Article

Spatial Prediction of Aftershocks Triggered by a Major Earthquake: A Binary Machine Learning Perspective

Sadra Karimzadeh [1],[*] , **Masashi Matsuoka** [2] , **Jianming Kuang** [3] and **Linlin Ge** [3]

1   Department of Remote Sensing and GIS, University of Tabriz, Tabriz 5166616471, Iran
2   Department of Architecture and Building Engineering, School of Environment and Society, Tokyo Institute of Technology 4259-G3-2 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan; matsuoka.m.ab@m.titech.ac.jp
3   School of Civil and Environmental Engineering, The University of New South Wales (UNSW), Sydney, NSW 2052, Australia; jianming.kuang@geos.org.au (J.K.); linlin.ge@geos.org.au (L.G.)
*   Correspondence: sadra.karimzadeh@gmail.com or sa.karimzadeh@tabrizu.ac.ir

check for updates

Abstract: Small earthquakes following a large event in the same area are typically aftershocks, which are usually less destructive than mainshocks. These aftershocks are considered mainshocks if they are larger than the previous mainshock. In this study, records of aftershocks (M > 2.5) of the Kermanshah Earthquake (M 7.3) in Iran were collected from the first second following the event to the end of September 2018. Different machine learning (ML) algorithms, including naive Bayes, k-nearest neighbors, a support vector machine, and random forests were used in conjunction with the slip distribution, Coulomb stress change on the source fault (deduced from synthetic aperture radar imagery), and orientations of neighboring active faults to predict the aftershock patterns. Seventy percent of the aftershocks were used for training based on a binary ("yes" or "no") logic to predict locations of all aftershocks. While untested on independent datasets, receiver operating characteristic results of the same dataset indicate ML methods outperform routine Coulomb maps regarding the spatial prediction of aftershock patterns, especially when details of neighboring active faults are available. Logistic regression results, however, do not show significant differences with ML methods, as hidden information is likely better discovered using logistic regression analysis.

Keywords: aftershocks; machine learning; Kermanshah earthquake; Coulomb stress; synthetic aperture radar

1. Introduction

Machine learning (ML), as is implied by its name, represents a category of algorithms used within a computer system to "learn" based on previous experiences. ML is utilized to explore hidden features or relationships among datasets and to construct algorithms to learn how to address other dependent or independent datasets [1,2]. ML can classify datasets and make regressions to estimate some values. For the dataset classification task, we have two input categories: response variables (dependent variables), which can be either qualitative or quantitative, are the variables that the model is trying to predict for other datasets; and explanatory variables, which are used to explain various aspects of data that are gathered from different acquisition methods. A response variable should always be coupled with one or more explanatory groups to run an ML model. ML models are becoming popular in seismology, with a large range of applications from phase picking to ground motion prediction [3].

Previous studies have applied several statistical and ML algorithms for the classification of aftershocks. For example, Kortström et al. [4] classified blast-related and nature-related aftershocks and sources of noise using an automated support vector machine (SVM) technique in regions that suffer

from sparse seismic networks; their results showed that man-made (explosions) and natural (seismic) aftershocks can be identified with a high level of reliability using their approach. Discrimination analysis was used by Che et al. [5] to identify an explosion-induced event in North Korea. In addition, Lyubushin et al. [6] classified seismic records acquired from the Aswan Dam region in Egypt as either natural events or blasts using the spectral support widths method. Rouet-Leduc et al. [7] investigated laboratory earthquakes to predict the failure time of a fault based on acoustical information and an ML method called random forest (RDF) classification. A few works have attempted to predict the locations of earthquakes based on an adaptive neural fuzzy inference system (ANFIS) and a supervised radial basis function (RBF) network [8,9]. One reason for this paucity of research is the complexity associated with the extraction of dependent and independent variables within ML algorithms. Recently, however, DeVries et al. [10] gathered a comprehensive database of nearly 131,000 mainshock–aftershock pairs to use in neural network (NN) algorithms to predict aftershocks from independent datasets and found that the aftershock pattern derived from deep learning analysis can explain the shear stress changes of the surrounding areas in addition to other independent components (e.g., first invariant, second invariant, and third invariant parameters) in the stress change matrix. Mignan and Broccardo pointed out that for aftershock prediction, utilizing only one neuron can provide an accuracy as acceptable as that achieved by a deep neural network [11].

In this paper, we focus only on the spatial prediction of recorded aftershocks (M > 2.5) regardless of their depth and magnitude following the Kermanshah earthquake (M 7.3) in western Iran using simple/logistic regression analyses and four popular ML algorithms (i.e., naive Bayes (NB), k-nearest neighbors (KNN), SVM, and RDF algorithms), and their corresponding stress changes, slip distributions, and Euclidean distances between the source and receiver faults.

2. Pilot Area and Data

On 12 November 2017, a severe earthquake (M 7.3) struck western Iran, leaving 620 deaths in its wake [10,11]. According to the Iranian Seismological Center (ISC), the epicenter was identified at 34.772° latitude and 45.762° longitude, near the Iran–Iraq border, with a depth of 19 km (Figure 1a). According to the United States Geological Survey (USGS) and the Global Centroid Moment Tensor (CMT) catalog, the focal mechanism of the event exhibited oblique thrust faulting with a NNW–SSE orientation [12–15]. The results of differential interferometric synthetic aperture radar (DInSAR) analysis of datasets acquired from the Advanced Land Observing Satellite ALOS-2 and Sentinel-1 revealed that the maximum displacements in the line-of-sight (LOS) direction for the ascending and descending pairs of both satellites were approximately 90 cm and 49 cm, respectively [13]. Due to sparse vegetation, the study area is ideal for identifying and extracting coseismic deformation and slip distribution information with a low level of noise using the InSAR technique and SAR data acquired from both Sentinel-1 (C-band) and ALOS-2 (L-band) [16–18].

2.1. Aftershocks

We collected aftershocks from the ISN in an area of approximately 110 km × 120 km near the Iran–Iraq border from the first second following the M 7.3 event to the end of September 2018. As shown in Figure 1a, more than 1700 aftershock occurred in the area near or around the High Zagros Fault (HZF) and the Zagros Mountain Front Fault (MFF); among them, five aftershocks had a magnitude exceeding 4.8. The statistical plots in Figure 1b show that the aftershock depths were relatively shallow with mean and median depths of 9.3 and 9.2 km, respectively. The mean and median magnitudes of the aftershocks were 3 and 2.8, respectively. Figure S1, available in the electronic supplement to this article, shows the root-mean-square (RMS) travel times of the recorded aftershocks in seconds and the azimuthal gaps between the recorded aftershocks in degrees. Generally, a smaller RMS of travel time residuals and a smaller azimuthal gap can provide a more reliable estimation of the horizontal position of an aftershock. The majority of the aftershocks (55%) had an azimuthal gap of less than 180 degrees.

However, those aftershocks with azimuthal gaps greater than 180 degrees were weighted during the ML spatial prediction.
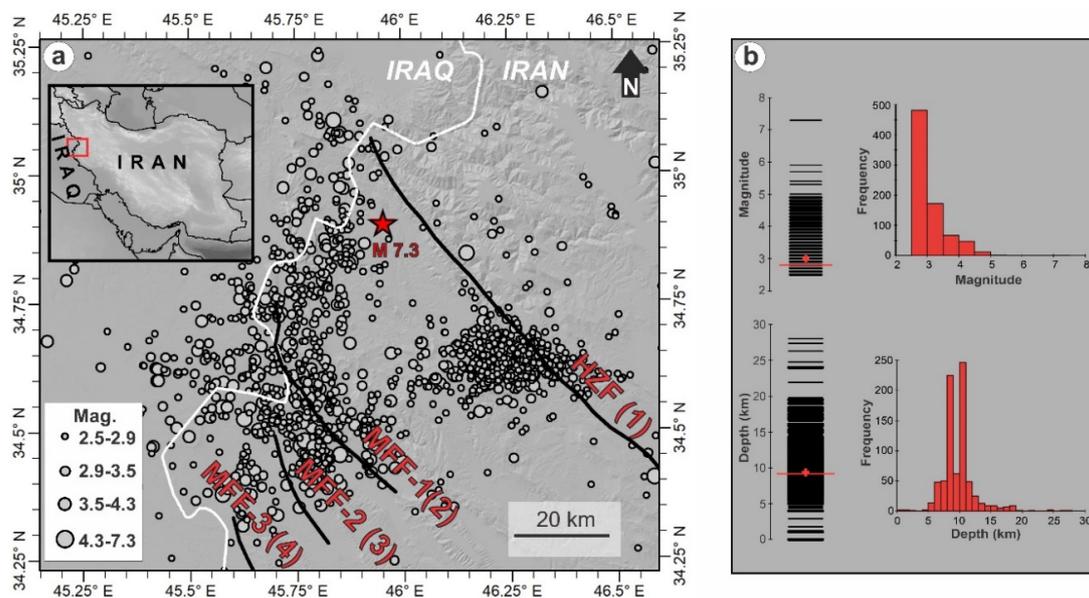

Figure 1. (**a**) Shaded relief of the Shuttle Radar Topography Mission (SRTM) version 4 digital elevation model near the Iran–Iraq border together with aftershocks (gray circles) on different scales. The black lines delineate recognized faults. The dashed white line represents the Iran–Iraq border. The red star shows the location of the main shock (M 7.3) according to the Iranian Seismological Network (ISN). The inset map shows the study area near the Iran–Iraq border. (**b**) Strip plots and histograms of the magnitudes and depths of the aftershocks. The red lines and red plus symbols show the median and mean values, respectively.

2.2. Slip Distribution, Coulomb Stress Change, and Fault Maps

In this study, we used maps of the slip distribution and Coulomb stress change associated with the Kermanshah Earthquake, together with fault maps of the surrounding region, as the explanatory variables in the ML algorithms [16]. The DInSAR phase unwrapping results of ALOS-2 data in ScanSAR mode (ascending orbit no. 180 and descending orbit no. 71) and Sentinel-1 data in interferometric wide swath mode (ascending orbit no. 72 and descending orbit nos. 6 and 79) were first subsampled and then inverted using a finite elastic half-space dislocation model (Table S1, available in the electronic supplement to this article) [16,19]. The inversion itself was performed in two steps. The first step involved a nonlinear inversion, and all the source parameters (i.e., the length, width, depth, longitude, latitude, strike, dip, rake, and slip) were initialized with the CMT solution with a uniform slip model. In the nonlinear inversion step, to find the minimum value of the function, the Gauss–Newton iteration and the Levenberg–Marquardt (LM) least-squares algorithms were applied by considering multiple random restarts [16,20,21]. The second step was a linear inversion, in which the fault geometry inverted from the nonlinear step was fixed to estimate the slip distribution along the fault plane. For this purpose, the fault plane was divided into 320 small segments (5 km × 5 km) in the strike and dip directions. Once the slip distribution maps for each dataset were inverted, a joint inversion was performed using Sentinel-1 and ALOS-2 datasets (both ascending and descending pairs) to optimize the results (Figure 2a) for the source fault while assuming that its length and width in the directions of the dip and strike were 100 km and 80 km, respectively [16]. This area was used for the procedure described in the next section to generate a mask for the aftershocks because the slip distribution beyond the inversion area was unknown. The slip distribution model deduced from the joint inversion was then used as the source map to calculate the amount of stress triggered by the mainshock. The Coulomb stress change on the source fault and its surrounding faults is an explanatory parameter that is relevant

to the distributions of aftershocks and earthquakes [22]. The Coulomb stress change (Δ*CFF*) can be defined from the slip distribution model and the friction coefficient in addition to Skempton's coefficient, which ranges between 0 and 1 depending on the soil situation (0 indicates dry soil, and 1 reflects fully saturated soil):
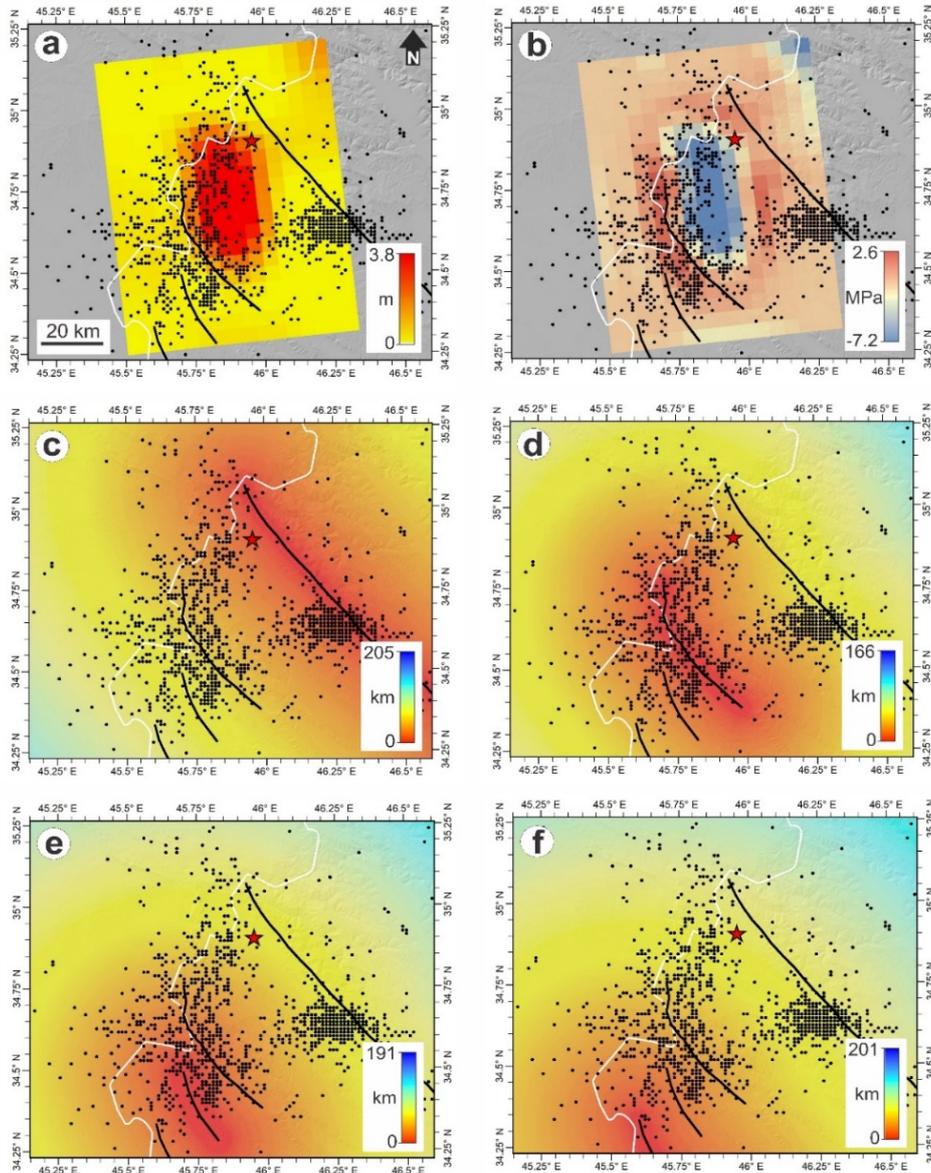

Figure 2. (**a**) Slip distribution map of the M 7.3 event deduced from joint inversion analysis of the Sentinel-1 and ALOS-2 data. (**b**) Coulomb stress change map of the M 7.3 event deduced from the subsampled phase unwrapped displacement results of the Sentinel-1 and ALOS-2 datasets in an elastic half-space [17]. (**c**–**f**) Euclidean distance maps of the neighboring faults.

$$\Delta CFF = \Delta\tau + \mu\cdot\left(\Delta\sigma - \beta\cdot\left(\frac{T}{3}\right)\right) \tag{1}$$

where $\Delta\tau$ is the variation in the shear stress, μ is the friction coefficient, $\Delta\sigma$ is the variation in the normal stress, β is Skempton's coefficient, and T is the stress matrix. Based on previous applications of Coulomb stress changes, an increase in the Coulomb stress change after an event can indicate a high failure risk; similarly, when the Coulomb stress change decreases, the risk of failure also decreases [23–26]. It should be noted that some studies have disputed the application of the stress change to explain aftershock sequences [27,28]. The Coulomb stress changes in this study were generated by assuming

that the friction coefficient μ and shear modulus were 0.4 and 3×10^{10} N/m², respectively. In addition, two segments of the MFF, namely, MFF-1 and MFF-2 (faults 2 and 3), and one segment of the HZF (fault 1) were considered receiver faults. The traces and orientations of the faults ($\approx 320°$) and their dip angles (NE 30–60°) were extracted from previous studies [29]. Figure 2b shows that the Coulomb stress change after the Kermanshah Earthquake ranged between -7.2 MPa (at a depth of 13 km) to $+2.6$ MPa (at a depth of 9 km). As a consequence of the Coulomb stress decrease in the upper part of the source fault plane (where the mainshock was located), the number of aftershocks around that area was relatively low. In contrast, the positive stress changed in the lower part of the source fault plane and all around faults 1 and 2, resulting in more aftershocks. For each pixel grid, we calculated the Euclidean distance to the specified source, where the pixel value of a legitimate source is 0. The Euclidean distance represents the distance from each location to the corresponding fault. Figure 2c–f shows the Euclidean distance map for each fault (i.e., the HZF, MFF-1, MFF-2, and MFF-3). This concept will be used in the next section to input the Euclidean distance of each aftershock on each fault into an applicable level in the ML algorithms.

3. Binary Aftershock Map and Classification

We applied four different ML algorithms—NB, KNN, SVM, and RDF algorithms—to assess how well each one can predict the locations of the aftershocks. Predicting the location of an aftershock using its geographic coordinates is rather difficult. ML algorithms tend to classify objects or features in a predictive analytical way so that an assumed function, such as (f), can project optimally independent variables, such as (x), to an output variable, such as (Y), in a functional state following $Y = f(x)$. To follow this structure, the geographic coordinates of the aftershocks were converted into a binary grid map; that is, if an aftershock occurred on a given grid, we assigned a "yes" statement; otherwise, we assigned a "no" statement.

As shown in Figure 3a, we bound the aftershocks within length and width dimensions of 100 km × 80 km where the InSAR modeling and Coulomb stress change results were available. We also assumed a point buffer of approximately 1.5 km with a value of 0 for the "no" condition and a value of 1 for the "yes" condition, and we assigned a weight for each aftershock using the RMS of the travel time residuals and the azimuthal gap information. Here, 4547 points out of 5331 points had a value of 0, and 784 points had a value of 1. The number of response variables in group 0 was six times more than that in group 1. To maintain a balance between these two groups, we selected the same number of samples from each group to train the four ML algorithms. Consequently, we randomly selected 530 points from each group (almost 70% of all aftershocks) to train all the algorithms. We considered the response variable and explanatory variables as qualitative and quantitative parameters, respectively, and created the predicted model for all selected points. Put simply, values of 0 and 1 qualitatively described a pixel as being an "aftershock" or "not an aftershock," respectively, whereas the slip distribution, Coulomb stress change, and Euclidean distance were represented by their specific pixel values. We defined a constraint for the Euclidean distance based on our prior knowledge about the locations of the 530 randomly selected aftershocks. For HZF (1), MFF-1 (2), MFF-2 (3), and MFF-3 (4), the accumulated Euclidean distances from the 530 aftershocks were 31,911,736, 24,493,352, 39,645,498, and 56,556,613 m, respectively (Figure 3b). These distances were an indicator of the possibility and closeness of the aftershocks to each fault, meaning that a fault with a smaller accumulated Euclidean distance was more likely to be triggered by the mainshock. Here, faults 2 and 1 had the lowest accumulated distances from the aftershocks. Thus, to perform a fair analysis, we considered only faults 1 and 2. Although it was possible to consider faults 3 and 4 as well, the ML algorithms become confused when a meaningful relationship does not exist among the Euclidean distance, Coulomb stress change, and slip distribution.
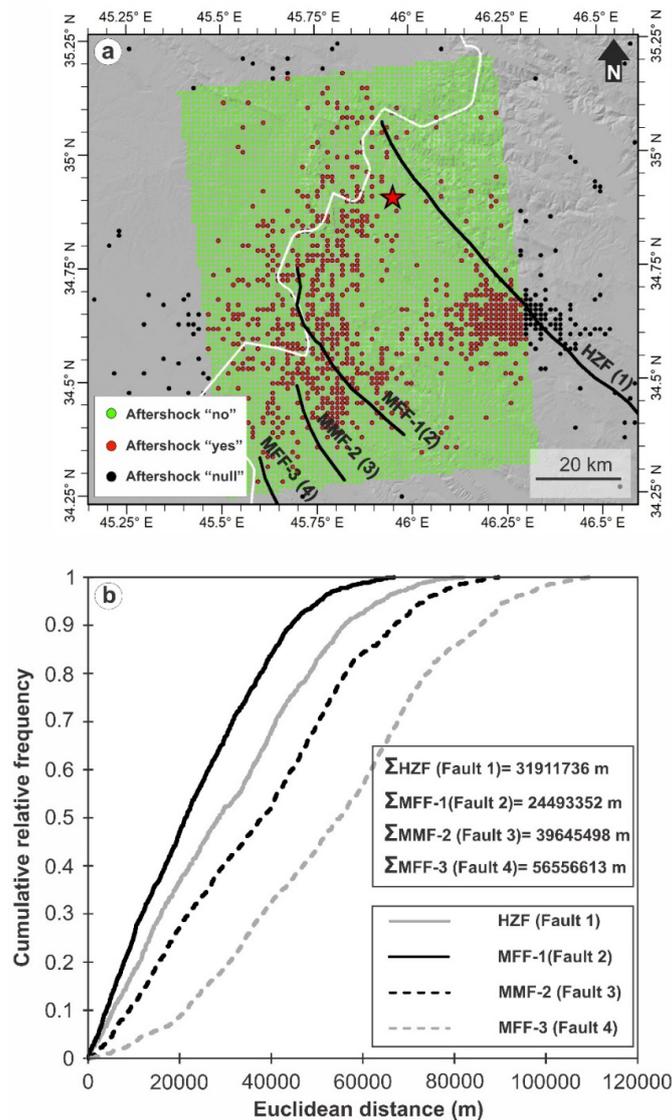
Figure 3. (**a**) Binary aftershock map of the study area with a grid resolution of approximately 1.5 km. Green and red dots represent "no aftershock" and "aftershock" statements, respectively. Black dots were not considered for the initial learning step due to the lack of overlap with the footprint of the Coulomb stress change map. (**b**) Accumulated Euclidean distance of each fault from the aftershocks.

We first applied the NB algorithm, which is a probabilistic classifier that makes classifications with a "naive" assumption that each input variable (either continuous or categorical) is independent. Since we had the same number of observations from groups 0 and 1, the prior probability of the training set was equal. We trained the algorithm by assuming that the quantitative variables followed a normal distribution and by considering a smoothing parameter that did not allow zero probabilities in subsequent computations. The NB algorithm calculated the probabilities of events 0 and 1, the probability of event 0 given 1, and the probability of event 1 given 0 based on Bayes' theorem (Figure 4). The NB algorithm was strong; if the probabilities of groups 0 and 1 were not equal, the final step in the aftershock classification was performed by combining the prior probability and the likelihood into a posterior probability. Second, we applied the KNN algorithm, which is applicable when prior knowledge is limited or lacking. A simple KNN algorithm works based on measuring the similarity among specifications and the average of the values of its k-nearest neighbors in a multidimensional environment (Figure 4). Here, the KNN model was formed based on the Euclidean distances of the response variables, and an inverse-squared distance method was used to generate

weights. The third algorithm that we trained in this study was the SVM, which was first presented by Vapnik [30]. The SVM classifier has attracted growing interest in recent years with its increasingly frequent application to remote sensing data and its use for predicting phenomena such as landslides and building damage [31–33]. The SVM is a nonparametric supervised classifier that segregates variables into a number of previously defined classes in a way that is consistent with the training dataset. The SVM classifier generally enables kernel functions to find a hyperplane or line near the extreme points, and kernel functions (e.g., linear, polynomial, Gaussian, and sigmoid) are advantageous because we can transfer datasets from a lower (e.g., 1D) dimension to a higher (e.g., 2D) dimension to simplify the construction of a hyperplane (Figure 4). Here, we applied only a linear kernel, which is the simplest model for separating data. To solve the SVM optimization problem, the sequential minimal optimization (SMO) algorithm was considered with a corresponding margin parameter (C) that controlled low training and low testing errors [34]. The value of C was set to 1; if C has a value greater than 1, the margin hyperplane selected by the SMO will be relatively small. The final classifier used herein was the RDF classifier, which is a more advanced type of decision tree classifier. In the RDF method, instead of a few trees, hundreds (or thousands) of trees are utilized, and thus, the weakness of each tree is compensated for by the others. Similar to the NB, KNN, and SVM algorithms, the RDF classifier also allowed us to predict the classes into which the observations fell on the basis of either quantitative or qualitative variables. The RDF algorithm can efficiently handle large datasets, and the number of trees can be selected depending on the size of the dataset. In general, the more trees there are in the forest, the more robust the prediction. The binary classification and regression trees (CART) method introduced by Breiman et al. [35] was used in this study. We defined 500 trees for 1060 samples (530 "yes" and 530 "no"), and the observations were chosen randomly; the same sample could be chosen several times using a bagging method, which means that we chose n' samples (i.e., the number of samples in a bag) for each bag by considering that the repeated samples were inevitable in each bag and that the number of whole samples (n) was always larger than n'. Thus, once we had a different ensemble of different learning algorithms in the RDF, we had 500 different models, and their corresponding average was used for the prediction, where the RDF classifier performed the prediction based on the most voted trees over all the other trees in the forest (Figure 4).

Alternatively, we also performed a logistic regression analysis suggested by Mignan and Broccardo to consider fault 1 and 2 and Coulomb stress values as independent variables. A logistic regression analysis tries to find the probability of being 0 or 1 with defined constraints to avoid violation of dummy values, so the interpretations in the logistic regression analysis are clearer. The logistic regression analysis can elucidate the relationship of binary values (0 and 1) with independent values that can be in different forms (e.g., nominal, ordinal, etc.). The logistic regression can be formed as follows:

$$z(p) = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_i x_i = \ln\left[\frac{p}{1-p}\right] \tag{2}$$

$$p = \frac{1}{1 + e^{-y(p)}} \tag{3}$$

where z is the dependent variable ("yes" or "no"), b_0 is a constant value, and b_1 through b_i are coefficients. $x_1, x_2, \ldots, x_i$ are independent variables (fault 1, fault 2 and Coulomb stress) calculated for each grid point, p is the probability of existence of an aftershock, and $1 - p$ is the probability of absence of an aftershock. Therefore, we could calculate the odds ratios as follows:

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ an\ aftershock}{probability\ of\ absence\ of\ an\ aftershock} = e^{b_0} \times e^{b_1 x_1} \times e^{b_2 x_2} \times \ldots e^{b_i x_i} \tag{4}$$

We could calculate $z(p)$ from a relationship between the dependent and independent variables and accordingly define a constraint on the sum of all possibilities to find b_0 and other coefficients (b_i).
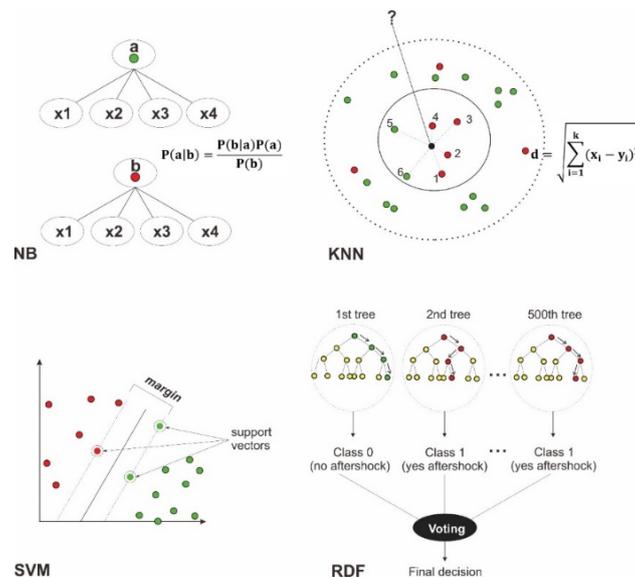
Figure 4. Schematic illustration of four ML methods. Red and green circles are "aftershock" and "not an aftershock," respectively. In the NB model, x_1, x_2, $\ldots$ are explanatory variables. The KNN model shows a simple illustration of how the close circles in a defined area can predict an unknown circle (black circle). The SVM model shows the contribution of the support vectors in the hyperplane for the prediction. The RDF model (formed using 500 decision trees) shows how the final classification of each tree is utilized in a majority voting process.

4. Results and Discussion

To evaluate the accuracies of the different models, the prediction values obtained from all the algorithms were rounded to either 0 or 1, except for those generated by the SVM, which did not provide predictions in a standard formulation; for the SVM, an absolute value (either 0 or 1) was used [33]. The overall accuracy (OA) of each classifier for the 5331 points within the study area is given in Table 1. Except for the KNN results, the OAs of the other classifiers were improved by the contribution from fault 1. The OA of the KNN algorithm was not improved by the contribution of auxiliary information, likely because the KNN is a lazy learner that uses only the training data for classification, which does not promote its learning capability. The KNN algorithm basically found the Euclidean distances of the nearest neighbors to a new instance within the training data to predict the corresponding class of each new instance. Accordingly, as shown in Figure 5e,f, the KNN patterns were more dependent on the Euclidean distances of the variables, which made the prediction difficult. The prediction patterns deduced from the SVM algorithm (Figure 5g–i) were also similar to those predicted from the NB classifier. The achieved accuracy was improved when the Euclidean information of fault 1 was considered. However, with the contribution of fault 2, the OA accuracy decreased from 74% to 57%. The OA accuracy of the RDF classifier also improved from 73% up to 76% and 75% for fault 1 and fault 2, respectively. As shown in Figure 5c,f,I,l, the predicted patterns were highly affected by the Euclidean values of fault 2. Thus, the cluster of aftershocks that formed near fault 1 was ignored by the classifiers. However, the accuracy of each pattern was lowered when fault 2 was utilized as an explanatory variable. A cursory glimpse of Table 1 reveals that the OA of the NB algorithm was better than those of the other classifiers; an OA of 78% was attainable without the contribution of any fault. The OA of the logistic regression method with both fault 1 and 2 was 74%, which was higher than simple regression for only Coulomb stress values (69%). However, the OA could not always fully address the accuracy of the aftershock locations. Additionally, as shown in Figure 5a,b,g,h, the NB and SVM classifiers with or without fault 1 presented similar patterns in the Coulomb map (Figure 5b). For a fair assessment, we performed receiver operating characteristic (ROC) analysis. ROC curves are efficient for diagnostic and sensitivity tests. We plotted the true positive rate (sensitivity) versus the

false positive rate (1 − specificity) for the different classifiers and conditions (Figure 6). Each point on the ROC curve represents a corresponding cutoff value of the sensitivity versus the specificity. For the Coulomb map, we applied a simple classification using a regression method (yellow lines in Figure 6). All the applied algorithms showed a better performance than the Coulomb failure criterion alone, which was deduced from a simple discrimination analysis. The area under the curve (AUC) criterion was employed to address how well the explanatory parameters were able to differentiate an aftershock ("yes") from a non-aftershock ("no"). For tests better than a random assignment, the AUC should be higher than 0.5 with a maximum of 1. The weakest and strongest performances belonged to the Coulomb regression and RDF classification models "with fault 1," in which the AUC values were 0.432 and 0.87, respectively (Figure 6). For the NB classifier, as shown in Figure 6a, the classifier "with fault 2" showed a better performance. For all classifiers except the KNN, the addition of the surrounding faults improved the AUC values. This improvement was more evident with the RDF, for which the AUC increased from 0.798 up to 0.869 and 0.87 with faults 1 and 2, respectively. Although the OA of the NB classifier was higher than those of the other classifiers, in terms of the AUC values, the KNN and RDF classifiers showed better performances than the NB algorithm. The performances of the KNN and RDF algorithms for classification at different thresholds were better, which meant a higher true positive rate and a lower false positive rate could be achieved. The NB and RDF models were able to fit our aftershock data well every single time, which meant that our models were high-variance models with overfitting. Therefore, high variance indicated that the algorithms were probably too flexible to find outliers, which is known to not generalize well. Furthermore, independent datasets are needed to validate the ML algorithms in future works. In Figure 6d, the AUC value of the logistic regression method (pink line) was 0.850, which was higher than most of the used ML algorithms. This also implied that the logistic regression method likely not only yielded the coefficient size, but also its negative or positive direction. This aspect is different than the RDF model or other decision tree methods. Such ML methods tell us which predictor is more important without any information on the direction of the association, while logistic regression can discover hidden relationships in the data.
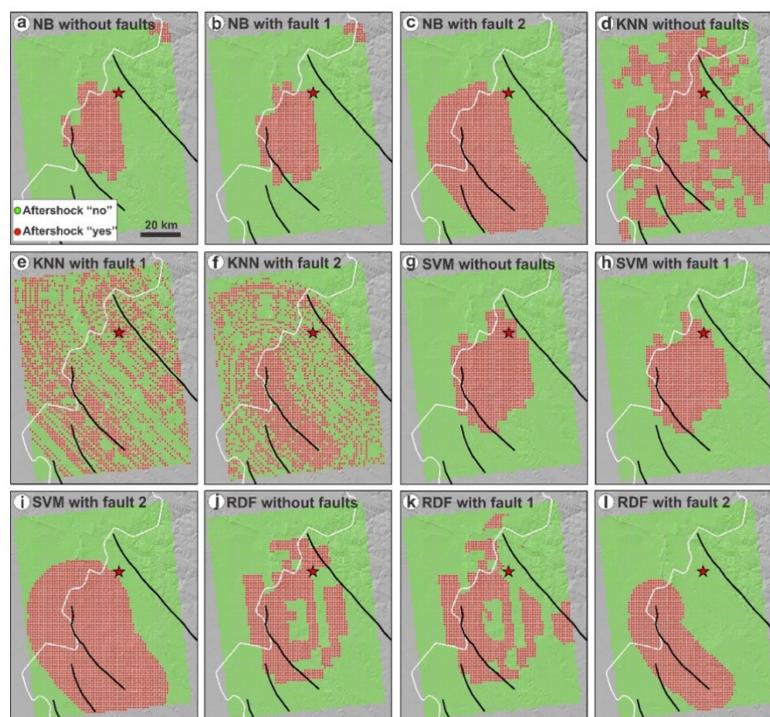

Figure 5. (**a**–**l**) Predicted aftershock patterns using different algorithms. The red star shows the location of the main shock. The small red dots are the aftershocks predicted by each classifier. The black lines delineate recognized faults.

Table 1. The overall accuracy (OA) of each different classifier.

	Logistic Regression (LR)	Naive Bayes (NB)	K-Nearest Neighbor (KNN)	Support Vector Machine (SVM)	Random Forest (RDF)
Without Fault	-	OA 78%	OA 58%	OA 74%	OA 73%
With Fault 1	OA 74%	OA 78%	OA 58%	OA 75%	OA 76%
With Fault 2	OA 74%	OA 65%	OA 61%	OA 57%	OA 75%
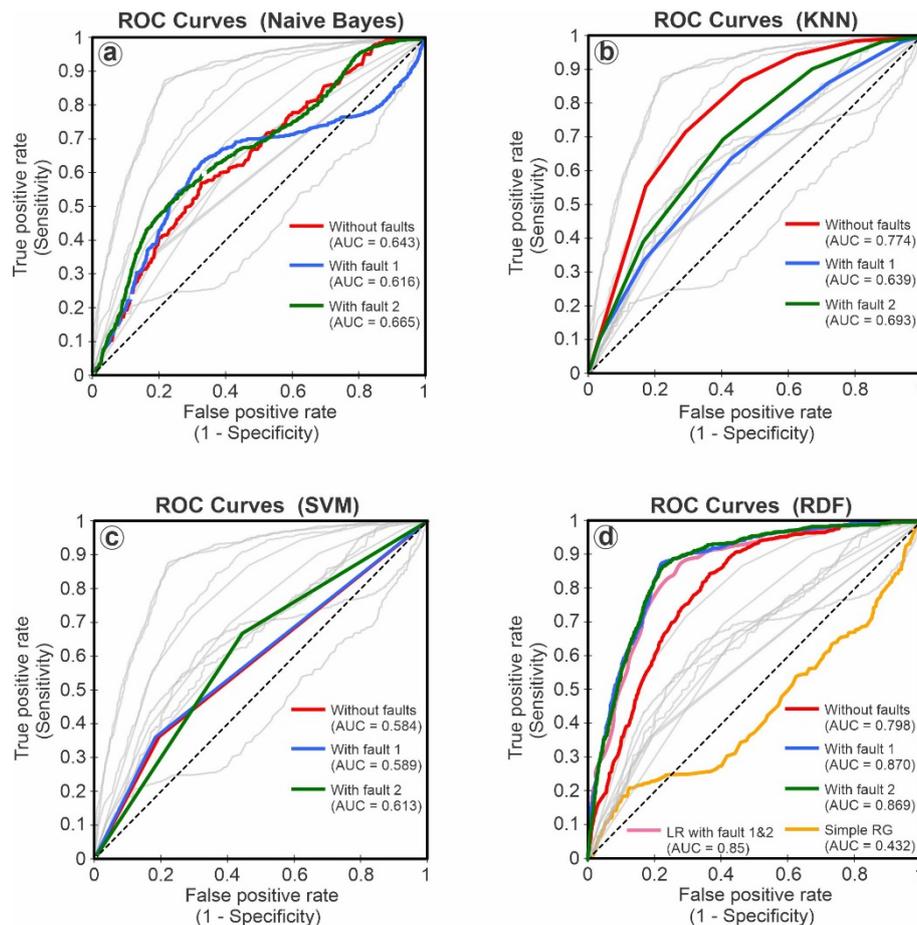
Figure 6. (**a**–**d**) ROC curves of the different classifiers. For the red curves, the Euclidean distances of the faults from the aftershocks were ignored. For the blue curves, the Euclidean distance of fault 1 was considered. For the green line, the Euclidean distance of fault 2 was considered. The gray curves in each figure show the ROC curves of the other classifiers. The pink and yellow curves in (**d**) indicate the ROC curve of the logistic regression and simple regression methods, respectively.

5. Conclusions

Here, we presented the ML patterns of the predicted spatial distributions of aftershocks. Since we evaluated the Coulomb failure map through the source model, the quality of the inverted source model was important for extracting reliable values for the training step. This shallow event was a suitable example of an InSAR-derived Coulomb failure map. However, source modeling of too shallow earthquakes can make the inversion procedure difficult. Thus, the source modeling should be addressed correctly before the ML analysis. We predicted the aftershock patterns using different ML and regression methods assuming available knowledge of the source and neighboring faults, the uncertainties in previous aftershocks (e.g., the azimuthal gap and RMS of travel time residuals), the slip

distribution, and Coulomb failure maps. Here, the defined buffer indicated that the spatial accuracy of the prediction was approximately 1.5 km. A denser seismic network and a larger number of records may help us to consider smaller spatial grids in future research. Nevertheless, this study showed that the aftershock pattern prediction was possible, even with a smaller database compared to a previous study [10].

This study does not prove that spatial prediction of independent datasets is possible because we did not validate the results with other independent datasets since the InSAR-based Coulomb stress map was not available for other seismic events. However, a receiver operating characteristic analysis for the same dataset showed that the ML methods outperformed routine Coulomb maps in the spatial prediction of aftershock patterns, especially when detailed information on neighboring active faults was available.

The AUC results from the logistic regression method were significantly greater than NB, KNN, and SVM ML methods, a finding that likely occurred because hidden information could also be better discovered in logistic regression analysis. In the next studies, the contribution of other reliable Coulomb maps of other regions deduced from InSAR or seismic approaches can also help us to develop a more comprehensive database that considers the predictions of aftershocks following other independent events if the seismicity in the target region is high. Since the Coulomb failure map is a spatially based first-order indicator of seismic events, the RDF and logistic regression methods yielded promising results in terms of spatial aftershock prediction, but the prediction of time and depth of the aftershocks remains a question.

Author Contributions: Conceptualization, Sadra Karimzadeh and Masashi Matsuoka; methodology, Sadra Karimzadeh, Jianming Kuang, and Linlin Ge; software, Sadra Karimzadeh and Jianming Kuang; validation, Sadra Karimzadeh; formal analysis, Jianming Kuang; investigation, Sadra Karimzadeh; resources, Sadra Karimzadeh, Masashi Matsuoka, Jianming Kuang, and Linlin Ge; data curation, Sadra Karimzadeh and Jianming Kuang; writing—original draft preparation, Sadra Karimzadeh; visualization, Sadra Karimzadeh; supervision, Sadra Karimzadeh and Masashi Matsuoka; project administration, Sadra Karimzadeh and Masashi Matsuoka; funding acquisition, Sadra Karimzadeh and Masashi Matsuoka.

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

References

1. Jia, Y.; Ma, J. What can machine learning do for seismic data processing? An interpolation application. *Geophysics* **2017**, *82*, V163–V177. [CrossRef]
2. Asim, K.M.; Idris, A.; Iqbal, T.; Martínez-Álvarez, F. Earthquake prediction model using support vector regressor and hybrid neural networks. *PLoS ONE* **2018**, *13*, e0199004. [CrossRef] [PubMed]
3. Kong, Q.; Trugman, D.T.; Ross, Z.E.; Bianco, M.J.; Meade, B.J.; Gerstoft, P. Machine Learning in Seismology: Turning Data into Insights. *Seismol. Res. Lett.* **2018**, *90*, 3–14. [CrossRef]
4. Kortström, J.; Uski, M.; Tiira, T. Automatic classification of seismic events within a regional seismograph network. *Comput. Geosci.* **2016**, *87*, 22–30. [CrossRef]
5. Che, I.-Y.; Jun, M.-S.; Jeon, J.-S. A compound linear discriminant method for small-magnitude seismic events and its application to the North Korea seismic event of October 9, 2006. *Earth Planets Space* **2007**, *59*, e41–e44. [CrossRef]
6. Lyubushin, A.; Kaláb, Z.; Lednická, M.; Mohamed, H. Discrimination of earthquakes and explosions using multi-fractal singularity spectrums properties. *J. Seismol.* **2013**, *17*, 975–983. [CrossRef]

7. Rouet-Leduc, B.; Hulbert, C.; Lubbers, N.; Barros, K.; Humphreys, C.J.; Johnson, P.A. Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.* **2017**, *44*, 9276–9282. [CrossRef]

8. Panakkat, A.; Adeli, H. Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators. *Comput. Aided Civ. Infrastruct. Eng.* **2009**, *24*, 280–292. [CrossRef]

9. Zamani, A.; Sorbi, M.R.; Safavi, A.A. Application of neural network and ANFIS model for earthquake occurrence in Iran. *Earth Sci. Inform.* **2013**, *6*, 71–85. [CrossRef]

10. DeVries, P.M.R.; Viégas, F.; Wattenberg, M.; Meade, B.J. Deep learning of aftershock patterns following large earthquakes. *Nature* **2018**, *560*, 632–634. [CrossRef]

11. Mignan, A.; Broccardo, M. One neuron versus deep learning in aftershock prediction. *Nature* **2019**, *574*, E1–E3. [CrossRef] [PubMed]

12. Karimzadeh, S.; Matsuoka, M.; Miyajima, M.; Adriano, B.; Fallahi, A.; Karashi, J. Sequential SAR coherence method for the monitoring of buildings in Sarpole-Zahab, Iran. *Remote Sens.* **2018**, *10*, 1255. [CrossRef]

13. Miyajima, M.; Fallahi, A.; Ikemoto, T.; Samaei, M.; Karimzadeh, S.; Setiawan, H.; Talebi, F.; Karashi, J. Site investigation of the Sarpole-Zahab earthquake, Mw 7.3 in SW Iran of November 12, 2017. *JSCE J. Disaster FactSheets* **2018**, FS2018-E-0002. Available online: http://committees.jsce.or.jp/disaster/system/files/FS2018-E0002_0.pdf (accessed on 10 October 2019).

14. Global Centroid Moment Tensor (GCMT). Global CMT Catalog Search. Available online: https://www.globalcmt.org/CMTsearch.html (accessed on 11 October 2018).

15. United States Geological Survey. Mw 7.3—Iran/Iraq border. Available online: https://www.usgs.gov/news/magnitude-73-earthquake-iraniraq-border (accessed on 11 October 2018).

16. Kuang, J.; Ge, L.; Metternicht, G.I.; Ng, A.H.-M.; Wang, H.; Zare, M.; Kamranzad, F. Coseismic deformation and source model of the 12 November 2017 MW 7.3 Kermanshah Earthquake (Iran–Iraq border) investigated through DInSAR measurements. *Int. J. Remote Sens.* **2018**, *40*, 532–554. [CrossRef]

17. Vajedian, S.; Motagh, M.; Mousavi, Z.; Motaghi, K.; Fielding, E.; Akbari, B.; Wetzel, H.-U.; Darabi, A. Coseismic deformation field of the Mw 7.3 12 November 2017 Sarpol-e Zahab (Iran) earthquake: A decoupling horizon in the Northern Zagros Mountains inferred from InSAR observations. *Remote Sens.* **2018**, *10*, 1589. [CrossRef]

18. Yang, Y.H.; Hu, J.C.; Yassaghi, A.; Tsai, M.C.; Zare, M.; Chen, Q.; Wang, Z.G.; Rajabi, A.M.; Kamranzad, F. Midcrustal thrusting and vertical deformation partitioning constraint by 2017 Mw 7.3 Sarpol Zahab earthquake in Zagros Mountain belt, Iran. *Seismol. Res. Lett.* **2018**, *89*, 2204–2213. [CrossRef]

19. Okada, Y. Surface deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* **1985**, *75*, 1135–1154.

20. Atzori, S.; Hunstad, I.; Chini, M.; Salvi, S.; Tolomei, C.; Bignami, C.; Stramondo, S.; Trasatti, E.; Antonioli, A.; Boschi, E. Finite fault inversion of DInSAR coseismic displacement of the 2009 L'Aquila earthquake (Central Italy). *Geophys. Res. Lett.* **2009**, *36*, L15305. [CrossRef]

21. Marquardt, D.W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [CrossRef]

22. Harris, R.A. Introduction to special section: Stress triggers, stress shadows, and implications for seismic hazard. *J. Geophys. Res. Solid Earth* **1998**, *103*, 24347–24358. [CrossRef]

23. Nostro, C.; Belardinelli, M.E.; Cocco, M. Static stress changes in extensional regimes: An application to Southern Apennines (Italy). *Bull. Seismol. Soc. Am.* **1997**, *87*, 234–248.

24. Parsons, T.; Stein, R.S.; Simpson, R.W.; Reasenberg, P.A. Stress sensitivity of fault seismicity: A comparison between limited-offset oblique and major strike-slip faults. *J. Geophys. Res. Solid Earth* **1999**, *104*, 20183–20202. [CrossRef]

25. Stein, R.S.; Barka, A.A.; Dieterich, J.H. Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophys. J. Int.* **1997**, *128*, 594–604. [CrossRef]

26. Toda, S.; Stein, R.S.; Reasenberg, P.A.; Dieterich, J.H.; Yoshida, A. Stress transferred by the 1995 Mw = 6.9 Kobe, Japan, shock: Effect on aftershocks and future earthquake probabilities. *J. Geophys. Res. Solid Earth* **1998**, *103*, 24543–24565. [CrossRef]

27. Hardebeck, J.L.; Nazareth, J.J.; Hauksson, E. The static stress change triggering model: Constraints from two Southern California aftershock sequences. *J. Geophys. Res.* **1998**, *1032*, 24427–24438. [CrossRef]

28. Mallman, E.P.; Zoback, M.D. Assessing elastic coulomb stress transfer models using seismicity rates in southern California and Southwestern Japan. *J. Geophys. Res.* **2007**, *112*, B03304. [CrossRef]

29. Hessami, K.; Jamali, F.; Tabassi, H. *Major Active Faults of Iran*; International Institute of Earthquake Engineering and Seismology: Tehran, Iran, 2003.

30. Vapnik, V. *Estimation of Dependences Based on Empirical Data*; Springer: New York, NY, USA, 1979.

31. Motevalli, A.; Pourghasemi, H.R.; Zabihi, M. Assessment of GIS-based machine learning algorithms for spatial modeling of landslide susceptibility: Case study in Iran. *Ref. Modul. Earth Syst. Environ. Sci.* **2018**, 258–276. [CrossRef]

32. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]

33. Wieland, M.; Liu, W.; Yamazaki, F. Learning change from synthetic aperture radar images: Performance evaluation of a support vector machine to detect earthquake and Tsunami-induced changes. *Remote Sens.* **2016**, *8*, 792. [CrossRef]

34. Platt, J. Sequential minimal optimization: A fast algorithm for training support vector machines. *CiteSeerX* **1998**. Available online: https://www.semanticscholar.org/paper/Sequential-Minimal-Optimization%3A-A-Fast-Algorithm-Platt/53fcc056f79e04daf11eb798a7238e93699665aa (accessed on 10 October 2019).

35. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; Chapman and Hall, Wadsworth: New York, NY, USA, 1984.