Performance of Logistic Regression and Support Vector Machines for Seismic Vulnerability Assessment and Mapping: A Case Study of the 12 September 2016 ML5.8 Gyeongju Earthquake, South Korea

Jihye Han 1, Soyoung Park 2, Seongheon Kim 1, Sanghun Son 1, Seonghyeok Lee 1 and Jinsoo Kim 3,*

1 Division of Earth Environmental System Science (Major of Spatial Information Engineering), Pukyong National University, 45 Yongso-ro, Busan 48513, Korea
2 BK21 Plus Project of the Graduate School of Earth Environmental Hazards System, Pukyong National University, 45 Yongso-ro, Busan 48513, Korea
3 Department of Spatial Information System, Pukyong National University, 45 Yongso-ro, Busan 48513, Korea
* Correspondence jinsookim@pknu.ac.kr; Tel.: +82-51-629-6658

Received: 27 September 2019; Accepted: 5 December 2019; Published: 9 December 2019

Abstract: In this study, we performed seismic vulnerability assessment and mapping of the ML5.8 Gyeongju Earthquake in Gyeongju, South Korea, as a case study. We applied logistic regression (LR) and four kernel models based on the support vector machine (SVM) learning method to derive suitable models for assessing seismic vulnerabilities; the results of each model were then mapped and evaluated. Dependent variables were quantified using buildings damaged in the 9.12 Gyeongju Earthquake, and independent variables were constructed and used as spatial databases by selecting 15 sub-indicators related to earthquakes. Success and prediction rates were calculated using receiver operating characteristic (ROC) curves. The success rates of the models (LR, SVM models based on linear, polynomial, radial basis function, and sigmoid kernels) were 0.652, 0.649, 0.842, 0.998, and 0.630, respectively, and the prediction rates were 0.714, 0.651, 0.804, 0.919, and 0.629, respectively. Among the five models, RBF-SVM showed the highest performance. Seismic vulnerability maps were created for each of the five models and were graded as safe, low, moderate, high, or very high. Finally, we examined the distribution of building classes among the 23 administrative districts of Gyeongju. The common vulnerable regions among all five maps were Jungbu-dong and Hwangnam-dong, and the common safe region among all five maps was Gangdong-myeon.

Keywords: seismic vulnerability; support vector machine; kernel function; logistic regression; machine learning; Gyeongju Earthquake; GIS

1. Introduction

Natural disasters such as earthquakes, landslides, or tsunamis damage buildings and cause loss of human life as well as environmental and economic losses due to unexpected changes in the environment [1]. Earthquakes are considered the most devastating natural disaster in most countries, causing a serious threat to human life and safety [2,3]. According to a UN report, about 10% of all natural disasters from 1998 to 2017 were related to earthquakes and volcanic eruptions [4], and about 23% of economic losses due to natural disasters was due to earthquakes, which also resulted in about 56% of all casualties. Thus, despite their low occurrence rate compared to other natural disasters, earthquakes cause considerable damage [5].
The Korean Peninsula is located inside the Eurasian plate, and has the characteristics of the Intraplate as it is close to the Japan and Ryukyu trenches where the Pacific plate and the Philippine plate enter the lower part of the Eurasian plate. It has accumulated local seismological stress due to plate tectonic movement. Compared to interplate earthquakes, intraplate earthquakes exhibit irregular spatiotemporal distribution, making it difficult to predict their occurrence, which is less frequent and accidental [6]. Since the beginning of the twentieth century, no major changes in seismic activity have occurred in the Korean Peninsula, which exhibits a low frequency of mid- to large-level earthquakes [7].

A magnitude 5.8 earthquake occurred in Gyeongju, South Korea, at 20:32:54 on September 12, 2016; this earthquake was preceded by a 5.1 foreshock, followed by many aftershocks, the largest of which (4.5) occurred at 11:33:58 on September 19, 2016 [8,9]. The Mw 5.8 Gyeongju Earthquake was recorded as the largest earthquake since South Korea began measuring earthquakes in 1978 [8,9]. Tremors from this earthquake were detected in most parts of the country; although no surface ruptures occurred, 23 people were injured and 5368 properties were destroyed [9,10].

In densely populated cities with large infrastructure, earthquake effects are easily amplified, and the corresponding ripple effect can continue for long periods, causing a considerable economic blow to the country [11]. To alleviate such losses in various areas, the management of natural disasters is essential [12]. To promote the sustainability of disaster management, the overall degree of earthquake damage can be reduced by identifying regions at high risk of earthquake occurrence and conducting disaster response and preparation activities in these regions [13,14].

In the past several years, many studies of natural disasters have implemented various methods to evaluate vulnerability and susceptibility, resulting in the development of risk-assessment maps. These methods have included probabilistic and statistical methodologies [15–19] and machine learning approaches [20–27]. Recent research has actively compared the suitability of machine-learning methodologies [28–32]. Compared to active research on landslides or floods, fewer studies have been conducted on seismic vulnerability.

Seismic vulnerability is best addressed by multi-criteria decision analysis and sustainable development [33]. To evaluate seismic vulnerability, many studies have applied the analytical hierarchy process (AHP) and multi-criteria decision analysis (MCDA) in conjunction with a geographic information system (GIS). If multiple targets require evaluation, they are stratified and their importance is quantified to determine the relative priorities of their criteria by using factor weighting [3,14,34–37].

To date, few studies of seismic vulnerability have applied machine-learning approaches, although some have used this technique for data mining. Machine learning analyzes and predicts data based on the automatic learning of statistical rules and patterns from large datasets [38], and its applicability has been demonstrated in various fields [39]. Şengezer et al. (2008) [40] evaluated parameters affecting earthquake damage using decision trees (DTs). Borfecchia et al. (2010) [41] analyzed urban seismic vulnerability parameters using DT and artificial neural network (ANN) data mining. To evaluate the seismic vulnerability of buildings, Tesfamariam and Liu (2010) [42] used support vector machine (SVM), random forest (RF), and other categorization methods, whereas Guettiche et al. (2017) [43] used association-rule learning (ARL). Riedel et al. (2015) [44] and Liu et al. (2019) [45] proposed building seismic vulnerability prediction methods based on building characteristics using the SVM and ARL approaches. Alizadeh et al. (2018) [2] studied the social vulnerability of Tabriz, Iran, using an ANN-based seismic-threat model, and Ahmed and Morita (2018) [46] analyzed the seismic vulnerability of residential buildings in Dhaka, Bangladesh, based on RF and DT approaches.

Studies based on machine learning have assessed the seismic vulnerability of target areas using seismic factors, with a focus on buildings. Although these studies have typically used single factors, such as geographic and building factors, few studies have considered the combined effects of multiple factors. Few studies have assessed seismic vulnerability by creating models incorporating SVM kernels, which are among the most widely used methodologies that are used to examine vulnerability to natural disasters.
Therefore, the objective of the present study was to evaluate and map seismic vulnerability among all buildings in Gyeongju, South Korea. Model performance was compared and analyzed based on four SVM kernels (linear, polynomial, radial basis function, and sigmoid), as well as logistic regression (LR), using 15 sub-indicators related to geotechnical, physical, structural, and capacity indicators. The accuracy of each model was verified using the receiver operating characteristic (ROC) curve, and a seismic vulnerability map was produced to evaluate the target regions according to administrative district.

2. Study Area and Data

2.1. Study Area

The target region of this study was Gyeongju, Gyeongsangbuk-do, South Korea (35°39′–36°04′N, 128°58′–129°31′E), which is bounded by the East Sea to the east; Cheongdo County and Yeongcheon City, Gyeongsangbuk-do, to the west; Ulju County, Ulsan, to the south; and Pohang City, Gyeongsangbuk-do, to the north. Comprising 23 districts, Gyeongju is 1324.82 km² in area and has a population of 256,141 (Figures 1a and 1b) [47].

Of the total area, 67.4% is forested and 14.8% is farmland, and 17.8% is for other purposes (e.g., commercial, residential, and industrial districts) [47]. Several faults are distributed throughout the region, including the Ulsan and Yangsan faults, and quaternary fault movement has been reported along the Dongrae, Moryang, Miryang, and Ilkwang faults [48]. These geographic properties imply that there is a high probability of earthquake occurrence in the future, which is expected to cause secondary natural disasters.

According to historical records, 75 earthquakes have occurred in Gyeongju, with 21 instrumental earthquakes having occurred prior to the Ml 5.8 Gyeongju Earthquake [10].

Gyeongju contains 396 cultural assets, including Yangdong Village, which is a UNESCO World Heritage Site; therefore, this city has very high preservation value [49]. Several national infrastructure facilities, including nuclear plants and nuclear-waste treatment facilities, are also located within the region; thus, it is necessary to minimize and prevent secondary damage propagation in the event of an earthquake by establishing preparatory measures.

2.2. Gyeongju Earthquake Inventory

The dependent variable used in this study is comprised of a dataset of the 3896 buildings damaged by the 9.12 Gyeongju Earthquake. These buildings were converted to 9847 cells at a spatial resolution of 10 m; 70% of the data (6893) were used as a training dataset to create the model, and 30% (2954) were used to test model accuracy. The data were randomly sampled; we included the same numbers of undamaged buildings (Figure 1c).

2.3. Spatial Database Preparation

To evaluate seismic vulnerability comprehensively, we considered all seismic-related factors. As a preceding study to this study, Han and Kim (2019) [50] evaluated seismic vulnerabilities using the analytic hierarchy process (AHP) technique. Factors were selected through prior study in relation to seismic vulnerabilities, and the survey was conducted to weight the factors. Thus, we selected the five main indicators to be geotechnical, physical, structural, and capacity factors. We then selected 15 sub-indicators related to these main indicators, and established a raster-type (10 m spatial resolution) spatial database related to these factors. Finally, all buildings were converted to cells and used as independent variables (Figure 2).
Figure 1. (a) Study areas in Gyeongju, South Korea. (b) Administrative districts within Gyeongju (1: Angang-eup, 2: Gangdong-myeon, 3: Seo-myeon, 4: Hyungok-myeon, 5: Cheonbuk-myeon, 6: Geoncheon-eup, 7: Seondo-dong, 8: Seonggun-dong, 9: Hwangseong-dong, 10: Yonggang-dong, 11: Junghu-dong, 12: Hwangoh-dong, 13: Dongcheon-dong, 14: Hwangnam-dong, 15: Wolseong-dong, 16: Bodeok-dong, 17: Bulguk-dong, 18: Yangbuk-myeon, 19: Gampo-eup, 20: Sannae-myeon, 21: Naenam-myeon, 22: Oedong-eup, and 23: Yangnam-myeon). (c) Training and validation datasets.
Figure 2. Fifteen sub-indicators related to geotechnical, physical, structural, and capacity seismic characteristics.
2.3.1. Geotechnical Indicators

Geotechnical indicators are the most influential factors affecting the vulnerability of a city to earthquake [36]. We considered three sub-indicators, among which slope and altitude cause secondary damage by increasing the probability of rocks and structures falling as well as ground failure [13]. Groundwater level is used as an important factor in the impact of a seismic response in the event of a large scale earthquake [51,52]; the groundwater level data were collected according to tubular well locations, and interpolated throughout Gyeongju.

2.3.2. Physical Indicators

The epicenter, or location where an earthquake occurs, is the most important indicator related to earthquake occurrence; the level of damage is different depending on ground condition or the structure of the fault plane, on which the greatest damage often occurs at the epicenter. Therefore, we used distance data from earthquake epicenters for January 2015 to April 2018, including the 9.12 Gyeongju Earthquake. Peak ground acceleration (PGA) is the degree to which the ground shakes at the Earth's surface [10]; it is generally the most important indicator for evaluating seismic vulnerability because it is related to the amount of fault activity [35]. In this study, raw data measured at each National Weather Services observatory in South Korea were converted to acceleration data and interpolated throughout Gyeongju [8,9]. We also used distance data from each fault to evaluate how the degree of damage changes with the structure of the fault plane.

2.3.3. Structural Indicators

The Ml:5.8 Gyeongju Earthquake resulted in 5368 cases of property damage; since then, the importance of buildings with anti-seismic design has been recognized. Since the introduction of anti-seismic design in South Korea in 1988, it has been mandatory only for buildings that are three stories or higher [53]. As of November 2016, 29.9% of residential buildings and 23.7% of non-residential buildings in Seoul were designed to be anti-seismic [53]. Because there is no guarantee that future earthquakes will not exceed the magnitude of the 9.12 earthquake, most buildings in South Korea are considered highly vulnerable. To assess their vulnerability, we identified four structural indicators of seismic vulnerability: building age, number of floors, construction materials, and building density. Construction materials included masonry, wood, concrete, steel, and a mixture of concrete and steel.

2.3.4. Capacity Indicators

Since disaster accommodation facilities are irregularly distributed, not all people have equivalent access. Thus, it is difficult to predict the scale of damage that may be caused by a disaster that results in considerable economic losses. To determine the accessibility of such facilities, we identified the locations of social infrastructure facilities that offer aid in the event of an earthquake, and of hazardous facilities that have the potential to cause huge damage. The degree of accessibility following a disaster was analyzed by considering the physical distances to five indicators, including four types of social infrastructure facility (hospital, fire station, police station, and road network) and one hazardous facility (gas station).
3. Methodology

The detailed workflow for the production and evaluation of the seismic vulnerability map is shown in Figure 3.

![Figure 3. Methodology flowchart for this study.](image)

### 3.1. Logistic Regression

The logistic regression (LR) model, developed by McFadden (1973) [54], is a multivariate regression analysis model that describes the relationship between a bivariate dependent parameter and several independent parameters [55] through the estimation of an optimal model. The addition of a link function suitable for a general linear regression model allows the parameter type to be continuous, discrete, or mixed, thus obviating the requirement of a normal distribution [56,57]. Some studies have shown that the LR model is more accurate than other types of models constructed for the same purpose [31,58,59]. The LR model based on a general linear model can be derived from the following equation:

\[ y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_n x_n \]  

(1)

\[ P = \frac{e^y}{1 + e^y} \]  

(2)

where \( y \) is the linear logistic model, \( b_0 \) is the y-intercept, \( b_n \) is the logistic coefficient of each factor, \( n \) is the number of factors controlling a seismic event, \( x \) is the earthquake conditioning factor, and \( P \) is the probability of damage (ranging from 0 to 1) in the event of an earthquake [31].

### 3.2. Support Vector Machine

The support vector machine (SVM) method is a supervised machine-learning method for solving problems of complex categorization and regression based on statistical-learning theory and structural-risk minimization principles [60–62].

The SVM was developed to determine an optimal hyperplane that can distinguish between two classes and \( r \) using a training dataset. The hyperplane with the largest margin between the two classes is the optimal hyperplane; the closest point to it is called a support vector [62].
For data that allow for linear separation $x_i$, a group of training vectors should be considered ($i = 1, 2, \ldots, n$), which are categorized into two classes, such that $y_i = \pm 1$. This process is shown in the following set of equations [63].

$$\frac{1}{2} ||w||^2$$ (3)

$$y_i((w \cdot x_i) + b) \geq 1$$ (4)

where $w$ is a coefficient vector that defines the orientation of the hyperplane in the feature space and $b$ is the offset of the hyperplane from the origin [64,65].

A cost function using the Lagrangian multiplier is defined as follows:

$$L = \frac{1}{2} ||w||^2 - \sum_{i=1}^{n} \lambda_i (y_i((w \cdot x_i) + b) - 1)$$ (5)

where $\lambda_i$ is the Lagrangian multiplier.

Because it is difficult to classify data linearly into various categories for regression, it is permitted to transform a nonlinear space into a linear space for optimal separation of two classes [66]. The constraints can be revised by introducing slack variables $\xi_i$ ($\xi_i \geq 0$), and $\nu(0,1)$ is introduced to explain misclassification.

$$y_i((w \cdot x_i) + b) \geq 1 - \xi_i$$ (6)

$$L = \frac{1}{2} ||w||^2 - \sum_{i=1}^{n} \xi_i$$ (7)

Vapnik (1995) [60] explained the nonlinear decision-making border using the kernel function $K(x_i,x_j)$. In SVM, selection of the kernel function is very important; generally, four kernel types are used: linear (LN), polynomial (PL), radial basis function (RBF), and sigmoid (SIG), which are defined as follows:

Linear: $K(x_i,x_j) = x_i^T x_j$

PL: $K(x_i,x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$

RBF: $K(x_i,x_j) = e^{-\gamma(x_i-x_j)^2}, \gamma > 0$

SIG: $K(x_i,x_j) = \tanh(\gamma x_i^T x_j + r)$ (8)

where the $\gamma$ term controls the width of the Gaussian kernel and is present in all functions except the linear function, and $d$ is a degree term that applies only to the polynomial function, and $r$, a bias term in the polynomial and sigmoid functions, is entered manually to improve the accuracy of SVM. Cost (C), a common parameter applied to all functions, is the reciprocal of the normalization parameter $\lambda$. For each controlling factor, higher C values correspond to less influence.

### 4. Results

#### 4.1. Model Validation

LR model creation and accuracy verification were performed using using IBM SPSS Statistics ver. 25.0 (Foundation for IBM Corp, Armonk, NY, USA). Based on the training dataset, we calculated the coefficient between seismic vulnerability and seismic factors (Table 1), and used the test dataset to create the LR model. The resulting model had a success rate of 0.649 and a prediction rate of 0.655 (Figure 4a and 4b).

We then used R ver. 3.6.0 (Foundation for Statistical Computing, Vienna, Austria) to create models for each of the four SVM kernels (LN-SVM, PL-SVM, RBF-SVM, and SIG-SVM), and the models were verified using SPSS. We used a dataset identical to LR; based on the training dataset, C and $\gamma$ values were adjusted and default values were applied to the $d$ and $r$. The accuracy of the generated models was determined using the ROC method to calculate the area under the curve (AUC). AUC values closer to 1 indicate higher accuracy [64].

The success rates of the models based on the four kernel types using the training dataset (LN-SVM, PL-SVM, RBF-SVM, and SIG-SVM models) were 0.649, 0.842, 0.998, and 0.630, respectively.
(Figure 4a). The test dataset was then applied to the generated models to calculate their prediction accuracy. The prediction accuracy values for the LN-SVM, PL-SVM, RBF-SVM, and SIG-SVM models were 0.651, 0.804, 0.919, and 0.629, respectively (Figure 4b). The success and prediction rates of the five models are listed in Table 2.

Table 1. Coefficients of the logistic regression model.

| Sub-Indicators               | Logistic Coefficient | Significance Coefficient |
|------------------------------|----------------------|--------------------------|
| Slope                       | 0.02403              | 0.000                    |
| Elevation                   | 0.00122              | 0.004                    |
| Groundwater level           | -0.01183             | 0.000                    |
| Peak ground acceleration    | -5.47077             | 0.000                    |
| Distance to faults          | -0.00009             | 0.000                    |
| Distance to epicenters      | 0.00008              | 0.000                    |
| Age of buildings            | -0.00082             | 0.351                    |
| Number of floors            | -0.02691             | 0.013                    |
| Density of buildings        | -0.00032             | 0.000                    |
| Construction materials      |                      |                          |
| Materials1 (masonry)        | 0.48755              | 0.744                    |
| Materials2 (concrete)       | 0.25250              | 0.866                    |
| Materials3 (wood)           | 0.43746              | 0.770                    |
| Materials4 (steel)          | 0.52131              | 0.727                    |
| Materials5 (mix)            | -0.14710             | 0.922                    |
| Distance to fire stations   | -0.00006             | 0.001                    |
| Distance to hospitals       | -0.00023             | 0.000                    |
| Distance to gas stations    | 0.00023              | 0.000                    |
| Distance to roads           | 0.00002              | 0.774                    |
| Distance to police stations | 0.00014              | 0.000                    |
| Constant                    | 1.00179              |                          |

Table 2. Area under the curve (AUC) for the five models (unit: %).

|                  | LR     | LN-SVM | PL-SVM | RBF-SVM | SIG-SVM |
|------------------|--------|--------|--------|---------|---------|
| Success rate     | 64.9   | 64.9   | 84.2   | 99.8    | 63.0    |
| Prediction rate  | 65.5   | 65.1   | 80.4   | 91.9    | 62.9    |

LR: Logistic regression; LN-SVM: Linear-support vector machine; PL-SVM: Polynomial-support vector machine; RBF-SVM: Radial basis function-support vector machine; SIG-SVM: Sigmoid-support vector machine.
Figure 4. Analysis of the receiver operating characteristic (ROC) curves of five models. (a) Success rate. (b) Prediction rate.

4.2. Seismic Vulnerability Mapping and Assessment

Five seismic-vulnerability maps were produced based on the four SVM kernel models and the LR model. Maps obtained using the SVM models were based on predicted values, and the LR map was produced by applying the logistic coefficient calculated above to the following equation:

\[
\text{LR map} = 1.00179 + (0.02403 \times \text{Slope}) + (0.00122 \times \text{Elevation}) \\
+ (-0.01183 \times \text{Groundwater level}) + (-5.47077 \times \text{PGA}) \\
+ (-0.00009 \times \text{Distance to fault}) + (0.00008 \times \text{Distance to epicenter}) \\
+ (-0.00082 \times \text{Building age}) + (-0.02691 \times \text{Number of floors}) \\
+ (-0.00032 \times \text{Building density}) + (0.48755 \times \text{Material 1}) + (0.25250 \times \text{Material 2}) \\
+ (0.43746 \times \text{Material 3}) + (0.52131 \times \text{Material 4}) + (-0.14710 \times \text{Material 5})
\]
The seismic vulnerability maps were classified as safe, low, moderate, high, or very high. The values of each map were normalized from 0 to 1 and then divided into five equal intervals to apply grades 1 to 5 (Figure 5).

We then examined the percentages of buildings assigned to each grade in the vulnerability maps (Table 3). In the LR map, the buildings in the “safe” class among the total 40,621 buildings were 2505 (6.17%), those in the “low” class were 4324 (10.64%), those in the “moderate” class were 11,537 (28.40%), those in the “high” class were 5454 (13.43%), and those in the “very high” class were 16,801 (41.36%). In the LN-SVM map, the buildings in the “safe” class were 24 (0.06%), those in the “low” class were 4597 (11.32%), those in the “moderate” class were 12,063 (29.70%), those in the “high” class were 20,346 (50.09%), and those in the “very high” class were 3591 (8.84%), whereas in the PL-SVM map, the buildings in the “safe” class were 282 (0.69%), those in the “low” class were 15,636 (38.49%), those in the “moderate” class were 24,624 (60.62%), those in the “high” class were 37 (0.09%), and those in the “very high” class were 42 (0.10%). In the RBF-SVM map, the buildings in the “safe” class were 268 (0.69%), those in the “low” class were 18,571 (45.72%), those in the “moderate” class were 14,502 (35.70%), those in the “high” class were 7141 (17.58%), and those in the “very high” class were 139 (0.34%), whereas in the SIG-SVM map, those in the “safe” class were 97 (0.24%), those in the “low” class were 5635 (13.87%), those in the “moderate” class were 19,739 (48.59%), those in the “high” class were 14,867 (36.60%), and those in the “very high” class were 283 (0.70%).

**Table 3. Distribution of building classes determined by the SVM and LR maps.**

| Model  | Safe  | Low   | Moderate | High   | Very High | Sum  |
|--------|-------|-------|----------|--------|-----------|------|
| LR     | Number of buildings | 2505  | 4324     | 11,537 | 5454      | 16,801 | 40,621 |
|        | Ratio   | 6.17  | 10.64    | 28.40  | 13.43     | 41.36  | 100.00 |
| LN-SVM | Number of buildings | 24    | 4597     | 12,063 | 20,346    | 3591   | 40,621 |
|        | Ratio   | 0.06  | 11.32    | 29.70  | 50.09     | 8.84   | 100.00 |
| PL-SVM | Number of buildings | 282   | 15,636   | 24,624 | 37        | 42     | 40,621 |
|        | Ratio   | 0.69  | 38.49    | 60.62  | 0.09      | 0.10   | 100.00 |
| RBF-SVM| Number of buildings | 268   | 18,571   | 14,502 | 7141      | 139    | 40,621 |
|        | Ratio   | 0.66  | 45.72    | 35.70  | 17.58     | 0.34   | 100.00 |
| SIG-SVM| Number of buildings | 97    | 5635     | 19,739 | 14,867    | 283    | 40,621 |
|        | Ratio   | 0.24  | 13.87    | 48.59  | 36.60     | 0.70   | 100.00 |
Finally, we examined the distribution of seismic vulnerability among the 23 administrative districts of Gyeongju (Figure 6). Based on the LR map, the regions most vulnerable to earthquake (i.e., high and very high classes) were districts 2, 5, and 12; based on the LN-SVM map, they were districts 11, 12, and 13; based on the RBF-SVM map, they were districts 11, 14, and 15; and based on the SIG-SVM map, they were districts 11, 12, and 13. The safest regions (i.e., safe and low classes) based on the LR map were districts 1, 3, and 8; based on the LN-SVM map, they were districts 1, 2, and 23; based on the PL-SVM map, they were districts 19, 20, and 23; based on the RBF-SVM map, they were districts 1, 2, and 13; and based on the SIG-SVM map, they were districts 1, 2, and 23.

Figure 5. Seismic vulnerability maps created using the support vector machine (SVM) and logistic regression (LR) models.
Figure 6. Ratio of vulnerable buildings among Gyeongju districts determined by each model.
5. Discussion

We produced seismic vulnerability maps using five models based on SVM and LR techniques and explored their functional differences. The LN-SVM, SIG-SVM, and LR models were very similar in terms of success and prediction rates. The negligible difference in accuracy between the training and verification datasets may indicate underfitting problems for these two models, such as they are too simplistic to extract data diversity. Therefore, the LN-SVM, SIG-SVM, and LR models may be inappropriate for predicting seismic vulnerability. The success rates of the PL-SVM (84.2%) and RBF-SVM (99.8%) models were very high, and their prediction rates were also high, at 80.4% and 91.9%, respectively. These two models incorporate nonlinear SVM kernels, and would, therefore, be useful for creating complex decision-making borders, even with a small number of features, and are advantageous in that they operate smoothly for various datasets. The PL-SVM and RBF-SVM models showed high accuracy using the training and verification datasets, and are therefore considered reliable. Using the AUC of the ROC, the prediction accuracy of the functional models was determined. The RBF-SVM model was the most reliable, followed by PL-SVM; the SIG-SVM model had the lowest prediction efficiency, and is, therefore, unfit for evaluating seismic vulnerability.

The results of the present study are consistent with those of previous studies. Xu and Xu (2012) [59] compared the performance of SVM kernels in spatial prediction models for landslides caused by earthquakes, and found that the RBF kernel function (0.843) yielded the highest model accuracy, followed by PL (0.837), LN (0.801), and SIG (0.655). Xu et al. (2016) [67] generated models using ANN and SVM kernels, and found that the RBF (0.882) and PL (0.888) kernels had similarly high accuracy values, followed by ANN (0.864), LN (0.795), and SIG (0.502). Feizizadeh et al. (2017) [62] evaluated the accuracy of SVM kernels and reported that the RBF kernel function (0.893) was the most appropriate for evaluating landslide susceptibility, whereas the SIG function (0.828) had the lowest prediction accuracy. Hong et al. (2018) [68] performed flood susceptibility mapping using fuzzy weight-of-evidence (fuzzy-WofE) and data-mining methods (SVM, LR, and RF), and found that the SVM model based on the RBF kernel had the highest accuracy and the LR model was less accurate than the other machine-learning methods.

The performance of the SVM model with respect to each kernel showed that the RBF kernel-based SVM model had the highest prediction accuracy. The LR coefficient values confirmed the significance of the factors. When significance coefficient values exceed 0.05, factors have less influence on the seismic vulnerability of buildings. Herein, the significance coefficient values with respect to the age of buildings and distance to roads were 0.351 and 0.774, respectively, and the corresponding factors were considered inappropriate for analyzing seismic vulnerability. The construction material factors were also considered inappropriate because their corresponding p-values all exceeded 0.05. The buildings were between 1 and 562 years old. Of the 40,621 buildings considered in this study, 950 (2.34%) were more than 100 years old and 121 (12.74%) of these were actually damaged, which is a high proportion. However, these buildings were all one-story buildings, including several cultural assets and traditional Korean-style houses. Further, buildings likely have less influence on seismic vulnerability because they continue to undergo maintenance. Based on the construction material factors, 22,325 masonry- and wood-based buildings were classified as very vulnerable and these buildings were all one to four stories in height. The distance to roads ranged from 0 to 3.31 km. Of the 40,621 buildings considered in this study, 39,762 (97.89%) were less than 1 km from a road. Because road accessibility was very high for almost all buildings within the study area, distance is not a significant factor for evaluating seismic vulnerability.

Subsequently, the maps obtained based on five models were compared. The LN–SVM and SIG–SVM models produced similar risky and safe regions. Central Gyeongju was a risky region, whereas northern and southwestern Gyeongju were safe regions. The outskirts of southeastern and southwestern Gyeongju were also safe regions in the PL–SVM map. In the RBF–SVM map, central Gyeongju was considered a vulnerable region, whereas most of the northern region was a safe region. In the PL–SVM map, which classified risk into five levels, the majority of the buildings belonged to low and moderate risk classes. Because only 0.19% of the buildings belonged to the high and very high classes, which is negligible for evaluating risky regions, only four maps were evaluated.
Angang-eup District, which was considered to be a safe region in four maps excluding the PL-SVM map, is located to the north of Gyeongju. This area was far from the epicenter, and the majority of the buildings within this area are less than 100 years old. Further, it is a flat region with an average slope with respect to the ground surface of 0.97° and an average altitude of 28.63 m. However, central Gyeongju was evaluated as a risky region in the majority of the maps. Additionally, 11–15 districts belonging to this region had an average building density of 660.04, which is the greatest building density and is very high compared with the building density of Angang-eup (1) (171.05). In addition, the average distance to the epicenter in central Gyeongju (3.79 km) was shorter than the average distance in Angang-eup (1), which was 19.29 km. Further, central Gyeongju had an average distance to faults of 0.74 km compared with an average of 4.72 km for Angang-eup (1), indicating the presence of numerous faults near the central region.

The seismic vulnerability evaluation presented here can be augmented by adding or excluding other parameters, and this model can be applied to evaluate building vulnerability in other regions.

6. Conclusions

In this study, the seismic vulnerability of Gyeongju was analyzed by applying 15 seismic sub-indicators related to geotechnical, physical, structural, and capacity indicators including slope, elevation, groundwater level, PGA, building age, construction materials, building density, number of floors, and distances to epicenter, fault, hospital, fire station, police station, road, and gas station. This research is important because it considered various components in evaluating seismic vulnerability and it showed that seismic vulnerability maps can be developed based on various models. Evaluation of the significant factors using logistic regression revealed that building age, distance to roads, and construction materials had less influence on the evaluation of seismic vulnerability in the relevant region. These results can be used as an important reference for seismic vulnerability evaluations in other regions or for selecting additional factors to be considered in the future. Further, the risky and safe regions based on the five seismic vulnerability maps should be given priority in terms of management. The seismic vulnerability map described here is useful for intuitively identifying regions with high vulnerability. An evaluation of seismic vulnerability should help to manage the environment, properties, buildings, and facilities to prepare for future earthquakes. These results can be used as important basic data to establish earthquake-related policies, which are expected to reduce the economic damage and fatalities caused by earthquakes.

Author Contributions: conceptualization, J.H., S.P. and J.K.; methodology, J.H. and S.P.; software, J.H., S.P., S.S and S.L.; validation, J.H. and S.P.; formal analysis, J.H. and S.P.; investigation, J.H.; data curation, J.H. and S.K.; writing—original draft preparation, J.H.; writing—review and editing, J.H. and J.K.; visualization, J.H.; supervision, J.K.; project administration, J.K.

Funding: This research was funded by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20171510101960).

Acknowledgments: This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20171510101960).

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

References
1. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. Landslide Susceptibility Mapping using GIS-Based Multi-Criteria Decision Analysis, Support Vector Machines, and Logistic Regression. Landslides 2014, 11, 425–439.
2. Alizadeh, M.; Alizadeh, E.; Asadollahpour Ketonee, S.; Shahabi, H.; Beiranvand Pour, A.; Panahi, M.; Bin Ahmad, B.; Saro, L. Social Vulnerability Assessment using Artificial Neural Network (ANN) Model for Earthquake Hazard in Tabriz City, Iran. Sustainability 2018, 10, 3376.
3. Bahadori, H.; Hasheminezhad, A.; Karimi, A. Development of an Integrated Model for Seismic Vulnerability Assessment of Residential Buildings: Application to Mahabad City, Iran. J. Build. Eng. 2017, 12, 118–131.
4. Wallemacq, P. Centre for Research on the Epidemiology of Disasters, Economic Losses; Poverty & Disasters: Brussels, Belgium, 2018.
5. Duk, K.L. Application of Satellite Imagery to Research on Earthquake and Volcano. *Korean J. Remote Sens.* **2018**, *34*, 1469–1478.
6. Shin, D.H.; Hong, S.; Kim, H. Investigation on Effective Peak Ground Accelerations Based on the Gyeongju Earthquake Records. *J. Earthq. Eng. Soc. Korea* **2016**, *20*, 425–434.
7. Lee, K. Comments on Seismicity and Crustal Structure of the Korean Peninsula. *Geophys. Geophys. Explor.* **2010**, *13*, 256–267.
8. Kim, K.H.; Kang, T.S.; Rhie, J.; Kim, Y.; Park, Y.; Kang, S.Y.; Han, M.; Kim, J.; Park, J.; Kim, M. The 12 September 2016 Gyeongju Earthquakes: 2. Temporary Seismic Network for Monitoring Aftershocks. *Geosci. J. Seoul* **2016**, *20*, 753–757.
9. Kim, Y.; Rhie, J.; Kang, T.S.; Kim, K.H.; Kim, M.; Lee, S.J. The 12 September 2016 Gyeongju Earthquakes: 1. Observation and Remaining Questions. *Geosci. J. Seoul* **2016**, *20*, 747–752.
10. Ministry of Public Safety and Security (MPSS). Report on the 9.12 Earthquake and Countermeasures; Ministry of Public Safety and Security: Seoul, Korea, 2017.
11. National Emergency Management Agency (NEMA). *Construction of Active Fault Map and Seismic Risk Map*; National Emergency Management Agency: Seoul, Korea, 2012; p. 899.
12. Pourghasemi, H.R.; Gayen, A.; Panahi, M.; Rezaie, F.; Blaschke, T. Multi-Hazard Probability Assessment and Mapping in Iran. *Sci. Total Environ.* **2019**, *692*, 556–571.
13. Moustafa, S.S.; Al-Arifi, N.S.; Jafari, M.K.; Naeem, M.; Alawadi, E.A.; Metwaly, M.A. First Level Seismic Microzonation Map of Al-Madinah Province, Western Saudi Arabia using the Geographic Information System Approach. *Environ. Earth Sci.* **2016**, *75*, 251.
14. Walker, B.B.; Taylor-Noonan, C.; Tabbernor, A.; Bal, H.; Bradley, D.; Schuurman, N.; Clague, J.J. A Multi-Criteria Evaluation Model of Earthquake Vulnerability in Victoria, British Columbia. *Nat. Hazards* **2014**, *74*, 1209–1222.
15. Yalcin, A.; Reis, S.; Aydinoglu, A.C.; Yomralioglu, T. A GIS-Based Comparative Study of Frequency Ratio, Analytical Hierarchy Process, Bivariate Statistics and Logistics Regression Methods for Landslide Susceptibility Mapping in Trabzon, NE Turkey. *Catena* **2011**, *85*, 274–287.
16. Regmi, A.D.; Devkota, K.C.; Yoshida, K.; Pradhan, B.; Pourghasemi, H.R.; Kumamoto, T.; Akgun, A. Application of Frequency Ratio, Statistical Index, and Weights-of-Evidence Models and their Comparison in Landslide Susceptibility Mapping in Central Nepal Himalaya. *Arab. J. Geosci.* **2014**, *7*, 725–742.
17. Rahmati, O.; Pourghasemi, H.R.; Zeinivand, H. Flood Susceptibility Mapping using Frequency Ratio and Weights-of-Evidence Models in the Golistan Province, Iran. *Geocarto Int.* **2016**, *31*, 42–70.
18. Youssef, A.M.; Pradhan, B.; Sefry, S.A. Flash Flood Susceptibility Assessment in Jeddah City (Kingdom of Saudi Arabia) using Bivariate and Multivariate Statistical Models. *Environ. Earth Sci.* **2016**, *75*, 12.
19. Khosravi, K.; Nohani, E.; Maroufinia, E.; Pourghasemi, H.R. A GIS-Based Flood Susceptibility Assessment and its Mapping in Iran: A Comparison between Frequency Ratio and Weights-of-Evidence Bivariate Statistical Models with Multi-Criteria Decision-Making Technique. *Nat. Hazards* **2016**, *83*, 947–987.
20. Yeon, Y.; Han, J.; Ryu, K.H. Landslide Susceptibility Mapping in Injae, Korea, using a Decision Tree. *Eng. Geol.* **2010**, *116*, 274–283.
21. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Spatial Prediction of Flood Susceptible Areas using Rule Based Decision Tree (DT) and a Novel Ensemble Bivariate and Multivariate Statistical Models in GIS. *J. Hydrol.* **2013**, *504*, 69–79.
22. Conforti, M.; Pascale, S.; Robustelli, G.; Sdao, F. Evaluation of Prediction Capability of the Artificial Neural Networks for Mapping Landslide Susceptibility in the Turbolo River Catchment (Northern Calabria, Italy). *Catena* **2014**, *113*, 236–250.
23. Shrestha, S.; Kang, T.; Suwal, M. An Ensemble Model for Co-Seismic Landslide Susceptibility using GIS and Random Forest Method. *ISPRS Int. J. Geo Inf.* **2017**, *6*, 365.
24. Kim, J.; Lee, S.; Jung, H.; Lee, S. Landslide Susceptibility Mapping using Random Forest and Boosted Tree Models in Pyeong-Chang, Korea. *Geocarto Int.* **2018**, *33*, 1000–1015.
25. Kadavi, P.; Lee, C.; Lee, S. Application of Ensemble-Based Machine Learning Models to Landslide Susceptibility Mapping. *Remote Sens.* **2018**, *10*, 1252.
26. Taalab, K.; Cheng, T.; Zhang, Y. Mapping Landslide Susceptibility and Types using Random Forest. *Big Earth Data* **2018**, *2*, 159–178.
27. Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Bui, D.T. A Comparative Assessment of Decision Trees Algorithms for Flash Flood Susceptibility Modeling at Haraz Watershed, Northern Iran. Sci. Total Environ. 2018, 627, 744–755.

28. Yilmaz, I. Comparison of Landslide Susceptibility Mapping Methodologies for Koyulhisar, Turkey: Conditional Probability, Logistic Regression, Artificial Neural Networks, and Support Vector Machine. Environ. Earth Sci. 2010, 61, 821–836.

29. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating Machine Learning and Statistical Prediction Techniques for Landslide Susceptibility Modeling. Comput. Geosci. 2015, 81, 1–11.

30. Pham, B.T.; Pradhan, B.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. A Comparative Study of Different Machine Learning Methods for Landslide Susceptibility Assessment: A Case Study of Uttarakhand Area (India). Environ. Model. Softw. 2016, 84, 240–250.

31. Wang, L.; Guo, M.; Sawada, K.; Lin, J.; Zhang, J. A Comparative Study of Landslide Susceptibility Maps using Logistic Regression, Frequency Ratio, Decision Tree, Weights of Evidence and Artificial Neural Network. Geosci. J. 2016, 20, 117–136.

32. Chen, W.; Sun, Z.; Han, J. Landslide Susceptibility Modeling using Integrated Ensemble Weights of Evidence with Logistic Regression and Random Forest Models. Appl. Sci. 2019, 9, 171.

33. Amiri, A.; Delavar, M.; Zahrai, S.; Malek, M. Tehran Seismic Vulnerability Assessment using Dempster–Shafer Theory of Evidence. In Proceedings of the Map Asia Conference, Kuala Lumpur, Malaysia, 14–16 August 2007; pp. 14–16.

34. Armaj, I. Multi-Criteria Vulnerability Analysis to Earthquake Hazard of Bucharest, Romania. Nat. Hazards 2012, 63, 1129–1156.

35. Rezaie, F.; Panahi, M. GIS Modeling of Seismic Vulnerability of Residential Fabrics Considering Geotechnical, Structural, Social and Physical Distance Indicators in Tehran using Multi-Criteria Decision-Making Techniques. Nat. Hazards Earth Syst. Sci. 2015, 15, 461–474.

36. Panahi, M.; Rezaie, F.; Meshkani, S. Seismic Vulnerability Assessment of School Buildings in Tehran City Based on AHP and GIS. Nat. Hazards Earth Syst. Sci. 2014, 14, 969–979.

37. Aliabadi, S.F.; Sarsangi, A.; Modiri, E. The Social and Physical Vulnerability Assessment of Old Texture Against Earthquake (Case Study: Fahadan District in Yazd City). Arab. J. Geosci. 2015, 8, 10775–10787.

38. Kim, K.; Yoon, S. Assessment of Building Damage Risk by Natural Disasters in South Korea using Decision Tree Analysis. Sustainability 2018, 10, 1072.

39. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine Learning in Geosciences and Remote Sensing. Geosci. Front. 2016, 7, 3–10.

40. Şengezer, B.; Ansal, A.; Bilen, Ö. Evaluation of Parameters Affecting Earthquake Damage by Decision Tree Techniques. Nat. Hazards 2008, 47, 547–568.

41. Borfecchia, F.; De Cecco, L.; Pollino, M.; La Porta, L.; Lugari, A.; Martini, S.; Ristoratore, E.; Pascale, C. Active and Passive Remote Sensing for Supporting the Evaluation of the Urban Seismic Vulnerability. Ital. J. Remote Sens. 2010, 42, 129–141.

42. Tesfamariam, S.; Liu, Z. Earthquake Induced Damage Classification for Reinforced Concrete Buildings. Struct. Saf. 2010, 32, 154–164.

43. Guettiche, A.; Guéguen, P.; Mimoune, M. Seismic Vulnerability Assessment using Association Rule Learning: Application to the City of Constantine, Algeria. Nat. Hazards 2017, 86, 1223–1245.

44. Riedel, I.; Guéguen, P.; Dalla Mura, M.; Pathier, E.; Leduc, T.; Chanussot, J. Seismic Vulnerability Assessment of Urban Environments in Moderate-to-Low Seismic Hazard Regions using Association Rule Learning and Support Vector Machine Methods. Nat. Hazards 2015, 76, 1111–1141.

45. Liu, Y.; Li, Z.; Wei, B.; Li, X.; Fu, B. Seismic Vulnerability Assessment at Urban Scale using Data Mining and GIScience Technology: Application to Urumqi (China). Geomatics. Nat. Hazards Risk 2019, 10, 958–985.

46. Ahmed, M.; Morita, H. An Analysis of Housing Structures’ Earthquake Vulnerability in Two Parts of Dhaka City. Sustainability 2018, 10, 1106.

47. Gyeongju City Hall. Available online: http://www.gyeongju.go.kr (accessed on 12 July 2019).

48. Kim, Y.; Kim, T.; Kyung, J.B.; Cho, C.S.; Choi, J.; Choi, C.U. Preliminary Study on Rupture Mechanism of the 9.12 Gyeongju Earthquake. J. Geol. Soc. Korea 2017, 53, 407–422.

49. Wikipedia. Available online: https://ko.wikipedia.org/wiki%EA%B2%BD%EC%A3%BC%EC%8B%9C (accessed on 12 July 2019).
50. Han, J.; Kim, J. A GIS-Based Seismic Vulnerability Mapping and Assessment using AHP: A Case Study of Gyeongju, Korea. Korean J. Remote Sens. 2019, 35, 217–228.
51. Karimzadeh, S.; Miyajima, M.; Hassanzadeh, R.; Amiraslanzadeh, R.; Kamel, B. A GIS-Based Seismic Hazard, Building Vulnerability and Human Loss Assessment for the Earthquake Scenario in Tabriz. Soil Dyn. Earthq. Eng. 2014, 66, 263–280.
52. Thaker, T.P.; Savaliya, P.K.; Patel, M.K.; Patel, K.A. GIS Based Seismic Risk Analysis of Ahmedabad City, India. In Proceedings of the GeoShanghai International Conference, Shanghai, China, 27 May 2018; pp. 108–116.
53. Kang, T.S.; Kim, D.K. Convergence Research Review; Convergence Research Policy Center: Seoul, Korea, 2017; pp. 38–63.
54. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior; Academic Press: New York, NY, USA, 1973; pp 105–142.
55. Kleinbaum, D.G. Survival Analysis, a self-learning Text. J. Math. Methods Biosci. 1998, 40, 107–108.
56. Colkesen, I.; Sahin, E.K.; Kavzoglu, T. Susceptibility Mapping of Shallow Landslides using Kernel-Based Gaussian Process, Support Vector Machines and Logistic Regression. J. Afr. Earth Sci. 2016, 118, 53–64.
57. Lee, S. Application of Logistic Regression Model and its Validation for Landslide Susceptibility Mapping using GIS and Remote Sensing Data. Int. J. Remote Sens. 2005, 26, 1477–1491.
58. Bui, D.T.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Landslide Susceptibility Assessment in the Hoa Binh Province of Vietnam: A Comparison of the Levenberg–Marquardt and Bayesian Regularized Neural Networks. Geomorphology 2012, 171, 12–29.
59. Xu, C.; Xu, X.; Dai, F.; Saraf, A.K. Comparison of Different Models for Susceptibility Mapping of Earthquake Triggered Landslides Related with the 2008 Wenchuan Earthquake in China. Comput. Geosci. 2012, 46, 317–329.
60. Vapnik, V. The Nature of Statistical Learning Theory; Springer: Berlin/Heidelberg, Germany, 1995.
61. Vapnik, V.; Vapnik, V. Statistical Learning Theory; Wiley: Hoboken, NJ, USA, 1998; pp. 156–160.
62. Feizizadeh, B.; Roodposhti, M.S.; Blaschke, T.; Aryal, J. Comparing GIS-Based Support Vector Machine Kernel Functions for Landslide Susceptibility Mapping. Arab. J. Geosci. 2017, 10, 122.
63. Lee, S.; Hong, S.; Jung, H. A Support Vector Machine for Landslide Susceptibility Mapping in Gangwon Province, Korea. Sustainability 2017, 9, 48.
64. Pradhan, B. A Comparative Study on the Predictive Ability of the Decision Tree, Support Vector Machine and Neuro-Fuzzy Models in Landslide Susceptibility Mapping using GIS. Comput. Geosci. 2013, 51, 350–365.
65. Tehrany, M.S.; Pradhan, B.; Mansor, S.; Ahmad, N. Flood Susceptibility Assessment using GIS-Based Support Vector Machine Model with different Kernel Types. Catena 2015, 125, 91–101.
66. Cortes, C.; Vapnik, V. Support-Vector Networks. Mach. Learn. 1995, 20.
67. Xu, C.; Shen, L.; Wang, G. Soft Computing in Assessment of Earthquake-Triggered Landslide Susceptibility. Environ. Earth Sci. 2016, 75, 767.
68. Hong, H.; Tsangaratos, P.; Ili, I.; Liu, J.; Zhu, A.; Chen, W. Application of Fuzzy Weight of Evidence and Data Mining Techniques in Construction of Flood Susceptibility Map of Poyang County. China Sci. Total Environ. 2018, 625, 575–588.

© 2019 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/).