



Originally published as:

Wieland, M., Liu, W., Yamazaki, F. (2016): Learning Change from Synthetic Aperture Radar Images: Performance Evaluation of a Support Vector Machine to Detect Earthquake and Tsunami-Induced Changes. - *Remote Sensing*, 8, 10.

DOI: <http://doi.org/10.3390/rs8100792>

Article

Learning Change from Synthetic Aperture Radar Images: Performance Evaluation of a Support Vector Machine to Detect Earthquake and Tsunami-Induced Changes

Marc Wieland ^{1,2,3,4,*}, Wen Liu ² and Fumio Yamazaki ²

¹ Centre for Early Warning Systems, GFZ German Research Centre for Geosciences, Telegrafenberg, Potsdam 14473, Germany

² Department of Urban Environment Systems, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan; wen.liu@chiba-u.jp (W.L.); fumio.yamazaki@faculty.chiba-u.jp (F.Y.)

³ International Research Fellow of the Japan Society for the Promotion of Science, Tokyo 102-0083, Japan

⁴ Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

* Corresponding: marc.wieland@zoo.ox.ac.uk

Academic Editors: Zhong Lu, Richard Gloaguen and Prasad S. Thenkabail

Received: 6 June 2016; Accepted: 20 September 2016; Published: 23 September 2016

Abstract: This study evaluates the performance of a Support Vector Machine (SVM) classifier to learn and detect changes in single- and multi-temporal X- and L-band Synthetic Aperture Radar (SAR) images under varying conditions. The purpose is to provide guidance on how to train a powerful learning machine for change detection in SAR images and to contribute to a better understanding of potentials and limitations of supervised change detection approaches. This becomes particularly important on the background of a rapidly growing demand for SAR change detection to support rapid situation awareness in case of natural disasters. The application environment of this study thus focuses on detecting changes caused by the 2011 Tohoku earthquake and tsunami disaster, where single polarized TerraSAR-X and ALOS PALSAR intensity images are used as input. An unprecedented reference dataset of more than 18,000 buildings that have been visually inspected by local authorities for damages after the disaster forms a solid statistical population for the performance experiments. Several critical choices commonly made during the training stage of a learning machine are being assessed for their influence on the change detection performance, including sampling approach, location and number of training samples, classification scheme, change feature space and the acquisition dates of the satellite images. Furthermore, the proposed machine learning approach is compared with the widely used change image thresholding. The study concludes that a well-trained and tuned SVM can provide highly accurate change detections that outperform change image thresholding. While good performance is achieved in the binary change detection case, a distinction between multiple change classes in terms of damage grades leads to poor performance in the tested experimental setting. The major drawback of a machine learning approach is related to the high costs of training. The outcomes of this study, however, indicate that given dynamic parameter tuning, feature selection and an appropriate sampling approach, already small training samples (100 samples per class) are sufficient to produce high change detection rates. Moreover, the experiments show a good generalization ability of SVM which allows transfer and reuse of trained learning machines.

Keywords: machine learning; change detection; synthetic aperture radar; earthquake; tsunami

1. Introduction

With the rapidly growing supply of multi-temporal satellite imagery and the demand for up-to-date situation awareness in disaster situations, the need for robust change detection methods is constantly increasing. Various change detection [1] and, more specifically, building damage detection methods [2–4] were published in recent years. With respect to the input data, Synthetic Aperture Radar (SAR) provides clear advantages over optical satellite imagery as its acquisition is largely illumination and weather independent with current and future satellite missions providing high revisit periods. Drawbacks are largely related to the change of appearance with various incidence angles, or the presence of speckle noise.

Generally, two steps in the process of change detection can be distinguished, namely the creation of change features and their classification, which can be either unsupervised or supervised [5]. Concerning change feature creation, direct image comparison shows large potential for change detection in SAR images [6]. However, the mixture of additive and multiplicative noise contributions may cause high false alarm rates, and the choice of robust change features becomes essential to reduce the effects of noise and to improve the detection rates for any application. Calculating features over a moving window can be a way of reducing the effects of noise [7]. In this regard, object-based approaches, which calculate the change features from summary statistics over aggregated clusters of pixels (also referred to as super-pixels or objects), seem promising to create a robust feature space from an initial image segmentation or from independent objects (e.g., building footprints) [8]. However, only a few studies have dealt so far with object-based SAR change image analysis and further research is needed to understand its particular benefits and limitations [9].

The actual detection of change is largely done by means of unsupervised thresholding of the change feature space. Many studies exist that define thresholds based on experience or trial-and-error procedures to separate a one-dimensional feature space derived from bi-temporal image pairs [10,11]. Several threshold approximation methods have been proposed in recent literature to overcome the subjective bias and poor transferability of manual thresholding [12,13]. Liu and Yamazaki (2011) [14] analyze the change feature histograms to pick a threshold value. Bruzzone and Prieto (2002) [15] propose an approach based on Bayesian theory to adapt thresholds for different images with and without spatial-contextual information. Bazi et al. (2005) [16] use a generalized Gaussian model for automated threshold optimization. Despite their apparent ease of use, such thresholding approaches are usually not applied to a higher dimensional feature space and are limited to binary classification schemes. This is largely due to the difficulty associated with finding suitable threshold values and the need to adjust these for each of the involved features and classes, which increases the complexity of finding an overall optimal solution.

To this end, supervised machine learning approaches can provide valuable alternatives for change detection. They classify a multi-dimensional feature space based on the characteristics of a limited set of labeled training samples. A review on machine learning and pattern recognition for change detection with a view beyond satellite image analysis can be found in Bouchaffra et al. (2015) [17]. The majority of remote sensing studies use optical satellite imagery [18–20], whereas very few studies exist that explicitly apply learning machines to SAR imagery [21,22]. Gokon et al. 2015 [23], for example, use a combination of thresholding and a decision tree classifier on TerraSAR-X (TSX) data to distinguish three building damage classes caused by the 2011 Tohoku earthquake and tsunami. The study uses a one-dimensional feature space and provides indications about the transferability of the classifier by applying it to different subsets of the study area. However, a further evaluation of the influence of critical choices commonly made during the training phase of a learning machine (e.g., sample approach, number and distribution of training samples, classification scheme, and feature space) is not provided. Jia et al. (2008) [24] propose a semi-supervised change detection approach that uses a kernel k-means algorithm to cluster labeled and unlabeled samples into two neighborhoods, for which statistical features are extracted and that are fed into a Support Vector Machine (SVM) to perform the actual change detection. The approach specifically addresses the problems associated with sparse

availability of training samples. SVM is also widely used and shows superior results for other tasks, such as landuse/landcover classification [25].

Based on a screening of the recent literature it becomes apparent that further research is needed to better understand the capabilities and limitations of machine learning, and SVM in particular, in the context of change detection in SAR images. A sound understanding of the benefits and limitations of supervised SAR change detection and guidance on how to train a powerful learning machine for the detection of changes induced by natural disasters becomes particularly important, given the growing demand for such methods in disaster risk management. Especially in post-disaster situations rapid assessments of damage-related changes over large areas are required. In this regard, a better understanding of the influence of the input data, feature space and training approach on the change detection results is needed in order to design operational tools that can utilize the growing amount of satellite data to generate robust and validated information products for situation awareness in case of disasters.

The objective of this study is therefore to evaluate the performance of a SVM classifier to detect changes in single- and multi-temporal X- and L-band SAR images under varying conditions. Its purpose is to provide guidance on how to train a powerful learning machine for change detection in SAR images and thus to contribute to a better understanding of potential and limitations of supervised change detection approaches in disaster situations. The detection of changes induced to the building stock by earthquake and tsunami impact is used as application environment. With respect to previous work in this direction, the study at hand covers a wide range of performance experiments within a common evaluation framework, and focuses on research questions that have so far not specifically been evaluated. Moreover, a very large reference dataset of more than 18,000 buildings that have been visually inspected by local authorities for damages after the 2011 Tohoku earthquake and tsunami provides an unprecedented statistical population for the performance experiments. The specific research questions that are being tackled include:

- (1) How do the training samples influence the change detection performance?
- (2) How many change classes can be distinguished?
- (3) How does the choice of the acquisition dates influence the detection of changes?
- (4) How do X-band and L-band SAR compare for the detection of changes?
- (5) How does a SVM compare to thresholding change detection?

The study is structured as follows. In the next section, we describe the study area, images and reference data. In the subsequent sections, we introduce the method and present the results of the experiments undertaken to answer the previously raised research questions. Finally, a discussion and conclusions section close this study.

2. Study Area, Data and Software

The study focuses on the coastal areas of the Southern Miyagi prefecture in Japan (Figure 1 left), which was amongst the most severely affected regions hit by a Mw 9.0 earthquake and subsequent tsunami on 11 March 2011. The earthquake led to significant crustal movement over large areas and caused a tsunami with maximum run-up of 40.1 m [26]. Major damages to buildings, infrastructure and the environment occurred and a large number of people were reported dead or missing.

Three X-band images from the TerraSAR-X sensor taken five months before (t1), as well as one day (t2) and three months (t3) after the disaster were acquired over the study area (Table 1). Figure 1 (left and upper right) shows a false-color composite of the TerraSAR-X images that highlights the differences in backscattering intensities between the different acquisition times. The images were captured in StripMap mode with HH polarization on a descending path with 37.3° incidence angle and delivered as Single Look Slant Range Complex (SSC) products. Two L-band images of the ALOS PALSAR sensor with HH polarization on a descending path with 34.3° incidence angle, taken five months before (t1) and one month (t2) after the disaster were acquired and delivered at processing

level 1.1 (Table 1). Image preprocessing steps for both image types included multi-look focusing (with four equivalent number of looks), orthorectification (UTM 54N/WGS84), resampling, radiometric correction and conversion of digital numbers to sigma naught (db). Co-registration to the pre-event images has been performed using an algorithm based on Fast Fourier Transform (FFT) for translation, rotation and scale-invariant image registration [27]. No speckle filtering has been applied.

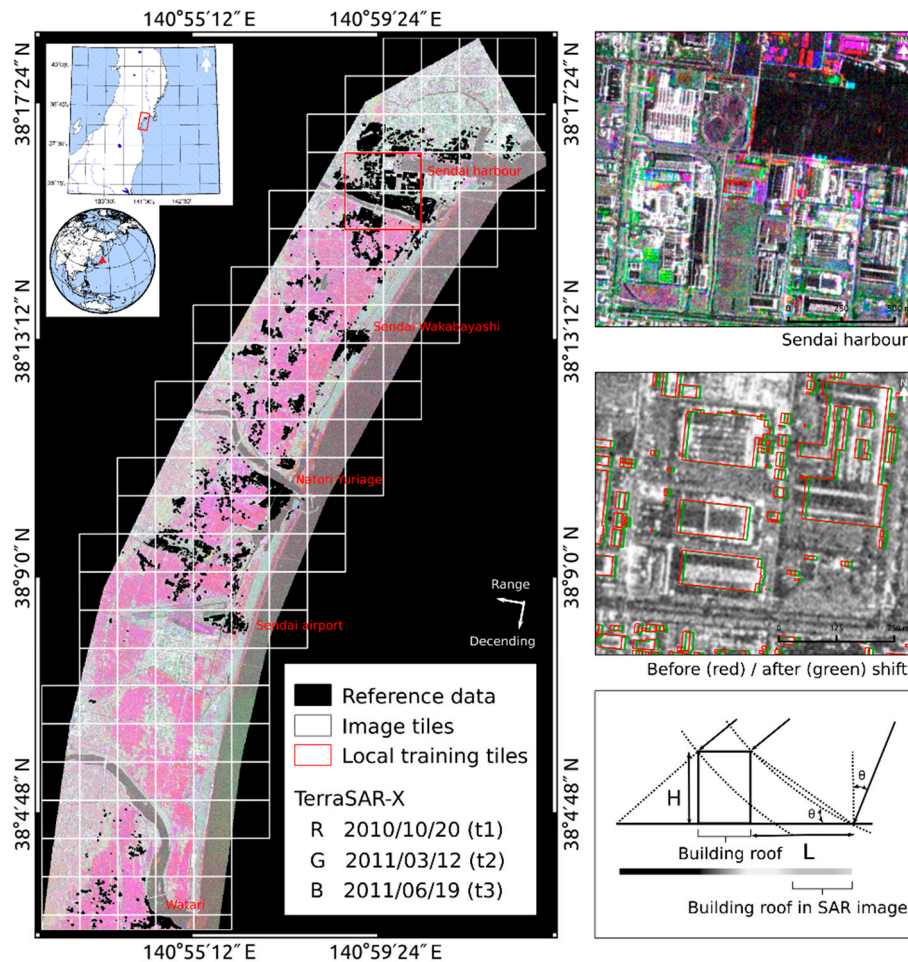


Figure 1. Study area covering the coastal areas of Southern Miyagi prefecture, Japan. Reference building footprints from damage surveys are superimposed in black on a RGB color-composite of the TerraSAR-X images (left and top, right). Image tiles are outlined in white. In red outline are the tiles that were used to create a local training dataset. Magnified view of a random location within the TerraSAR-X scene (middle, right) with building footprints superimposed before (in red) and after (in green) a shift has been applied to account for the mismatch between footprints and image due to SAR geometry (bottom, right).

Table 1. Overview of the satellite images used in this study.

Sensor	Acquisition	Incidence Angle	Pixel Spacing	Path	Polarization	Band
TerraSAR-X	20 October 2010 (t1)	37.3°	1.25	Descending	HH	X
TerraSAR-X	12 March 2011 (t2)	37.3°	1.25	Descending	HH	X
TerraSAR-X	19 June 2011 (t3)	37.3°	1.25	Descending	HH	X
ALOS Palsar	5 October 2010 (t1)	34.3°	6.25	Descending	HH	L
ALOS Palsar	7 April 2011 (t2)	34.3°	6.25	Descending	HH	L

Comprehensive reference data are used in this study from a database of building damages that were surveyed by the Japanese Ministry of Land Infrastructure, Transport and Tourism after the

disaster [28]. The data are referenced at the building footprint and include 18,407 buildings described by seven damage categories. Two reclassifications of the data have been performed as is depicted in Table 2. An overview of potential tsunami damage patterns and their characteristics in SAR imagery can be found in [23]. The building geometries have, moreover, been shifted to match the building outlines in the SAR images. Figure 1 (middle, right) shows the original building geometries in red and the shifted ones in green over the pre-event image (t1) for TerraSAR-X. Parts of the building walls with highest backscatter from corner reflection are outside the original footprints for most of the buildings. This can be attributed largely to the fact that a building in a TerraSAR-X image shows layover from the actual position to the direction of the sensor (Figure 1 bottom, right). The layover (L) is proportional to the building height (H) and can be calculated as

$$L = \frac{H}{\tan\theta} \quad (1)$$

where θ is the incident angle of the microwave. In order to account for this effect, the building geometries were shifted towards the direction of the sensor to match the TerraSAR-X images. With $\theta = 37.3^\circ$ and an assumed average building height of $H = 6$ m (approximately the height of a two-storied building), the layover is approximately 7.9 m. The assumption that the majority of the buildings in the study area have two stories is based on field work as described in Liu et al. [11]. Considering the path of the satellite (190.4° clockwise from north), the layover can be decomposed into 7.8 m to the east and 1.4 m to the south, which results in a lateral shift of the building geometries of 6 px to the east and 1 px to the south on the basis of a resampled pixel spacing of 1.25 m. Comparing the adjusted building geometries with the original ones (Figure 1 middle, right) larger areas of high backscattering intensities are located within the building footprints. Similarly, a copy of the reference data was shifted for the ALOS imagery according to incidence angle and path of the satellite by 8.7 m to the east and 1.6 m to the south.

Preprocessing of the satellite images has been performed with the Sentinel-1 Toolbox. Co-registration and all other processing and analysis steps have been implemented in Python using the GDAL, NumPy, SciPy and Scikit-learn libraries.

Table 2. Classification schemes applied to the reference damage data.

Original Classification		Reclassification I		Reclassification II	
Class—Label	Samples	Class—Label	Samples	Class—Label	Samples
1—Washed away	7065	1—Major change	8775	1—Change	8775
2—Collapsed	1710				
3—Complete damage (flooded over first floor)	879	2—Moderate change	6228	2—No change	9632
4—Major damage	1950				
5—Moderate damage (flooded over ground floor)	3399				
6—Minor damage (flooded under ground floor)	1964	3—No change	3404		
7—No damage	1440				
Total	18,407		18,407		18,407

3. Methods

Figure 2 depicts a schematic overview of the classification and performance evaluation framework that has been set up for this study. Following an object-based approach to image analysis, the available building footprints (Section 2) are used to cluster neighboring pixels and thus to segment the image into higher-order computational units. This segmentation of the image is done by a simple spatial intersection where image pixels that intersect with a footprint are considered to belong to the same segment. A SVM (Section 3.1) is used to classify the feature space (Section 3.2) based on training datasets, which represent samples of labelled feature vectors. Both multi-temporal and mono-temporal features are considered for the task of classifying changes in the images. In order

to evaluate the performance of the classifier with respect to decisions typically introduced during the data preparation and training stages, cross-validated and non-cross-validated accuracy measures are reported (Section 3.3). The results of this study (Section 4) shall thus provide guidance to train a powerful SVM for change detection.

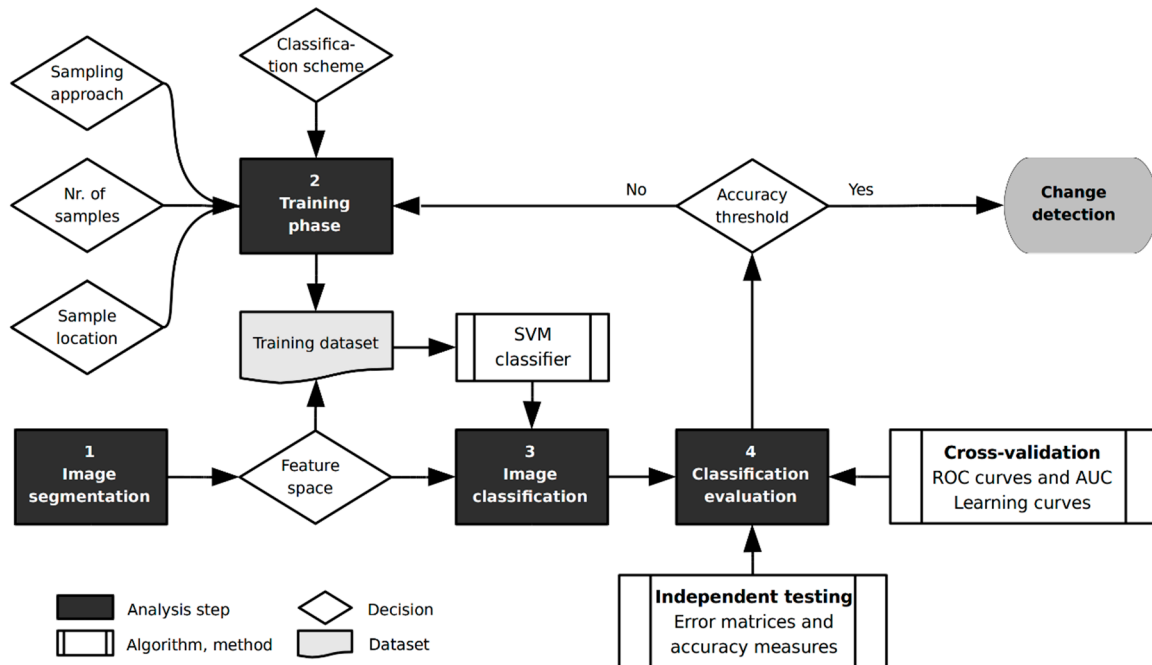


Figure 2. Schematic overview of the classification and performance evaluation framework.

3.1. Support Vector Machine (SVM)

Support Vector Machine (SVM) has been selected as promising classifier to be used within this study for the task of detecting changes from multi-temporal SAR images. The choice of the classifier is based on previous work of the authors that highlighted the superior performance of SVM for the classification of built-up areas in multi-spectral images [25]. SVM is a non-parametric classifier [29] that utilizes kernel functions to project non-linearly separable classes into higher dimensional feature space to make them separable by a linear hyperplane. Margin maximization is used to choose the optimal separating hyperplane so that only the closest feature vectors (support vectors) to the edge of the class distribution are used. A soft-margin parameter allows some data points to violate the separation through the hyperplane without affecting the final result. Multi-class problems are solved by applying a one-against-one scheme. In this study, optimal SVM parameters (kernel function ϕ , kernel coefficient γ and penalty or regularization parameter C) are tuned for each classification according to a ten-fold cross-validation and grid-search method during the training phase of the classifier. Kernel functions that have been considered in the grid-search include linear, polynomial, radial basis function and sigmoid. Ranges of values for γ and C have been selected based on the literature. An optimal parameter selection is reached when the cross-validation estimate of the test samples error is minimal.

The standard formulation of SVM does not provide class membership probabilities, which would be needed to get an estimate of the classifier's confidence, expressed by, for example, the Shannon entropy [30]. The probabilities can, however, be calibrated through logistic regression on the SVM scores and fit by an additional cross-validation on the training data as is described in Platt (1999) [31]. Therefore, the class membership probability of a sample x is computed from its distances to the optimal separating hyperplanes for each of the $n(n - 1)/2$ binary SVMs. This is done by fitting a sigmoid function to the decision values of each of the binary classifiers. The probabilistic output of the binary

classifiers is then combined into a vector that contains the estimated class memberships associated with the sample defined as

$$pk(x) = \{pk_1(x), pk_2(x), \dots, pk_i(x), \dots, pk_n(x)\} \quad (2)$$

where $pk_i(x)$ is the estimated membership degree of x to class i , and n is the number of classes [32]. From the probability vector $pk(x)$, the Shannon entropy H can be calculated with Equation (3).

$$H(x) = -\sum_{i=1}^n pk_i \log_2 pk_i \quad (3)$$

3.2. Feature Space and Feature Selection

Both mono-temporal and multi-temporal features are considered within this study. In case of a mono-temporal analysis where changes are classified solely on the information content of a single image, the feature space is limited to mean, mode, standard deviation, minimum and maximum values of the backscatter coefficients computed per building footprint. For multi-temporal change detection, an object-based change feature space is created by computing summary statistics of pixel-based change features for each building footprint. First, pixel-based change features are computed between two images which were acquired over the same spatial subset but taken at different times. Second, mean, mode, standard deviation, minimum and maximum values are computed per building footprint for the resulting change images. Using three pixel-based change features and five summary statistics per feature results in a total of 15 object-based change features. The multi-temporal change features for which the per-building statistics are computed include the averaged difference over a moving window, the correlation coefficient, and the change index as a combination of difference and correlation. The difference (d) is calculated by Equation (4) and the correlation coefficient (r) is calculated by Equation (5).

$$d = \bar{I}b - \bar{I}a \quad (4)$$

$$r = \frac{N \sum_{i=1}^N I a_i I b_i - \sum_{i=1}^N I a_i \sum_{i=1}^N I b_i}{\sqrt{\left(N \sum_{i=1}^N I a_i^2 - \left(\sum_{i=1}^N I a_i \right)^2 \right) \cdot \left(N \sum_{i=1}^N I b_i^2 - \left(\sum_{i=1}^N I b_i \right)^2 \right)}} \quad (5)$$

where i is the pixel number, $I a_i$ and $I b_i$ are the backscattering coefficient of the second (post) and first (pre) images, $\bar{I}a$ and $\bar{I}b$ are the corresponding averaged values over the $N = 5 \times 5$ pixel window surrounding the pixel i .

Difference and correlation are combined into a change index (z) as introduced by [14] and described by Equation (6).

$$z = \frac{|d|}{\max(|d|)} - w \cdot r \quad (6)$$

where $\max(|d|)$ is the maximum absolute value in difference and w is the weight between the difference and the correlation coefficient. A weight of $w = 1$ has been chosen in order to equally weight difference and correlation for the calculation of the z .

In order to identify the most significant features for each classification task and training dataset, a recursive feature selection algorithm as proposed by Guyon et al. (2002) [33] is used in this study. Given an external classifier that assigns weights to features, recursive feature selection considers iteratively smaller and smaller sets of features. With each iteration the features are used to train a classifier and are assigned weights according to their discriminating power. The features with the smallest weights are eliminated from the feature set for the next iteration. The feature space that maximizes a scoring value is selected. Feature selection has been performed in a ten-fold cross-validation loop and all feature values were standardized to zero mean and unit variance.

3.3. Performance Measures

To assess the classification performance, standard accuracy measures (precision, recall and F1 score [34]) are reported as average and standard deviation over the results of cross-validation iterations. Cross-validation increases the reliability of the results by reducing the bias resulting from specific training-testing datasets. In case of multi-class classification problems, the weighted average for each performance measure and class is provided. Additionally, final map accuracies are evaluated by deriving error matrices and standard accuracy measures in a non-cross-validated manner from independently sampled testing data.

Receiver Operating Characteristic (ROC) curves are used to visualize the performance of a classifier against reference labels while a discrimination threshold is varied. Plotting the true positive rate against the false positive rate at various thresholds results in the ROC curve. Compared to precision, recall and F1 score, ROC curves are solely based on true positive and false positive rates, and thus are insensitive to changes in class distribution of the test dataset [34]. The Area Under the Curve (AUC) is used as a single scalar value to describe the classifier performance as derived from an ROC curve.

A learning curve shows the training and validation score of a classifier for varying numbers of training samples. Learning curves are used to evaluate the influence of the training data size on the classification performance and to find out whether the classifier suffers more from a variance or a bias error. The training data size is iteratively increased and ten-fold cross-validation is applied to derive the mean score and the range of scores for each iteration.

4. Results

The performance of statistical pattern recognition systems, which includes classification accuracy, generalization ability, computational efficiency and learning convergence, can be influenced by several decisions commonly made during the training phase of a classifier. These include feature selection, training sampling approach, number and location of training samples, classification scheme and the choice of classifier parameters (Figure 2). SVM parameters are tuned according to a ten-fold cross-validation and grid-search method where an optimal parameter selection is reached when the cross-validation estimate of the test samples error is minimal. Feature selection is performed automatically using the above described recursive feature selection algorithm. The following experiments assess the influence of the training samples and the classification scheme. Moreover, the influence of image date and type are evaluated and a statistical learning approach is compared to a commonly used threshold method.

4.1. Influence of the Training Samples

The objective of this experiment is to find the optimal sampling approach and number of samples. Moreover, it aims at providing indication about the influence of the spatial distribution of training samples. Results are presented for TSX using the t1t3 image pair. The results of using three different sampling approaches, each with 3000 samples taken over the whole study area, are compared to each other under consideration of varying training-testing datasets as part of a cross-validation procedure. Simple random sample (SRS) and stratified random sample (STRS) follow the natural distribution of the classes in the study area, whereas the balanced random sample (BRS) represents a random sample with the number of samples per class being equally balanced (1500 per class). From Figure 3 it can be seen that a balanced random sample (BRS) clearly outperforms stratified random sampling (STRS) and simple random sampling (SRS). This result is further highlighted by Table 3, which shows the results from a comparison against an independent testing dataset. Precision, recall and F1 score can be improved by up to 0.1 by using a balanced random sample. Main difficulties that arise when using STRS or SRS are related to low recall values for the “no change” class.

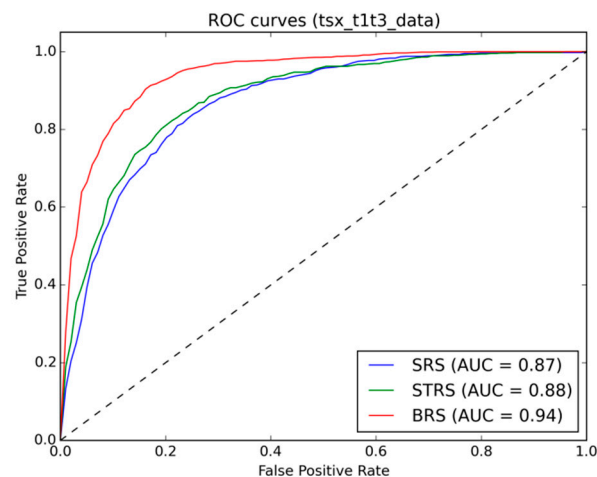


Figure 3. Receiver operating characteristic (ROC) curves with area under the curve (AUC) comparing simple random sampling (SRS), stratified random sampling (STRS) and balanced random sampling (BRS) for training data selection.

Table 3. Comparison of different training data sampling approaches (simple random sample—SRS, stratified random sample—STRS, balanced random sample—BRS) against an independent testing dataset. Training and testing have been performed over the full image scene.

Class	SRS			STRS			BRS			Samples
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	
1—Change	0.68	0.92	0.78	0.67	0.93	0.78	0.88	0.81	0.85	1500
2—No Change	0.88	0.56	0.68	0.89	0.54	0.67	0.83	0.89	0.86	1500
Total	0.78	0.74	0.73	0.78	0.74	0.73	0.86	0.85	0.85	3000

Figure 4 (top) shows the feature selection and learning curve for the SVM classifier that has been trained with a BRS of 3000 samples distributed over the whole study area. It can be seen that almost all features are used for the classification (14 out of 15 features) and that a sample size of 3000 is sufficient for stable classification performance at high accuracy values. The learning curve shows a potentially good generalization ability of the classifier at more than 1500 samples. It indicates that adding more training samples would not significantly improve the classification performance. The larger gap between validation and training score at small training sample size indicates a tendency of the classifier to over-fit the data on small training datasets. In order to further test the transferability of the classifier we took a BRS of 800 samples (400 per class) from a small and geographically clustered subset (see local training tiles in Figure 1) and tested the classifier against an independent testing dataset selected over the rest of the study area.

Figure 4 (bottom) shows feature selection and learning curve for the locally trained SVM. It can be seen from the learning curve that, even with a strongly reduced and locally selected training sample set, relatively high accuracy can be achieved. Evaluation against an independent testing dataset covering the full image scene underlines the good performance of the locally trained SVM (Table 4). However, a clear performance decrease for the locally trained classifier can be observed with respect to a globally trained one (Table 3). Adding more data could, however, potentially increase the generalization ability of the classifier. Other possible strategies include using less features or increasing the regularization of the classifier. We already optimize both parts automatically in the analysis chain through feature selection and tuning of the SVM parameters. A regularization parameter of $C = 100.0$ and a reduced feature space of 8 dimensions underline the fact that the implemented analysis chain is reacting to a limited training sample size.

Table 4. Accuracy assessment of a locally trained SVM with a balanced random sample (BRS) against an independent testing dataset covering the full image scene.

Class	BRS (Local Trained)			Samples
	Prec.	Rec.	F1	
1—Change	0.78	0.79	0.78	1500
2—No Change	0.79	0.78	0.78	1500
Total	0.78	0.78	0.78	3000

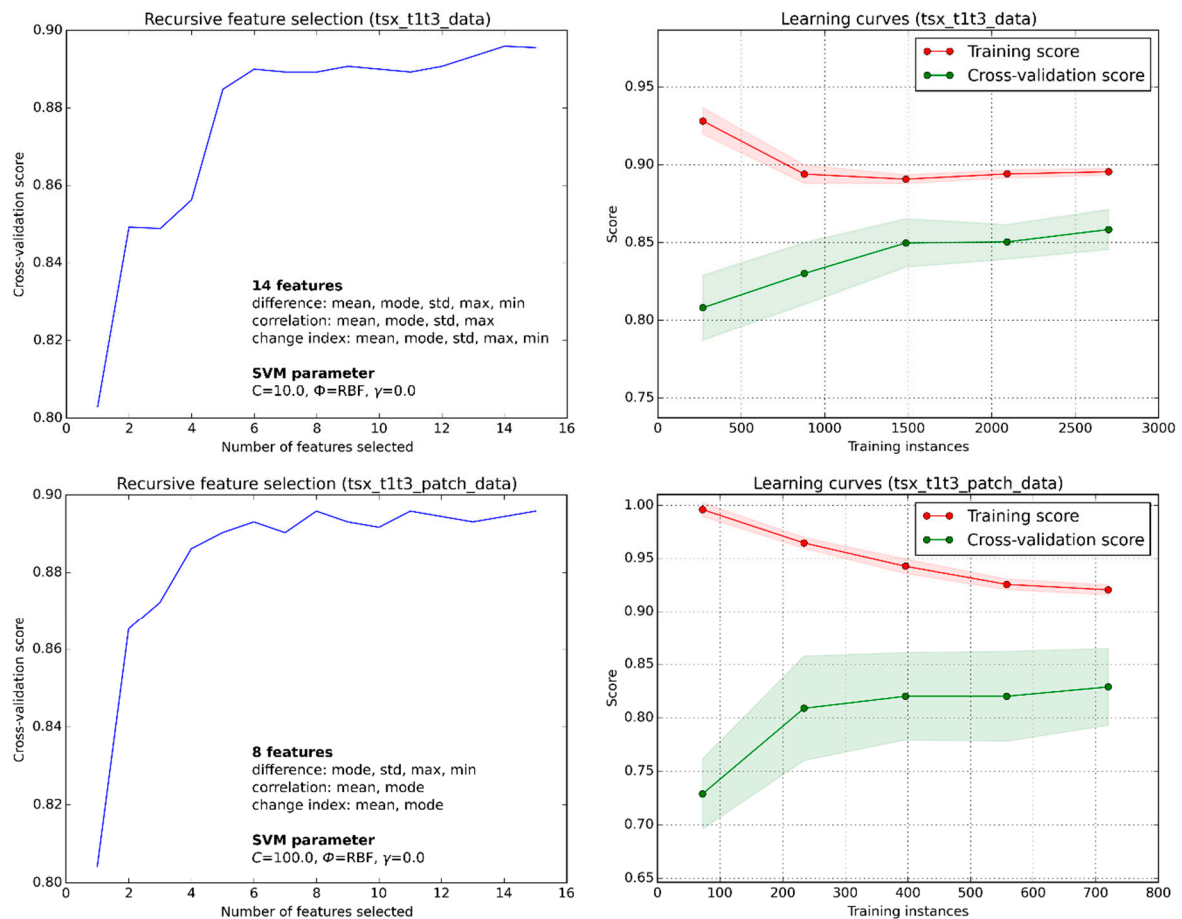


Figure 4. Feature selection and learning curves for an SVM that has been trained and tested over the full image scene (top), and for an SVM that has been locally trained on selected image tiles (bottom).

4.2. Influence of the Acquisition Date

Figure 5 shows different acquisition dates for a representative subset of optical satellite images and the respective TSX images. The TSX images have been speckle filtered using an enhanced Lee filter and stacked into a false-color composite for better visualization of changes. The acquisitions are centered on the earthquake and tsunami disaster with t1 being before, t2 being a few days after and t3 being a few months after the disaster. As a spatial reference, the building footprints are superimposed. On the optical images, clear differences in scene settings can be observed between the timestamps. Large amounts of debris can be found in the t2 image scene taken immediately after the disaster. By comparison with t3 where debris and damaged buildings have been removed, it can be seen that some of the severely damaged (and later removed) buildings seem unchanged in t2. This is largely due to the viewing perspective of the satellite sensor that mainly captures the roof structures, whereas especially tsunami-induced damages largely affect the lower floors and leave the roof untouched (unless the building gets washed away).

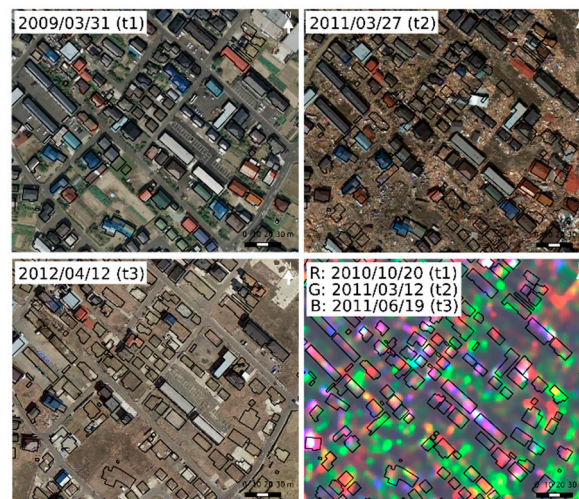


Figure 5. Comparison of different acquisition dates as seen in optical satellite images (GoogleEarth) and in TerraSAR-X images (as speckle filtered false-color composite for better visualization) for three different timestamps (t1: pre-event; t2: post-event damage and debris; t3: post-event cleaned up).

In this experiment, change features have been computed for two different image pairs (t1t2 and t1t3). Moreover, single date classifications have been performed on the post-event images in t2 and t3. For these classifications, the feature space has been reduced to mean, mode, standard deviation, minimum and maximum values of the backscatter coefficients computed per building footprint. Figure 6 shows the ROC plots for the different classifications. It can be seen that classifications involving the t2 image perform worse than using the later t3 image, both for the multi-temporal and the single date classifications. Moreover, it can be seen that despite the reduced feature space, single date classifications can achieve reasonably good performance. Nevertheless, a multi-temporal classification approach would be preferable. This is confirmed by the independent testing results in Table 5. The results further indicate that even the side-looking nature of the SAR images may not be sufficient to reliably detect tsunami-induced damages to building side-walls using the proposed approach and feature space.

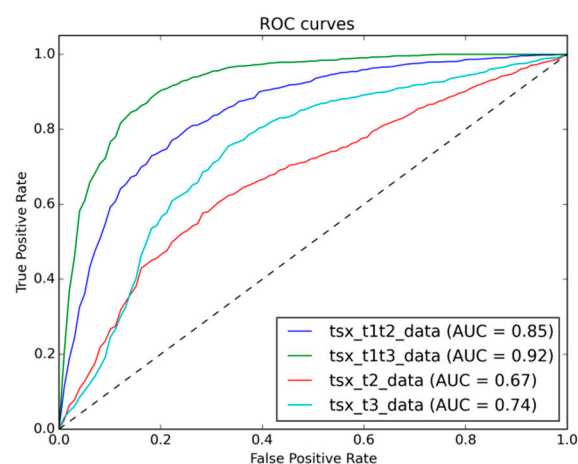


Figure 6. Receiver operating characteristic (ROC) curves with area under the curve (AUC) comparing the classification performance on images of different acquisition dates. Image comparisons between pre- and post-event dates (t1t2 and t1t3), and single date classifications of post-event images (t2 and t3) are considered.

Table 5. Comparison of using different acquisition dates for the change detection against an independent testing dataset. Training and testing have been performed over the full image scene. A balanced random sample has been drawn for training the classifier.

Class	t1t2			t1t3			Samples
	Prec.	Rec.	F1	Prec.	Rec.	F1	
1—Change	0.75	0.77	0.76	0.88	0.81	0.85	1500
2—No Change	0.76	0.74	0.75	0.83	0.89	0.86	1500
Total	0.76	0.76	0.76	0.86	0.85	0.85	3000

Class	t2			t3			Samples
	Prec.	Rec.	F1	Prec.	Rec.	F1	
1—Change	0.64	0.71	0.68	0.70	0.69	0.69	1500
2—No Change	0.68	0.60	0.64	0.69	0.71	0.70	1500
Total	0.66	0.66	0.66	0.70	0.70	0.70	3000

4.3. Influence of the Classification Scheme

In this experiment we use the TSX t1t2 image pair, which describes changes that occurred in direct consequence to the disaster and should include buildings of different damage classes as described by the reference dataset. For this experiment we did not use the TSX t1t3 image pair, since in t3 already a large number of damaged buildings have been removed. Figure 7 shows a comparison of classifiers being trained with the same sampling approach and number of samples per class but with varying classification schemes (Table 2). The sample size has been set to 450 samples per class, which equals 50% of the samples in the smallest class of the original classification scheme. The other 50% are needed to form the independent testing data. ROC curves are derived from cross-validation on the training data, whereas error matrices are created by comparison against the independent testing data. It can be seen that as soon as more than just two classes (“change,” “no change”) are considered, the classifier fails to properly separate them and the classification performance decreases drastically. This indicates not necessarily a deficiency of the classifier itself but rather a deficiency either of the input images and/or the derived feature space to describe the more detailed change classes.

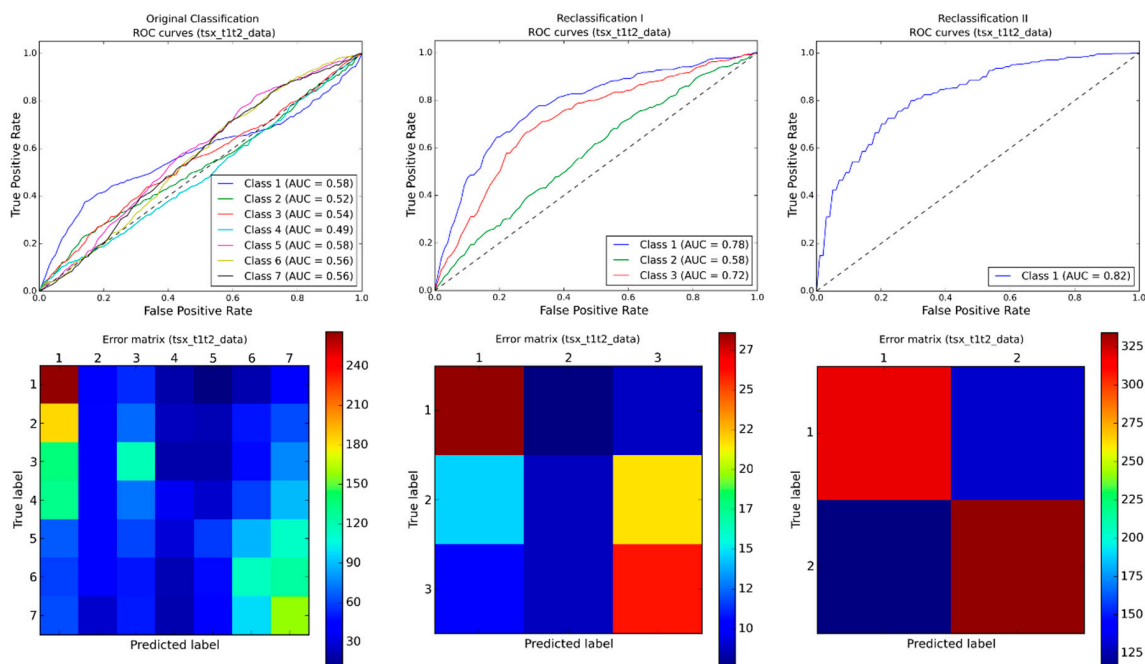


Figure 7. Receiver operating characteristic (ROC) curves with area under the curve (AUC) from ten-fold cross-validation on the training data and error matrices from comparison with independent testing dataset for different classification schemes (Table 2).

4.4. Influence of the Image Type

Figure 8 (left) shows the ROC curves from cross-validation on a random sample of 3000 buildings applied to ALOS and TSX. The acquisition dates between ALOS and TSX were selected as close as possible together in time, with five days difference for the pre-event and 26 days for the post-event imagery. Despite the apparent differences in spatial resolution of the sensors, the performance on both image types is sufficiently high with AUC values of 0.77 (ALOS) and 0.81 (TSX). To account for differences in spatial resolution of the sensors, a second sample has been drawn from the reference buildings filtered by building footprint area. Only buildings with a footprint area larger than 160 m² are considered, which is approximately the area covered by four pixels of the ALOS imagery. Figure 8 (right) shows the accordant ROC curves. A better performance when focusing solely on larger buildings can be observed on both image types with AUC values of 0.84 (ALOS) and 0.86 (TSX). The performance difference between the image types slightly decreases, and comparable performance can be observed on both image types.

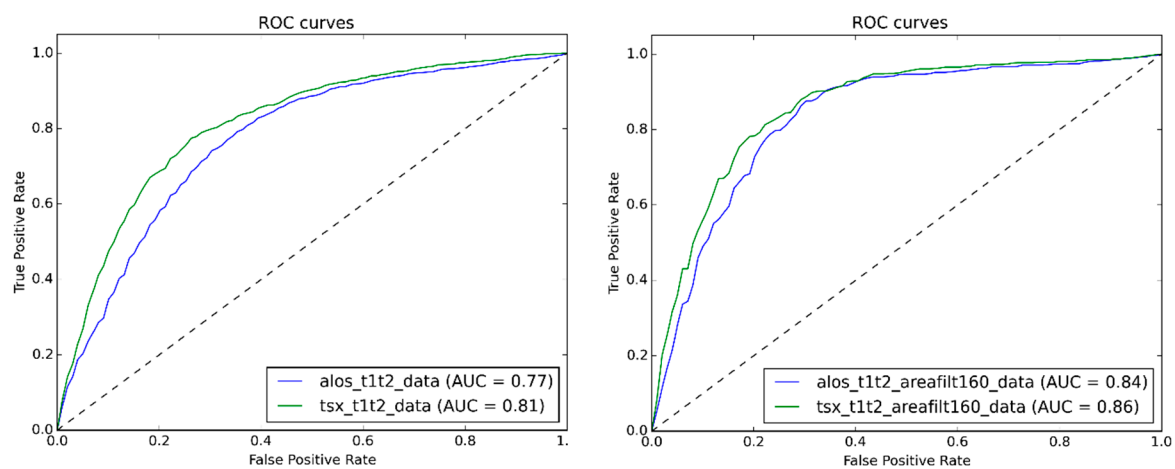


Figure 8. Receiver operating characteristic (ROC) curves with area under the curve (AUC) for a comparison between TSX and ALOS using unfiltered (left) and filtered (right) reference data. The filtered reference data includes only buildings with a footprint area larger than 160 m².

4.5. Comparison of SVM with Threshold Change Detection

A typical change detection method is to threshold a change feature. In this study, we iteratively change the threshold v over the whole range of the change index z (Equation (7)).

$$v = \{z | \min(z) \leq z \leq \max(z)\} \quad (7)$$

For 100 iterations, the predicted changes are compared to the changes identified by the respective reference dataset. The comparison is done on a per-building basis, where the mean change index per building is used as change feature. The mean change index per building combines the difference and correlation features in a single metric and has been used for a similar task in a previous study [11]. The same reference datasets with a balanced random sample of 3000 buildings are used for threshold and SVM change detection. ROC curves are produced by cross-validation for SVM, and by varying the threshold for the threshold change detection. Figure 9 shows the ROC curves for the two image types and different acquisition dates comparing SVM with threshold change detection (upper row). In all considered experimental set-ups, SVM outperforms the threshold method by 0.04 to 0.13 in terms of AUC values. The lower row of Figure 9 depicts the relation between accuracy (measured by F1 score) and threshold values. It highlights the sensitivity of this change detection method to the selection of

the threshold value. Consequently, we also tested a commonly used threshold selection method that uses the change feature value distribution to approximate a threshold v as described by Equation (8).

$$v = \mu(z) + 2 \times \sigma(z) \tag{8}$$

From the plots in Figure 9 it can be seen that approximating a threshold is not a trivial task and the tested method does not necessarily provide an optimal solution for the detection of changes. The performance difference of the two change detection methods is further underlined by comparison against an independent testing dataset (Table 6). In this case, the change detection resulting from the best threshold over all iterations was compared with results obtained from a trained SVM for the different image types and acquisition dates.

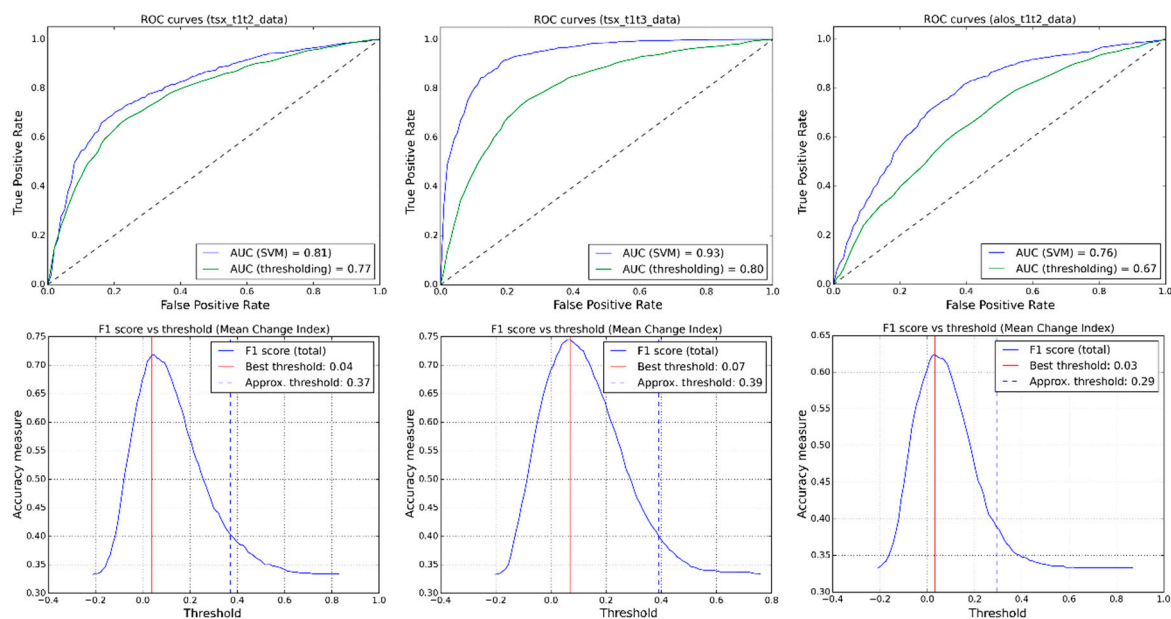


Figure 9. Receiver operating characteristic (ROC) curves with area under the curve (AUC) for a comparison of SVM change detection with thresholding method (**top**). Relationship between F1 score and threshold selection (**bottom**).

Table 6. Comparison of SVM and threshold change detection against an independent testing dataset for different image types and acquisition dates.

Class	SVM (tsx_t1t2)			Threshold (tsx_t1t2)			Samples
	Prec.	Rec.	F1	Prec.	Rec.	F1	
1—Change	0.75	0.75	0.75	0.68	0.78	0.72	1500
2—No Change	0.75	0.75	0.75	0.74	0.63	0.68	1500
Total	0.75	0.75	0.75	0.71	0.70	0.70	3000

Class	SVM (tsx_t1t3)			Threshold (tsx_t1t3)			Samples
	Prec.	Rec.	F1	Prec.	Rec.	F1	
1—Change	0.88	0.83	0.85	0.76	0.73	0.75	1500
2—No Change	0.84	0.89	0.86	0.74	0.77	0.75	1500
Total	0.86	0.86	0.86	0.75	0.75	0.75	3000

Class	SVM (alos_t1t2)			Threshold (alos_t1t2)			Samples
	Prec.	Rec.	F1	Prec.	Rec.	F1	
1—Change	0.77	0.65	0.71	0.62	0.62	0.62	1500
2—No Change	0.70	0.81	0.75	0.62	0.63	0.63	1500
Total	0.74	0.73	0.73	0.62	0.62	0.62	3000

4.6. Summary of Results

Figure 10 shows the final SVM change detection and classifier confidence maps derived with a balanced random sample of 3000 buildings for the different sensor types and acquisition dates over the whole study area. Of the 18,407 considered buildings in the study area, 58% were classified as being “changed” by ALOS (t1t2), 79% by TSX (t1t2) and 75% by TSX (t1t3). The reference data indicates 72% of changed buildings over the whole study area for comparison. Five hotspot areas where changes are densely clustered can be identified in all three change maps, namely Sendai harbor (1), Sendai Wakabayashi (2), Natori Yuriage (3), Sendai airport (4) and Watari (5) from north to south. The numbers of changed buildings in the five hotspot areas are presented in Figure 11 (left) and compared to the reference dataset. With the exception of TSX (t1t3) on hotspot area 3, a tendency to underestimate the number of changed buildings can be observed for all image pairs in the change hotspots. The difference with respect to the reference data is largest for ALOS (t1t2), followed by TSX (t1t2) and TSX (t1t3) which comes closest to the number of actually changed buildings both considering the change hotspots and the study area as a whole. The better performance on TSX (t1t3) is also reflected in the respective entropy value distributions for the changed buildings (Figures 10 and 11, right), that appear to be significantly lower. The relation between change detection performance (in terms of hits and misses) and classifier confidence (measured by Shannon entropy) is further outlined in Figure 12. A clear separation of entropy values between hits and misses can be observed over the whole study area (left), with high entropy values indicating likely misclassification. Also, the spatial distribution of hits and misses (middle) can be linked to the distribution of the confidence values (right) as illustrated for a randomly selected subset area.

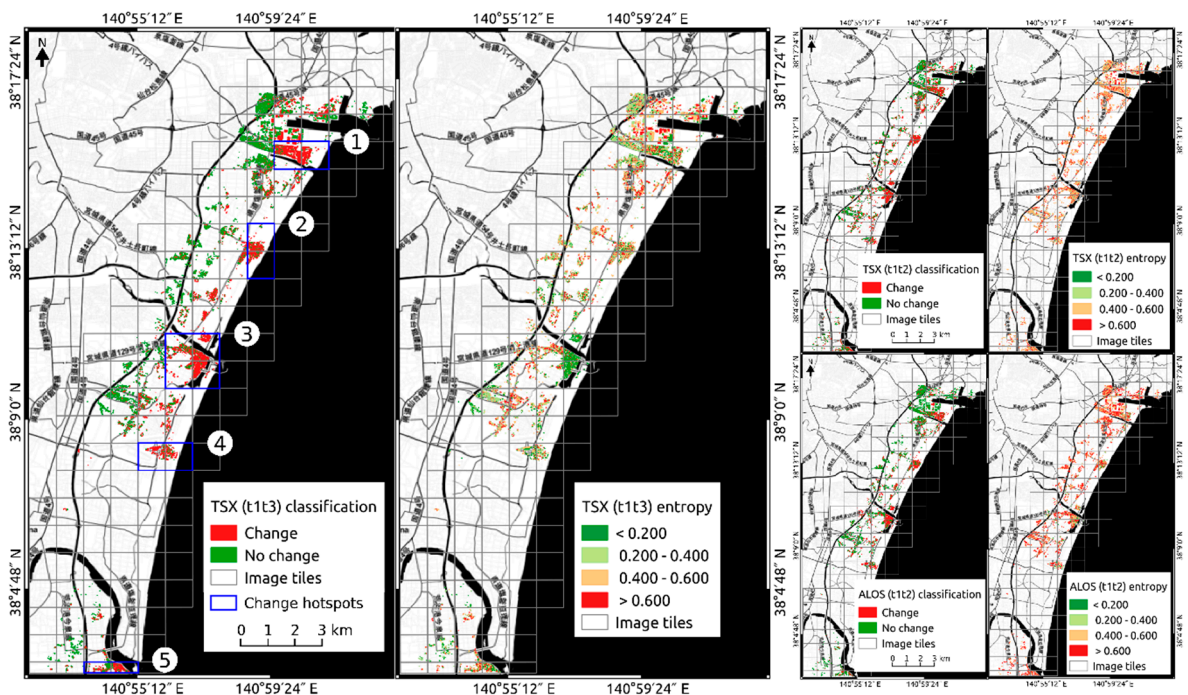


Figure 10. Change and confidence maps from TSX and ALOS for the whole study area.

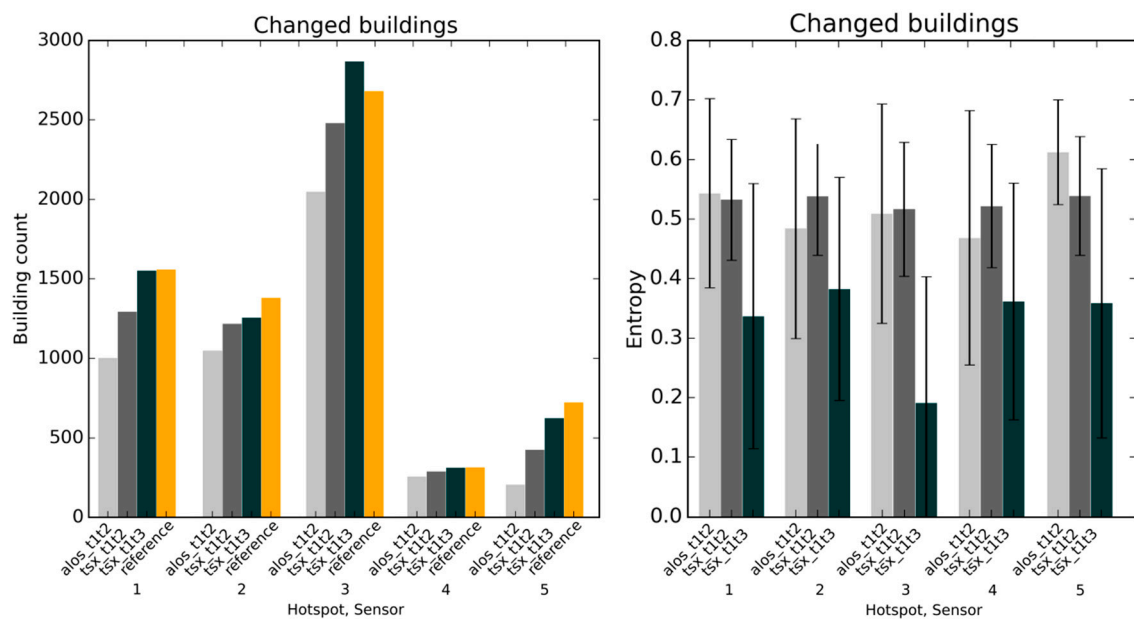


Figure 11. Changed buildings and related entropy value distributions for five hotspot areas, grouped by sensor type and acquisition date.



Figure 12. Relation between change detection performance (hits and misses) and classifier confidence (entropy).

5. Discussion

The study showed that major changes to the building stock are clearly described by the backscattering intensities of X- and L-band SAR images and can be detected by a trained and tuned SVM learning machine. A number of research questions have been considered and are discussed in the following in order to give guidance on how to train a powerful SVM for change detection.

(1) How do the training samples influence the change detection performance?

Given a large enough training sample size (>750 samples per class), good generalization ability of the SVM at high accuracy level (>0.85 F1 score) could be observed (Figure 4, top). Even when reducing the training sample size to 400 samples per class, good classification accuracy (>0.83 F1 score) could be measured when trained and tested over the whole study area. The learning curves indicate that further reducing the sample size to as low as 100 samples per class could still provide reliable results (>0.80 F1 score), albeit at the cost of losing generalization ability. To this regard, we could also show that it is possible to apply a locally trained classifier to different areas (Table 4) with minor loss of performance (0.78 F1 score) compared to a classifier that has been trained over the whole image scene (0.85 F1 score). The generalization experiment carried out in this study, however, could only be applied within the same image scene. Therefore, further tests involving different scenes are needed to strengthen and further constrain these findings. The training sample approach had a clear impact on the results, and a balanced random sample of training data produced superior results (0.85 F1 score)

over stratified (0.73 F1 score) or simple random samples (0.73 F1 score). This confirms the sensitivity of SVM to class imbalance as also observed in other classification domains [35] (Table 3). An alternative strategy to account for the class imbalance is to use the prior class distribution, as estimated from the training samples, to weight the penalty parameter C during classification. An in-depth discussion on imbalanced learning can be found in He and Ma (2013) [36].

(2) How many change classes can be distinguished?

The classifier performed well in case of a simple binary classification task (change–no change). Given the tested feature space and classification approach, a distinction between different types of changes in terms of damage grades led to poor performance of the classifier. The types of changes that are related to tsunami-induced damages represent a major difficulty to be detected by satellite images in general, since they occur mainly in the side-walls of the structures and not the roof. Yamazaki et al. (2013) [37] present an approach to tackle that problem by utilizing the side-looking nature of the SAR sensor. Their results seem promising for single buildings. Given the scenario and approach followed within the study at hand, however, such changes could not be detected in a robust manner over a large number of buildings. To this regard, additional change features such as texture, coherence and curvelet features [7,38] should be tested in more depth.

(3) How does the choice of the acquisition dates influence the detection of changes?

Different image acquisition dates have been tested for their influence on the classification performance (Figure 6, Table 5). The best results were observed on the t1t3 image pair (0.92 AUC), which uses a post-event image acquired three months after the disaster where large amounts of debris and collapsed buildings had already been removed. Image pairs that utilize shorter acquisition dates after the event (one day for TSX) showed a performance decrease (0.85 AUC). One reason for this is that, in many cases where the reference data reports a total collapse, only the lower floors of structures are damaged by the tsunami, whereas the roofs do not show any apparent changes from the satellite's point of view. As can be seen from Figure 5, these buildings are still present in the image acquired immediately after the disaster, but were then removed as part of the clean-up activities in the weeks after the disaster. Therefore, the aforementioned limitations related to the viewpoint of satellite-based change detection are further confirmed by this experiment and it could be shown that this becomes particularly important for any change detection application that aims at providing immediate post-disaster situation awareness.

A benefit of the proposed supervised change detection approach is that once training samples are defined and labelled according to the desired classification scheme, basically any change feature space can be processed without further adjustments. In this regard, also single date classifications that use only one post-disaster image have been successfully tested within the same framework. Classification of the mono-temporal feature space could achieve reasonably good performance on both t2 (0.67 AUC) and t3 (0.74 AUC) images that, however, could not compete with the multi-temporal approach (t1t2 image pair: 0.85 AUC; t1t3 image pair 0.92 AUC). It shows, nevertheless, that the proposed approach is flexible enough to deal with a multitude of possible data availability situations in case of a disaster.

(4) How do X-band and L-band SAR compare for the detection of changes?

The flexibility of the approach is further emphasized by the fact that it could successfully be applied to both X- and L-band images. A comparison of SVM on X- and L-band images showed slightly better performance of TSX (0.81 AUC) than ALOS (0.77 AUC). Also a test, for which only buildings with a footprint of larger than 160 m² were considered to account for the lower spatial resolution of ALOS, indicated an almost negligible performance difference (0.02 AUC). When looking at the final results over the whole study area it can be seen that the ALOS L-band tends to significantly underestimate the number of changed buildings. This is likely to be related to the characteristics of the ALOS L-band, which compared to the TSX X-band shows low backscatter intensity in dense residential areas. It may thus negatively affect the detection rate of damaged houses with small intensity changes. Even though the acquisition dates of ALOS and TSX were selected as close as possible to each other, still the t2 images are 26 days apart from each other. In order to avoid a possible bias by real-world

changes that may have occurred during this difference period, further tests with closer acquisition dates should be carried out.

(5) How does an SVM compare to thresholding change detection?

The proposed machine learning approach performed significantly better than a thresholding change detection over all tested training-testing scenarios and thresholds (Figure 9, Table 6). Testing of the thresholding approach has been carried out over all possible thresholds in order to avoid bias by a particular threshold approximation method. Also comparing the change detection resulting from the best threshold over all iterations with the results obtained from a trained SVM showed superior performance of SVM with respect to thresholding on all image types and acquisition dates. The window of possible thresholds that yield comparable results is, moreover, narrow and could not be identified by a simple threshold approximation method (Figure 9).

(6) Other considerations

Different kernel functions have been used within this study and were optimized for each classification by means of cross-validation and grid-search. For most of the classifications a radial basis function has been selected by the optimization procedure. Even though the influence of kernel functions and other SVM parameters on the performance of the change detection has not specifically been evaluated by this study, it is significant as has been shown by, for example, Camps-Valls [21] and should be considered by any study attempting to use SVM for change detection on SAR imagery.

Using an object-based approach with external building footprint data as computational unit needs to account for the SAR image geometry. Therefore, footprints should be adjusted accordingly before further analysis, which was done in this study by shifting them laterally. In case no independent building footprint data are available, the influence of the image segmentation on the classification performance should be evaluated more specifically as it can potentially have a significant impact on the classification [25].

6. Conclusions

This study evaluated the performance of a Support Vector Machine (SVM) classifier to learn and detect changes in single- and multi-temporal X- and L-band Synthetic Aperture Radar (SAR) images under varying conditions. The apparent drawback of a machine learning approach to change detection is largely related to the often costly acquisition of training data. With this study, however, we were able to demonstrate that given automatic SVM parameter tuning and feature selection, in addition to considering several critical decisions commonly made during the training stage, the costs for training an SVM can be significantly reduced. Balancing the training samples between the change classes led to significant improvements with respect to random sampling. With a large enough training sample size (>400 sample per class), moreover, good generalization ability of the SVM at high accuracy levels (>0.80 F1 score) can be achieved. The experiments further indicate that it is possible to transfer a locally trained classifier to different areas with only minor loss of performance. The classifier performed well in the case of a simple binary classification task, but distinguishing more complex change types, such as tsunami-related changes that do not directly affect the roof structure, lead to a significant performance decrease. The best results were observed on image pairs with a larger temporal baseline. A clear performance decrease could be observed for single-date change classifications based on post-event images with respect to a multi-temporal classification. Since the overall performances are still good (>0.67 F1 score) such a post-event classification can be a useful approach for the case when no suitable pre-event images are available or when very rapid change assessments need to be carried out in emergency situations. A direct comparison of SVM on X- and L-band images showed better performance of TSX than ALOS. Over all training and testing scenarios, however, the difference is minor (0.04 AUC). Compared to thresholding change detection, the machine learning approach performed significantly better on the tested image types and acquisition dates. This conclusion holds independent on the selected threshold approximation approach.

The benefits of a machine learning approach lie mainly in the fact that it allows a partitioning of a multi-dimensional change feature space that can provide more diverse information about the changed objects with respect to a single feature or a composed feature index. Also, the definition of the decision boundary by selection of relevant training samples is more intuitive from a user's perspective than fixing a threshold value. Moreover, the computation of Shannon entropy values from the soft answers of the SVM classifier proved to be a valid confidence measure that can aid the identification of possible false classifications and could support a targeted improvement of the change maps either by means of an active learning approach or by manual post-classification refinement. In case of a disaster, the proposed approach can be intuitively adjusted to local conditions in terms of applicable types of changes which may vary depending on the hazard type, the geographical region of interest and the objects of interest. Such an adjustment would commonly involve visual inspection of optical imagery from at least a subset of the affected area in order to acquire training samples. Detailed damage mapping as part of the operational emergency response protocols is largely done by human operators based on visual inspection of very high resolution optical imagery [39]. Thus, the findings of this study can be used to design a machine learning application that is coupled with such mapping operations and that can provide regular rapid estimates of the spatial distribution of most devastating changes while the detailed but more time-consuming damage mapping is in progress. The estimates from the learning machine can be used to guide and iteratively prioritize the manual mapping efforts. An example of a similar approach to prioritize data acquisition in order to improve post-earthquake insurance claim management is given in Pittore et al. [40].

Ongoing and future research efforts focus on extending the change feature space by texture, coherence and curvelet features [7,38] on testing the transferability of trained learning machines across image scenes and detection tasks, on further comparisons of the proposed method with other kernel-based methods [21,24] and on developing a prototype application for iterative change detection and mapping prioritization.

Acknowledgments: The authors would like to thank the editors and the anonymous reviewers for suggestions which helped to improve this paper. This study was supported by a research scholarship funded by the Japan Society for the Promotion of Science (JSPS). The ALOS images are the property of JAXA. The TerraSAR-X images used in this study are property of DLR and were provided through the Geohazard Supersites Service.

Author Contributions: Wen Liu and Fumio Yamazaki supervised this research activity and provided guidance and suggestions for the analysis and their revisions during the writing of the paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Lu, D.; Mausel, P.; Brondízio, E.; Moran, E. Change detection techniques. *Int. J. Remote Sens.* **2004**, *25*, 2365–2401. [[CrossRef](#)]
2. Dong, L.; Shan, J. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS J. Photogramm. Remote Sens.* **2013**, *84*, 85–99. [[CrossRef](#)]
3. Plank, S. Rapid damage assessment by means of multi-temporal SAR: A comprehensive review and outlook to Sentinel-1. *Remote Sens.* **2014**, *6*, 4870–4906. [[CrossRef](#)]
4. Brett, P.T.B.; Guida, R. Earthquake damage detection in urban areas using curvilinear features. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4877–4884. [[CrossRef](#)]
5. Bruzzone, L.; Prieto, D. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. *IEEE Trans. Image Process.* **2002**, *11*, 452–466. [[CrossRef](#)] [[PubMed](#)]
6. Hachicha, S.; Chaabane, F. On the SAR change detection review and optimal decision. *Int. J. Remote Sens.* **2014**, *35*, 1693–1714. [[CrossRef](#)]

7. Wieland, M.; Wen, L.; Yamazaki, F.; Sasagawa, T. A comparison of change features from multi-temporal SAR images for monitoring the built-environment in disaster situations. In Proceedings of the 59th Autumn Conference of the Remote Sensing Society of Japan, Nagasaki, Japan, 26–27 November 2015.
8. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [[CrossRef](#)]
9. Chen, G.; Hay, G.J.; Carvalho, L.M.T.; Wulder, M.A. Object-based change detection. *Int. J. Remote Sens.* **2012**, *33*, 4434–4457. [[CrossRef](#)]
10. Uprety, P.; Yamazaki, F.; Dell’Acqua, F. Damage Detection Using High-Resolution SAR Imagery in the 2009 L’Aquila, Italy, Earthquake. *Earthq. Spectra* **2013**, *29*, 1521–1535. [[CrossRef](#)]
11. Liu, W.; Yamazaki, F.; Gokon, H.; Koshimura, S. Extraction of tsunami-flooded areas and damaged buildings in the 2011 Tohoku-oki earthquake from TerraSAR-X intensity images. *Earthq. Spectra* **2013**, *29*, S183–S200. [[CrossRef](#)]
12. Quin, G.; Pinel-Puysegur, B.; Nicolas, J.M.; Loreaux, P. MIMOSA: An automatic change detection method for sar time series. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5349–5363. [[CrossRef](#)]
13. Bovolo, F.; Bruzzone, L. A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1658–1670. [[CrossRef](#)]
14. Liu, W.; Yamazaki, F. Urban monitoring and change detection of central Tokyo using high-resolution X-band SAR images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011.
15. Bruzzone, L.; Prieto, D. Automatic analysis of the difference image for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* **2002**, *38*, 1171–1182. [[CrossRef](#)]
16. Bazi, Y.; Bruzzone, L.; Melgani, F. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 874–887. [[CrossRef](#)]
17. Bouchaffra, D.; Cheriet, M.; Jodoin, P.M.; Beck, D. Machine learning and pattern recognition models in change detection. *Pattern Recognit.* **2015**, *48*, 613–615. [[CrossRef](#)]
18. Chan, J.; Chan, K.; Gar, A. Detecting the nature of change in an urban environment: A comparison of machine learning algorithms. *Photogramm. Eng. Remote Sens.* **2001**, *67*, 213–225.
19. Im, J.; Jensen, J. A change detection model based on neighborhood correlation image analysis and decision tree classification. *Remote Sens. Environ.* **2005**, *99*, 326–340. [[CrossRef](#)]
20. Shah-Hosseini, R.; Homayouni, S.; Safari, A. A hybrid kernel-based change detection method for remotely sensed data in a similarity space. *Remote Sens.* **2015**, *7*, 12829–12858. [[CrossRef](#)]
21. Camps-Valls, G.; Gomez-Chova, L.; Munoz-Mari, J.; Rojo-Alvarez, J.L.; Martinez-Ramon, M. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1822–1835. [[CrossRef](#)]
22. Le Saux, B.; Randrianarivo, H. Urban change detection in SAR images by interactive learning. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Melbourne, Australia, 21–26 July 2013.
23. Gokon, H.; Post, J.; Stein, E.; Martinis, S.; Twele, A.; Muck, M.; Geiss, C.; Koshimura, S.; Matsuoka, M. A Method for detecting buildings destroyed by the 2011 Tohoku Earthquake and Tsunami using multitemporal Terrasar-X Data. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1277–1281. [[CrossRef](#)]
24. Jia, L.; Li, M.; Wu, Y.; Zhang, P.; Chen, H.; An, L. Semisupervised SAR image change detection using a cluster-neighborhood kernel. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1443–1447. [[CrossRef](#)]
25. Wieland, M.; Pittore, M. Performance evaluation of machine learning algorithms for urban pattern recognition. *Remote Sens.* **2014**, *6*, 2912–2939. [[CrossRef](#)]
26. Liu, W.; Yamazaki, F.; Matsuoka, M.; Nonaka, T.; Sasagawa, T. Estimation of three-dimensional crustal movements in the 2011 Tohoku-Oki, Japan, earthquake from TerraSAR-X intensity images. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 637–645. [[CrossRef](#)]
27. Reddy, B.S.; Chatterji, B.N. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **1996**, *5*, 1266–1271. [[CrossRef](#)] [[PubMed](#)]
28. Damage Database of the Tohoku Earthquake and Tsunami 2010. Available online: http://fukkou.csis.u-tokyo.ac.jp/dataset/list_all (accessed on 11 August 2015).

29. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 2000.
30. Shannon, C.E. Communication in the presence of noise. *Proc. IEEE* **1998**, *86*, 447–457. [[CrossRef](#)]
31. Platt, J. Probabilistic outputs for support vector machines and comparison to regularized likelihood method. *Adv. Large Margin Classif.* **1999**, *3*, 61–74.
32. Wu, T.F.; Lin, C.J.; Weng, R.C. Probability estimates for multi-class classification by pairwise coupling. *J. Mach. Learn. Res.* **2004**, *5*, 975–1005.
33. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [[CrossRef](#)]
34. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
35. Wieland, M.; Torres, Y.; Pittore, M.; Belen, B. Object-based urban structure type pattern recognition from Landsat TM with a Support Vector Machine. *Int. J. Remote Sens.* **2016**, *37*, 4059–4083. [[CrossRef](#)]
36. He, H.; Ma, Y. *Imbalanced Learning: Foundations, Algorithms, and Applications*; Wiley-IEEE Press: New York, NY, USA, 2013.
37. Yamazaki, F.; Iwasaki, Y.; Liu, W.; Nonaka, T.; Sasagawa, T. Detection of damage to building side-walls in the 2011 Tohoku, Japan earthquake using high-resolution TerraSAR-X images. *Proc. SPIE* **2013**. [[CrossRef](#)]
38. Schmitt, A.; Wessel, B.; Roth, A. An innovative curvelet-only-based approach for automated change detection in multi-temporal SAR imagery. *Remote Sens.* **2014**, *6*, 2435–2462. [[CrossRef](#)]
39. Voigt, S.; Kemper, T.; Riedlinger, T.; Kiefl, R.; Scholte, K.; Mehl, H. Satellite image analysis for disaster and crisis-management support. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1520–1528. [[CrossRef](#)]
40. Pittore, M.; Wieland, M.; Errize, M.; Kariptas, C.; Güngör, I. Improving post-earthquake insurance claim management: A novel approach to prioritize geospatial data collection. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2401–2427. [[CrossRef](#)]



© 2016 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 (<http://creativecommons.org/licenses/by/4.0/>).