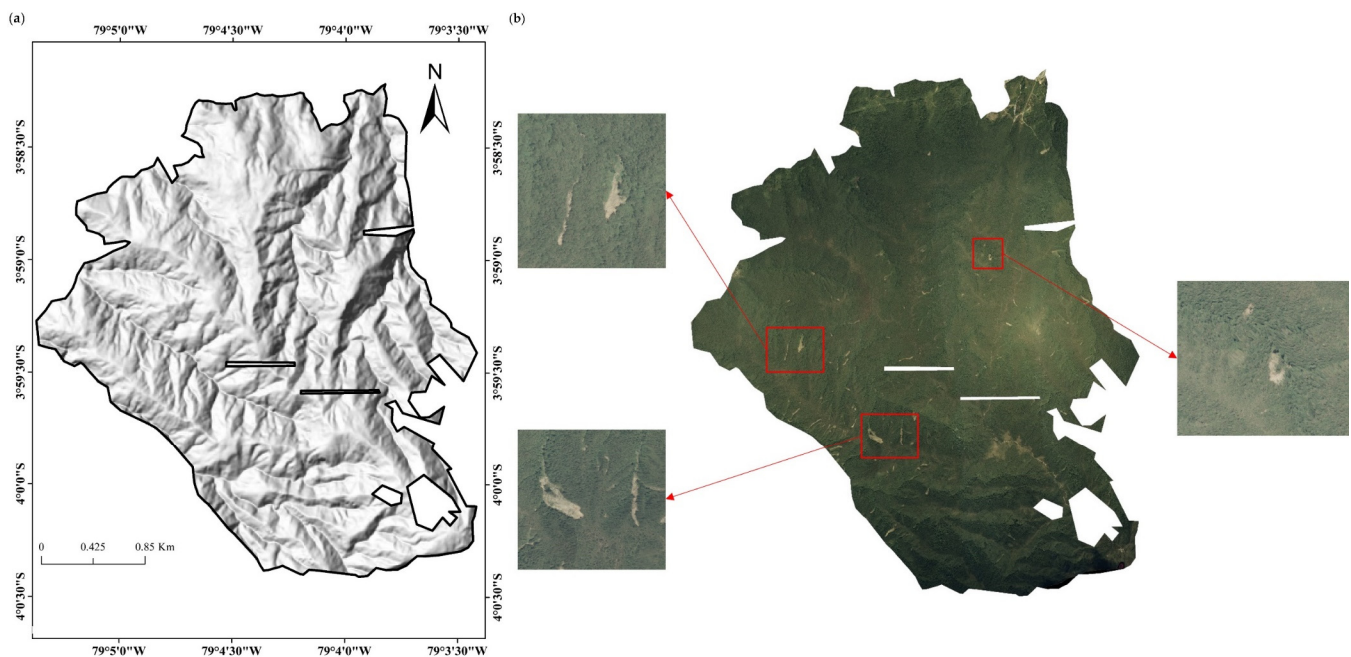
We used a high-resolution orthorectified aerial photograph of the study area as a direct optical indicator of vegetation disturbance by landslides. The image was acquired in 2001 and has a 0.3 m × 0.3 m spatial resolution (data source: E. Jordan and L. Ungerechts, Düsseldorf; DFG Research Unit FOR 816). Small cloudy patches and other errors were masked out manually (Figure 2). Landslides were mapped in this imagery by J. Muenchow (Erlangen), who analyzed the landslide distribution and characteristics in this study in more detail as part of a regional-scale comparison [25].

Vegetation indices (VIs) play an important role in mapping landslides and other forest disturbances [68,69]. Although the near-infrared part of the spectrum is particularly useful for identifying photosynthetically active, healthy plants, the imagery available for this case study is limited to the visible part of the spectrum, while having the benefit of offering the resolution required to detect the narrow, elongated landslides of this study area. Considering the spectral characteristics of the available orthophoto, we used the green chromatic coordinate (GCC) vegetation index [70], which has been shown to compare favourably to other indices in the visible part of the spectrum in distinguishing the forest from the soil [71]. The GCC is generally effective in suppressing the effects of changes in scene illumination [72], which is important in our mountainous study area. GCC is defined as

$$GCC = G / (R + B + G), \quad (9)$$

where R, G, and B represent the red, green, and blue bands of the ortho-photo. The red (RCC) and blue (BCC) chromatic coordinates are calculated in the same way. Because RCC

and BCC are strongly correlated (correlation coefficient: 0.95), we used GCC as a vegetation index as well as RCC.



**Figure 2.** The study area: the Reserva Biológica San Francisco (RBSF), (a) hillshade with  $10\text{ m} \times 10\text{ m}$  resolution in 1997, and (b) the orthophoto with  $0.3\text{ m} \times 0.3\text{ m}$  resolution in early 2001.

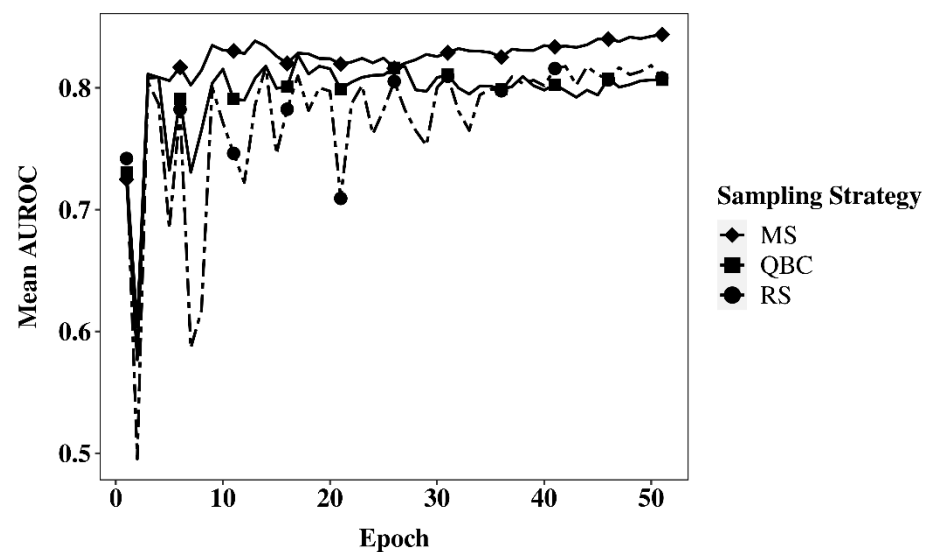
A digital elevation model (DEM) of the RBSF at a  $10\text{ m} \times 10\text{ m}$  resolution produced by E. Jordan and L. Ungerechts (Düsseldorf) was generated from stereo aerial photographs from the year 1997. Following Muenchow et al. (2012) [25], we derived the following terrain attributes from the DEM, as they are commonly included in landslide distribution models as preparatory factors: local slope angle (slope), plan and profile curvature (plancurv and profcurv), and the slope angle (cslope) and logarithm of the size of the upslope contributing area (log.carea) [15]. These terrain attributes are intended to act as proxies for destabilizing forces (slope, cslope), water availability (log.carea, concave curvatures), and exposure to wind (convex curvatures) as well as general variability in the characteristics of soil and vegetation [25]. Our expectation is that these terrain attributes will further improve landslide classification.

Overall, our feature set consisted of five terrain attributes and the GCC and RCC as remote-sensing variables. Predictors that presented outliers were winsorized at the 1st and 99th percentile.

### 3. Results

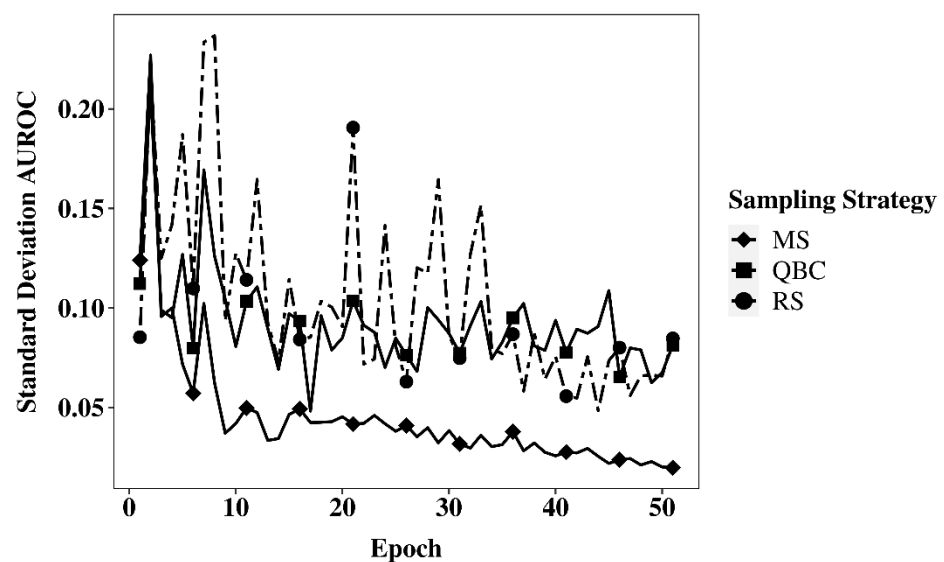
#### 3.1. Model Performance

Overall, active learning using margin sampling outperformed query by committee and random sampling after only four epochs, i.e., starting with a learning instance size of 310 grid cells (Figure 3). The SVM with MS increased continuously from this point, going from 0.80 in epoch 3 with only 285 grid cells to 0.83 after epoch 8, i.e., with label information for >410 grid cells. Mean AUROCs obtained with RS and QBC were very similar; they both reached  $\sim 0.79$  for only large sample sizes. The similar performances of QBC as an active-learning strategy and RS for passive learning suggest that the instances labelled in QBC-based active learning were no more informative than the ones retrieved with simple random sampling. Nevertheless, SVM performances with QBC were less variable than those completed with RS.



**Figure 3.** Performance versus epochs: mean AUROCs of SVMs across all 150 repetitions for different sampling methods in the target area.

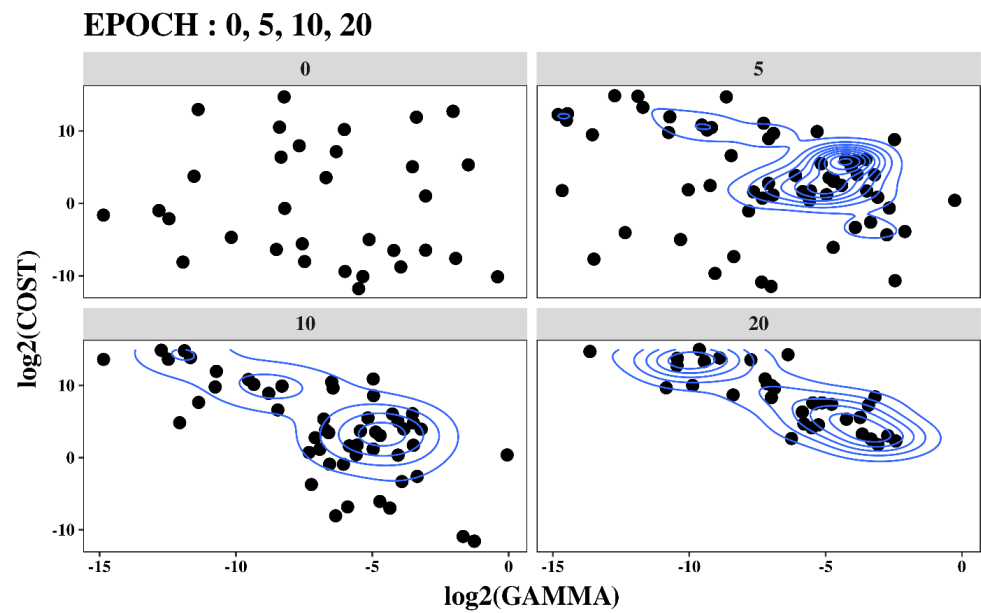
Similarly, the random variability of AUROC performances over the 150 repetitions revealed much less variable results for MS-based active learning than for QBC and RS, which show similar results (Figure 4). Differences in variability between MS and QBC/RS were at least twofold across all epochs, which indicates that informative instance data not only improves the performance but also reduces the probability of obtaining poor results due to random variability.



**Figure 4.** Standard deviation of AUROC estimates across all 150 repetitions for different sampling methods in the target area.

Considering the importance of the cost and gamma hyperparameters for the flexibility of the SVM, we examined the variability in optimal hyperparameters for different active-learning epochs in margin sampling. As the sample size increased, the optimal cost parameters across the repetitions were increasingly concentrated around the  $2^0$  to  $2^5$  region and the optimal  $\gamma$  to around  $2^{-5}$ , although the optimal region extended diagonally towards higher cost values when combined with larger  $\gamma$  values (Figure 5).

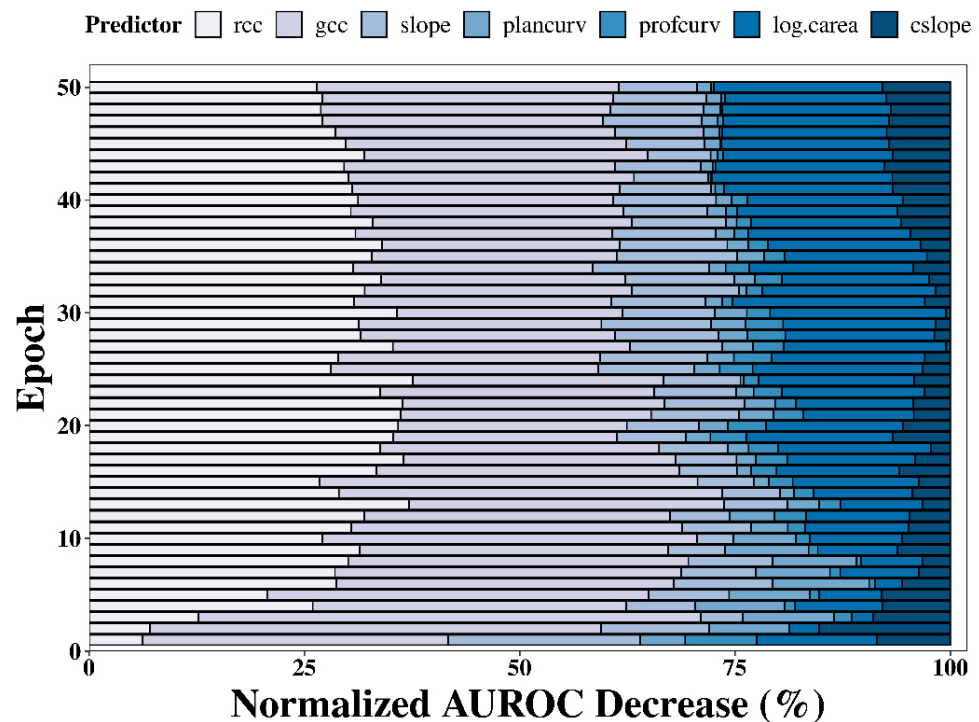




**Figure 5.** Optimal hyperparameters for the SVM with margin sampling in epochs 0, 5, 10, and 20 for all 150 repetitions.

### 3.2. Model Interpretation

A permutation-based variable importance assessment for the SVM with margin sampling revealed that the most important predictors were GCC and RCC, which was followed by a logarithm of the catchment area and catchment slope (Figure 6). Thus, predictors that are commonly used in landslide susceptibility modelling helped to improve the performance of models for landslide mapping consistently across all epochs. Note that there are moderate to strong correlations between the slope variables as well as between the upslope contributing area and the two curvature variables (Table 1).



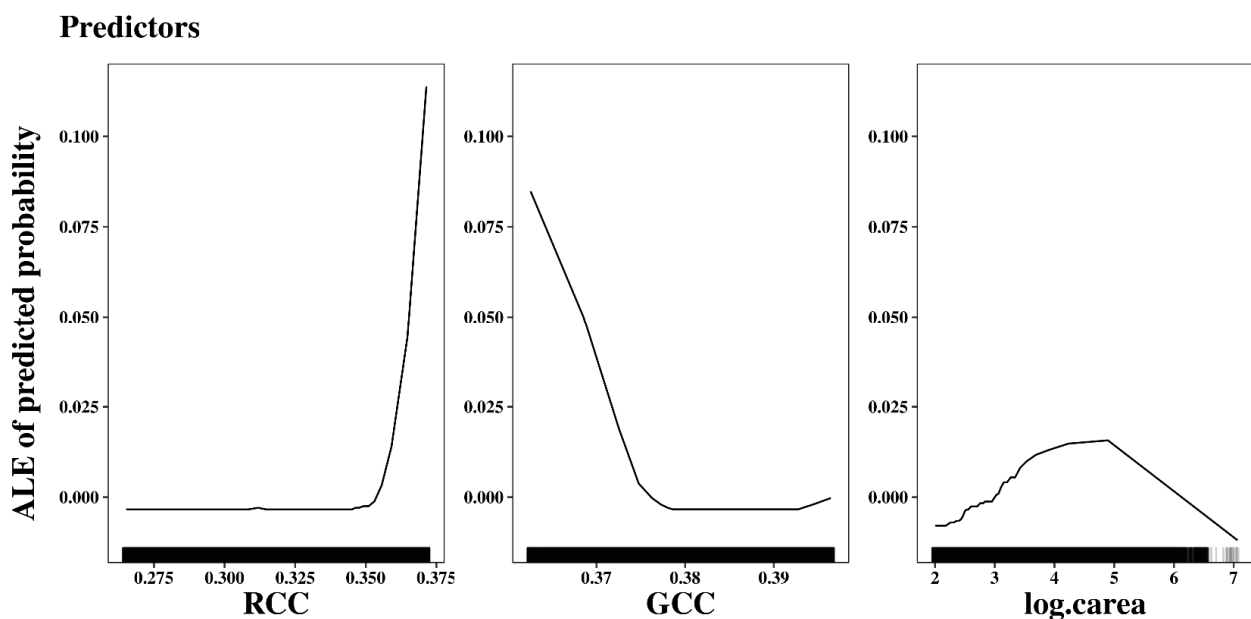
**Figure 6.** Variable importance plot for SVM using margin sampling in each epoch.

ALE plots for the 20th epoch of the 1st repetition display the averaged relationships between predictors and responses (Figure 7). Broadly speaking and as expected, landslides are primarily characterized by low vegetation vigour as represented by a low vegetation index, and to a smaller extent, by a steep upslope area. They are also rarely found in the valley bottoms, where the upslope contributing area is large, or directly on ridges or hilltops, where the upslope contributing area would be small. Ridges and hilltops often show reduced vegetation canopy due to factors other than landslides, such as windthrow, and the inclusion of the upslope contributing area therefore reduces confounding with these patterns, resulting in a geomorphologically more plausible classification.

**Table 1.** Correlations among the predictors (%).

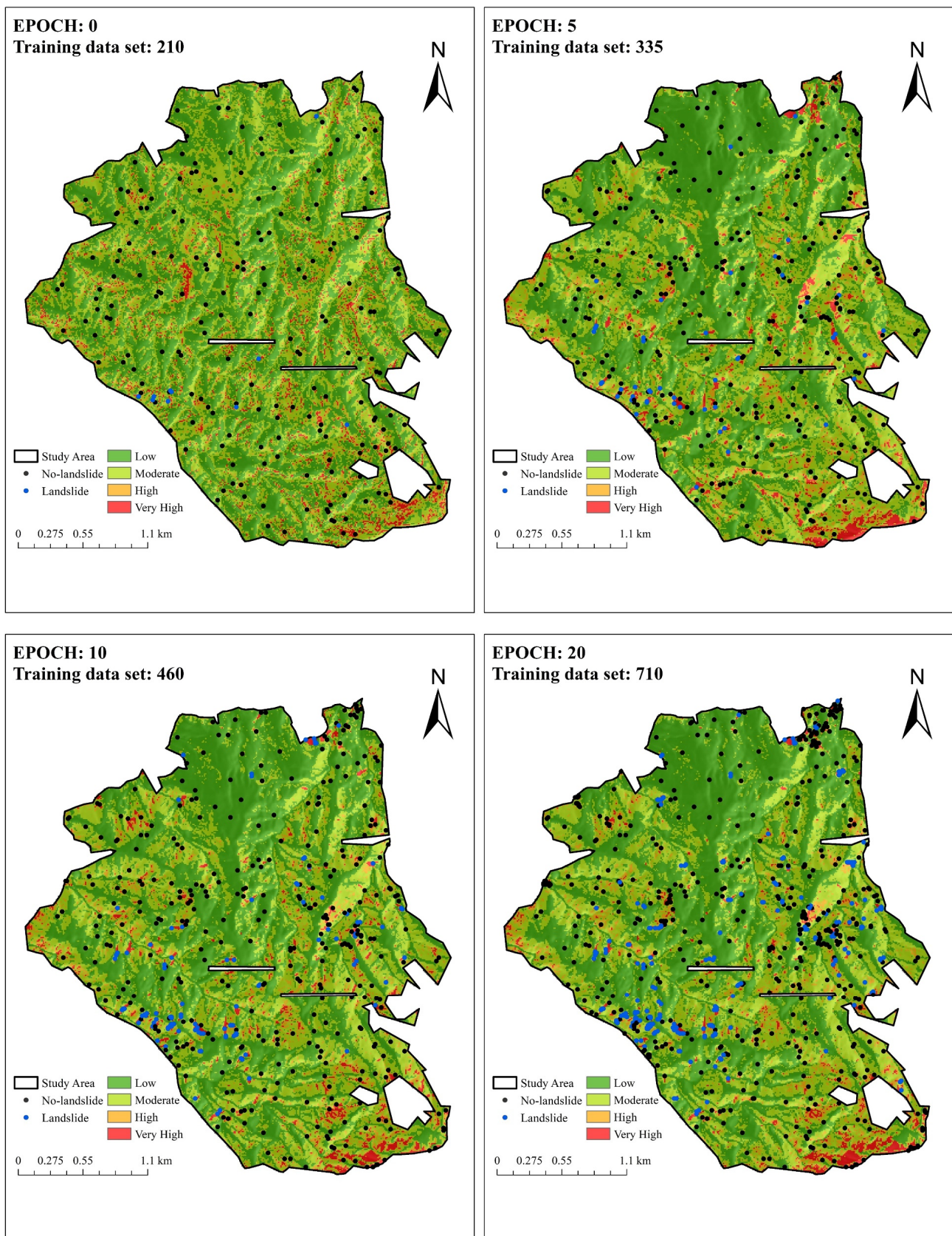
	RCC	GCC	Slope	Plancurv	Profcurv	Log.carea	Cslope
RCC	100	−36	−11	9	8	−16	−17
GCC	−36	100	14	−18	−18	34	28
Slope	−11	14	100	3	3	−3	75
plancurv	9	−18	1	100	52	−68	−17
profcurv	8	−18	3	52	100	−58	−23
log.carea	−16	34	−3	−68	−58	100	22
cslope	−17	28	75	−17	−23	22	100

Local slope angle (slope), plan and profile curvature (plancurv and profcurv), the slope angle (cslope) and logarithm of the size of the upslope contributing area (log.carea), and green chromatic coordinate (GCC) and red chromatic coordinate (RCC).



**Figure 7.** ALE plots of the most important predictors (green chromatic coordinate, GCC; red chromatic coordinate, RCC; logarithm of the size of the upslope contributing area, log.carea) for the SVM with margin sampling in repetition 1, epoch 20.

Landslide maps predicted by SVM with margin sampling clearly depict many of the landslide-affected areas even after only five epochs (Figure 8). There is little change in the spatial pattern of mapped landslides after more than five epochs, which is consistent with the relatively stable model performances and variable importance reported above. Despite the visual similarities between epochs 5 and 10, it should be remembered that quantitative performances in terms of the mean and standard deviation of AUROC did substantially improve from epoch 5 to epoch 10, as hyperparameter tuning started to stabilize as well around epoch 10.



**Figure 8.** Landslide classification maps of SVM with margin sampling in epochs 0, 5, 10, and 20. Predicted probabilities are classified into four classes (very high, high, moderate, and low) using the top 4th, 10th, and 50th percentile as class boundaries.

## 4. Discussion

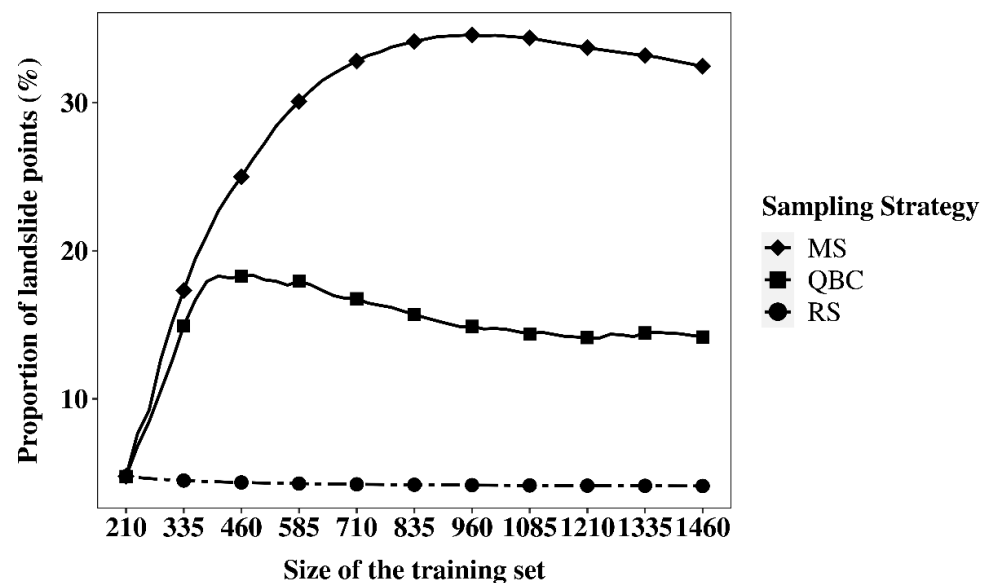
### 4.1. Potential of SVM with AL

Overall, our results confirm the potential of AL for remote-sensing applications and for landslide mapping in particular [19,22], and demonstrate the suitability of uncertainty sampling strategies. AL retrieves the data that it believes is more likely to be misclassified [26]. For the learning process of SVM, it builds a margin to classify the instances based on its features. If a new candidate point's distance from the margin is too small, this instance is more likely to be misclassified by the model and labelling these instances therefore has the greatest potential to improve model performance.

Landslide data is always imbalanced, which poses a particular challenge in classification modelling that can be addressed using uncertainty sampling. In this study, landslides covered only 4% of the study area, and consequently, random sampling is very poor at collecting information on positives. In contrast, AL strategies can reduce the impact of imbalance by retrieving more “useful” instances, including more positives (Table 2 and Figure 9). This is especially true for MS.

**Table 2.** Number of landslide and non-landslide instances in the training set after 50 epochs using different sampling strategies.

Sampling Strategy	Non-Landslide	Landslide
Margin sampling	1013	447
Query by committee	1280	180
Random sampling	1415	45



**Figure 9.** Number of landslide instances in the training set in each epoch using SVM with different sampling strategies.

### 4.2. Limitations of SVM with AL

A possible limitation of AL in the context of SVM classification is the interaction between the hyperparameters and the query strategy. Specifically, since AL queries depend on posterior probability, they are sensitive to SVM hyperparameters. If hyperparameters cannot be reliably estimated, as is the case in the initial epochs with small sample sizes, model performance can be highly variable and sometimes poor (Figures 3 and 4). When this occurs, AL query strategies may be close to random sampling, or they may even oversample irrelevant regions in the featured space [73,74]. The QBC strategy may be particularly sensitive to this problem since SVMs with 25 different hyperparameter settings were used to form a committee. In general, it can be difficult to strike the right balance between diversity

and goodness-of-fit of committee members, and the use of the top hyperparameter settings may not achieve an optimal committee. Although other strategies could be used, it was beyond the scope of this work to experiment with these additional design options. Xu et al. (2020) used three different pre-trained SVMs as committee members to conduct classifications and concluded that pre-trained SVMs made QBC more robust in iterative training [42]. Stumpf formed QBC committees using 500 fully grown trees from a random forest (RF) and concluded they can achieve good results [22], but no comparison with other, model-independent QBC approaches or with other query strategies was made. Considering these limitations and the positive results achieved with uncertainty sampling strategies, we suggest that the latter offers several advantages ranging from a simpler, model-agnostic implementation to fewer design decisions and reduced computational cost.

## 5. Conclusions

In this study, active-learning strategies for training landslide detection models outperformed models trained using randomly sampled data. The mean AUROCs of the SVM with margin sampling as an active-learning strategy was 0.80 with only 285 instances and 0.83 with 410 instances. In contrast, SVMs with query-by-committee and random sampling achieved AUROCs around 0.79 but only for large sample sizes. Meanwhile, the SVM with margin sampling was more robust than the other strategies. Therefore, uncertainty sampling is particularly promising as it achieved the best performance, was best able to handle imbalanced data, and is straightforward to implement regardless of the machine-learning model being used.

Labelling a large number of instances using human experts is a time-consuming process that cannot be executed in sufficient detail under time constraints, e.g., in emergency situations. Additionally, human experts cannot recognize which additional instances would be the most “useful” for predicting the response to, or in our case for identifying, landslides. Active-learning methods are therefore a promising strategy as part of an interactive landslide detection workflow, especially in an emergency response setting.

**Author Contributions:** Conceptualization, methodology, writing—original draft preparation, visualization, Z.W.; formal analysis, data curation, A.B. and Z.W.; writing—review and editing, supervision, A.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** Z.W. was funded through a China Scholarship Council PhD scholarship, which is gratefully acknowledged.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** We thank the DFG Research Unit FOR 816 (J. Bendix, Marburg) for providing the remote-sensing data created by E. Jordan and L. Ungerechts, Düsseldorf, and Raphael Knevels and José Cortés for their assistance with high-performance computing.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Highland, L.M.; Bobrowsky, P. *The Landslide Handbook—A Guide to Understanding Landslides*; Kidd, M., Ed.; U.S. Geological Survey Circular 1325: Reston, VA, USA, 2008; p. 129. ISBN 978-141-132-226-4.
2. Formetta, G.; Rago, V.; Capparelli, G.; Rigon, R.; Muto, F.; Versace, P. Integrated physically based system for modeling landslide susceptibility. *Procedia Earth Planet. Sci.* **2014**, *9*, 74–82. [[CrossRef](#)]
3. Aimaiti, Y.; Liu, W.; Yamazaki, F.; Maruyama, Y. Earthquake-induced landslide mapping for the 2018 Hokkaido eastern Iburi earthquake using PALSAR-2 data. *Remote Sens.* **2019**, *11*, 2351. [[CrossRef](#)]
4. Regmi, N.R.; Walter, J.I. Detailed mapping of shallow landslides in eastern Oklahoma and western Arkansas and potential triggering by Oklahoma earthquakes. *Geomorphology* **2020**, *366*, 106806. [[CrossRef](#)]
5. Fan, X.; Yang, F.; Subramanian, S.S.; Xu, Q.; Feng, Z.; Mavrouli, O.; Peng, M.; Ouyang, C.; Jansen, J.D.; Huang, R. Prediction of a multi-hazard chain by an integrated numerical simulation approach: The Baige landslide, Jinsha River, China. *Landslides* **2020**, *17*, 147–164. [[CrossRef](#)]

6. Peruccacci, S.; Brunetti, M.T.; Gariano, S.L.; Melillo, M.; Rossi, M.; Guzzetti, F. Rainfall thresholds for possible landslide occurrence in Italy. *Geomorphology* **2017**, *290*, 39–57. [[CrossRef](#)]
7. Lv, Z.Y.; Shi, W.Z.; Zhang, X.K.; Benediktsson, J.A. Landslide inventory mapping from bitemporal high-resolution remote sensing images using change detection and multiscale segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1520–1532. [[CrossRef](#)]
8. Kalantar, B.; Ueda, N.; Saeidi, V.; Ahmadi, K.; Halin, A.A.; Shabani, F. Landslide susceptibility mapping: Machine and ensemble learning based on remote sensing big data. *Remote Sens.* **2020**, *12*, 1737. [[CrossRef](#)]
9. Dao, D.V.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.C.; Moayed, H.; Phong, T.V.; Ly, H.B.; Le, T.T.; Trinh, P.T.; et al. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena* **2020**, *188*, 104451. [[CrossRef](#)]
10. Van Den Eeckhaut, M.; Kerle, N.; Poesen, J.; Hervás, J. Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data. *Geomorphology* **2012**, *173*, 30–42. [[CrossRef](#)]
11. Petschko, H.; Bell, R.; Glade, T. Effectiveness of visually analyzing LiDAR DTM derivatives for earth and debris slide inventory mapping for statistical susceptibility modeling. *Landslides* **2016**, *13*, 857–872. [[CrossRef](#)]
12. Knevels, R.; Petschko, H.; Leopold, P.; Brenning, A. Geographic object-based image analysis for automated landslide detection using open source GIS software. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 551. [[CrossRef](#)]
13. Brenning, A. Spatial prediction models for landslide hazards: Review, comparison and evaluation. *Nat. Hazards Earth Syst. Sci.* **2005**, *5*, 853–862. [[CrossRef](#)]
14. Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Alizadeh, M.; Chen, W.; Mohammadi, A.; Bin Ahmad, B.; Panahi, M.; Hong, H.Y.; et al. Landslide detection and susceptibility mapping by AIRSAR data using support vector machine and index of entropy models in Cameron highlands, Malaysia. *Remote Sens.* **2018**, *10*, 1527. [[CrossRef](#)]
15. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [[CrossRef](#)]
16. Pradhan, B.; Lee, S. Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Softw.* **2010**, *25*, 747–759. [[CrossRef](#)]
17. Huang, S.J.; Jin, R.; Zhou, Z.H. Active learning by querying informative and representative examples. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1936–1949. [[CrossRef](#)]
18. Bachman, P.; Sordani, A.; Trischler, A. Learning algorithms for active learning. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 301–310.
19. Demir, B.; Bovolo, F.; Bruzzone, L. Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1930–1941. [[CrossRef](#)]
20. Tuia, D.; Volpi, M.; Copa, L.; Kanevski, M.; Munoz-Mari, J. A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 606–617. [[CrossRef](#)]
21. Lin, J.Z.; Zhao, L.; Li, S.Y.; Ward, R.; Wang, Z.J. Active-learning-incorporated deep transfer learning for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4048–4062. [[CrossRef](#)]
22. Stumpf, A.; Lachiche, N.; Malet, J.P.; Kerle, N.; Puissant, A. Active learning in the spatial domain for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2492–2507. [[CrossRef](#)]
23. Shao, X.Y.; Ma, S.Y.; Xu, C.; Zhang, P.F.; Wen, B.Y.; Tian, Y.Y.; Zhou, Q.; Cui, Y.L. Planet image-based inventorying and machine learning-based susceptibility mapping for the landslides triggered by the 2018 Mw6.6 Tomakomai, Japan earthquake. *Remote Sens.* **2019**, *11*, 978. [[CrossRef](#)]
24. Peng, L.; Niu, R.Q.; Huang, B.; Wu, X.L.; Zhao, Y.N.; Ye, R.Q. Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the three gorges area, China. *Geomorphology* **2014**, *204*, 287–301. [[CrossRef](#)]
25. Muenchow, J.; Brenning, A.; Richter, M. Geomorphic process rates of landslides along a humidity gradient in the tropical Andes. *Geomorphology* **2012**, *139*, 271–284. [[CrossRef](#)]
26. Settles, B. *Active Learning Literature Survey*; Computer Sciences Technical Report 1648; University of Wisconsin: Madison, WI, USA, 2010.
27. Angluin, D. Queries and concept learning. *Mach. Learn.* **1988**, *2*, 319–342. [[CrossRef](#)]
28. Cohn, D.; Atlas, L.; Ladner, R. Improving generalization with active learning. *Mach. Learn.* **1994**, *15*, 201–221. [[CrossRef](#)]
29. Mackay, D.J.C. Information-based objective functions for active data selection. *Neural Comput.* **1992**, *4*, 590–604. [[CrossRef](#)]
30. Tong, S. Active Learning: Theory and Applications. Ph.D. Thesis, Stanford University, Stanford, CA, USA, August 2001.
31. Baum, E.B.; Lang, K. Query learning can work poorly when a human oracle is used. In Proceedings of the International Joint Conference on Neural Networks, Baltimore, MD, USA, 7–11 June 1992; p. 8.
32. Lewis, D.D.; Gale, W.A. A sequential algorithm for training text classifiers. In Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 1 July 1994; pp. 3–12.
33. Culotta, A.; McCallum, A. Reducing labeling effort for structured prediction tasks. In Proceedings of the 20th National Conference on Artificial Intelligence, Pittsburgh, PA, USA, 9–13 July 2005; pp. 746–751.
34. Scheffer, T.; Decomain, C.; Wrobel, S. Active hidden markov models for information extraction. In Proceedings of the International Symposium on Intelligent Data Analysis, Cascais, Portugal, 13–15 September 2001; pp. 309–318.

35. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
36. Seung, H.S.; Oppor, M.; Sompolinsky, H. Query by committee. In Proceedings of the 5th Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 287–294.
37. McCallum, A.K.; Nigam, K. Employing EM in pool-based active learning for text classification. In Proceedings of the 15th International Conference on Machine Learning, Madison, WI, USA, 24–27 July 1998; pp. 350–358.
38. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. Available online: <http://www.jstor.org/stable/2236703> (accessed on 23 April 2021). [[CrossRef](#)]
39. Dagan, I.; Engelson, S.P. Committee-based sampling for training probabilistic classifiers. In Proceedings of the 12th International Conference on Machine Learning, Tahoe, CA, USA, 9–12 July 1995; pp. 150–157.
40. Stańczyk, U.; Zielosko, B.; Jain, L.C. *Advances in Feature Selection for Data and Pattern Recognition*; Springer: Cham, Switzerland, 2018; p. 328. ISBN 978-3-319-67587-9.
41. Ramirez-Loaiza, M.E.; Sharma, M.; Kumar, G.; Bilgic, M. Active learning: An empirical study of common baselines. *Data Min. Knowl. Discov.* **2017**, *31*, 287–313. [[CrossRef](#)]
42. Xu, H.L.; Li, L.Y.; Guo, P.S. Semi-supervised active learning algorithm for SVMs based on QBC and tri-training. *J. Ambient Intell. Humaniz. Comput.* **2020**, 1–14. [[CrossRef](#)]
43. Vapnik, V. The support vector method of function estimation. In *Nonlinear Modeling*; Suykens, J.A.K., Vandewalle, J., Eds.; Springer: Boston, MA, USA, 1998; pp. 55–85.
44. Pawluszek, K.; Borkowski, A.; Tarolli, P. Sensitivity analysis of automatic landslide mapping: Numerical experiments towards the best solution. *Landslides* **2018**, *15*, 1851–1865. [[CrossRef](#)]
45. Dou, J.; Paudel, U.; Oguchi, T.; Uchiyama, S.; Hayakavva, Y.S. Shallow and Deep-Seated Landslide Differentiation Using Support Vector Machines: A Case Study of the Chuetsu Area, Japan. *Terr. Atmos. Ocean. Sci.* **2015**, *26*, 227–239. [[CrossRef](#)]
46. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [[CrossRef](#)]
47. Moguerza, J.M.; Munoz, A. Support vector machines with applications. *Stat. Sci.* **2006**, *21*, 322–336. [[CrossRef](#)]
48. Ruß, G.; Brenning, A. Data mining in precision agriculture: Management of spatial information. In Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Dortmund, Germany, 28 June–2 July 2010; pp. 350–359.
49. Begueria, S. Validation and evaluation of predictive models in hazard assessment and risk management. *Nat. Hazards* **2006**, *37*, 315–329. [[CrossRef](#)]
50. Frattini, P.; Crosta, G.; Carrara, A. Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* **2010**, *111*, 62–72. [[CrossRef](#)]
51. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 398, ISBN 978-0-470-58247-3.
52. Molnar, C. Interpretable Machine Learning. Available online: <https://christophm.github.io/interpretable-ml-book/> (accessed on 14 June 2021).
53. Ruß, G.; Brenning, A. Spatial variable importance assessment for yield prediction in precision agriculture. In Proceedings of the International Symposium on Intelligent Data Analysis, Tucson, AZ, USA, 19–21 May 2010; pp. 184–195.
54. Apley, D.W.; Zhu, J.Y. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B* **2020**, *82*, 1059–1086. [[CrossRef](#)]
55. R Core Team. *R: A Language and Environment for Statistical Computing*; R Version 3.6.3; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: <https://www.R-project.org/> (accessed on 30 June 2021).
56. Brenning, A. Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sprrorst. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 5372–5375.
57. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F.; Weingessel, A. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R Package Version 1.7-3. 2019. Available online: <https://CRAN.R-project.org/package=e1071> (accessed on 30 June 2021).
58. Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T. ROCR: Visualizing classifier performance in R. *Bioinformatics* **2009**, *21*, 3940–3941. [[CrossRef](#)]
59. Molnar, C.; Casalicchio, G.; Bischl, B. Iml: An R package for interpretable machine learning. *J. Open Source Softw.* **2018**, *3*, 786. [[CrossRef](#)]
60. Brenning, A.; Bangs, D.; Becker, M. RSAGA: SAGA Geoprocessing and Terrain Analysis. R package Version 1.3.0. 2018. Available online: <https://CRAN.R-project.org/package=RSAGA> (accessed on 30 June 2021).
61. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **2015**, *8*, 1991–2007. [[CrossRef](#)]
62. Beck, E.; Makeschin, F.; Haubrich, F.; Richter, M.; Bendix, J.; Valerezo, C. The Ecosystem (Reserva Biológica San Francisco). In *Gradients in a Tropical Mountain Ecosystem of Ecuador. Ecological Studies (Analysis and Synthesis)*, 198; Beck, E., Bendix, J., Kottke, I., Makeschin, F., Mosandl, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2008. [[CrossRef](#)]

63. Bussmann, R.W. The vegetation of Reserva Biológica San Francisco, Zamora–Chinchipec, southern Ecuador: A phytosociological synthesis. *Lyonia* **2003**, *3*, 145–254.
64. Emck, P. A Climatology of South Ecuador. with Special Focus on the Major Andean Ridge as Atlantic-Pacific Climate Divide. Ph.D. Thesis, University of Erlangen, Nuremberg, Germany, 2007.
65. Beck, E.; Bendix, J.; Kottke, I.; Makeschin, F.; Mosandl, R. *Gradients in a Tropical Mountain Ecosystem of Ecuador*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 198, p. 525. ISBN 978-3-540-73525-0.
66. Peters, T.; Diertl, K.H.; Gawlik, J.; Rankl, M.; Richter, M. Vascular plant diversity in natural and anthropogenic ecosystems in the Andes of southern Ecuador. *Mt. Res. Dev.* **2010**, *30*, 344–352. [[CrossRef](#)]
67. Brenning, A.; Schwinn, M.; Ruiz-Paez, A.P.; Muenchow, J. Landslide susceptibility near highways is increased by 1 order of magnitude in the Andes of southern Ecuador, Loja province. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 45–57. [[CrossRef](#)]
68. Mwaniki, M.; Möller, M.; Schellmann, G. Landslide inventory using knowledge based multi-sources classification time series mapping: A case study of central region of Kenya. *GL Forum* **2015**, *2015*, 209–219. [[CrossRef](#)]
69. Fernández, T.; Jiménez, J.; Fernández, P.; El Hamdouni, R.; Cardenal, F.; Delgado, J.; Irigaray, C.; Chacón, J. Automatic detection of landslide features with remote sensing techniques in the Betic Cordilleras (Granada, southern Spain). *Int. Soc. Photogramme* **2008**, *37*, 351–356.
70. Gillespie, A.R.; Kahle, A.B.; Walker, R.E. Color enhancement of highly correlated images. 2. Channel ratio and chromaticity transformation techniques. *Remote Sens. Environ.* **1987**, *22*, 343–365. [[CrossRef](#)]
71. Larrinaga, A.R.; Brotons, L. Greenness indices from a low-cost UAV imagery as tools for monitoring post-fire forest recovery. *Drones* **2019**, *3*, 6. [[CrossRef](#)]
72. Sonnentag, O.; Hufkens, K.; Teshera-Sterne, C.; Young, A.M.; Friedl, M.; Braswell, B.H.; Milliman, T.; O’Keefe, J.; Richardson, A.D. Digital repeat photography for phenological research in forest ecosystems. *Agric. For. Meteorol.* **2012**, *152*, 159–177. [[CrossRef](#)]
73. Vabalas, A.; Gowen, E.; Poliakov, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [[CrossRef](#)]
74. Wainer, J.; Cawley, G. Empirical evaluation of resampling procedures for optimising SVM hyperparameters. *J. Mach. Learn. Res.* **2017**, *18*, 475–509. [[CrossRef](#)]