

Article

Seismic Vulnerability Assessment and Mapping of Gyeongju, South Korea Using Frequency Ratio, Decision Tree, and Random Forest

Jihye Han ¹ , Jinsoo Kim ^{2,*}, Soyoung Park ¹, Sanghun Son ³ and Minji Ryu ³

¹ Geomatics Research Institute, Pukyong National University, 45 Yongso-ro, Busan 48513, Korea; jhhan931025@gmail.com (J.H.); yac100@pknu.ac.kr (S.P.)

² Department of Spatial Information System, Pukyong National University, 45 Yongso-ro, Busan 48513, Korea

³ Division of Earth Environmental System Science (Major of Spatial Information Engineering), Pukyong National University, 45 Yongso-ro, Busan 48513, Korea; ssh3496@naver.com (S.S.); rmj12@pukyong.ac.kr (M.R.)

* Correspondence: jinsookim@pknu.ac.kr; Tel.: +82-51-629-6658

Received: 21 August 2020; Accepted: 20 September 2020; Published: 21 September 2020



Abstract: The main purpose of this study was to compare the prediction accuracies of various seismic vulnerability assessment and mapping methods. We applied the frequency ratio (FR), decision tree (DT), and random forest (RF) methods to seismic data for Gyeongju, South Korea. A magnitude 5.8 earthquake occurred in Gyeongju on 12 September 2016. Buildings damaged during the earthquake were used as dependent variables, and 18 sub-indicators related to seismic vulnerability were used as independent variables. Seismic data were used to construct a model for each method, and the models' results and prediction accuracies were validated using receiver operating characteristic (ROC) curves. The success rates of the FR, DT, and RF models were 0.661, 0.899, and 1.000, and their prediction rates were 0.655, 0.851, and 0.949, respectively. The importance of each indicator was determined, and the peak ground acceleration (PGA) and distance to epicenter were found to have the greatest impact on seismic vulnerability in the DT and RF models. The constructed models were applied to all buildings in Gyeongju to derive prediction values, which were then normalized to between 0 and 1, and then divided into five classes at equal intervals to create seismic vulnerability maps. An analysis of the class distribution of building damage in each of the 23 administrative districts showed that district 15 (Wolseong) was the most vulnerable area and districts 2 (Gangdong), 18 (Yangbuk), and 23 (Yangnam) were the safest areas.

Keywords: seismic vulnerability assessment; frequency ratio; decision tree; random forest; machine learning; Gyeongju Earthquake; geographic information system (GIS)

1. Introduction

An ML 5.8 earthquake occurred 8.7 km south–southwest of Gyeongju, South Korea (35°46′36″ N, 129°11′24″ E) at 11:32:55 UTC (20:32:54 Korea Standard Time; GMT + 9 h) on 12 September 2016 [1,2]. The earthquake was accompanied by 601 aftershocks, including an ML 5.1 foreshock that occurred 8.2 km south–southwest of Gyeongju (35°46′12″ N, 129°11′24″ E) at 10:44:32 UTC (19:44:32 Korea Standard Time) and the largest aftershock (ML 4.5), which occurred at 11:33:58 UTC (20:33:58 Korea Standard Time) on 19 September. As of 31 March 2017 [3], the Gyeongju Earthquake was the largest earthquake among those recorded by the domestic seismic observation network; it consisted of a shock wave with concentrated energy, in which strong ground motion lasted for only 1–2 s, 15 km beneath the surface. Due to these characteristics, the initial reporting indicated that the earthquake did not significantly damage structures; however, it resulted in 5368 damaged properties, 111 victims,

and 23 injured people. The Gyeongju Earthquake represented a new disaster type that provoked a number of economic and social problems and revealed the limitations of established countermeasures. This disaster also made it impossible to rule out the possibility of similar earthquakes in the future, highlighting the importance of precautions to prevent greater losses.

The Korean Peninsula is located within the Eurasian plate and, therefore, has a lower earthquake occurrence frequency and longer recurrence period than countries located at the plate boundary. The geological structures of the Korean Peninsula include weak crust and many fault structures, which have led to increased earthquake occurrence frequency in recent years, partly because the peninsula is affected by earthquakes occurring in neighboring China and Japan [3]. The 2016 Gyeongju Earthquake occurred approximately five months after the occurrence of the ML 7.3 Kumamoto Earthquake on 16 April 2016, which followed the ML 9.0 Great East Japan Earthquake on 11 March 2011. Earthquake occurrence frequency and size are increasing globally; according to a UN report, disasters related to earthquakes and volcanoes accounted for approximately 10% of the natural disasters that occurred from 1998 to 2017 [4]. Although the proportion of earthquakes is low compared to those of other natural disasters, economic damage caused by earthquakes represented approximately 23% of that of total natural disasters and 56% of total human casualties during the same period. Property damage caused by domestic earthquakes totaled approximately 9.5 million USD in 2016 and 70.6 million USD in 2017, representing significant national losses. Despite continuous damage from earthquakes, it remains impossible to predict earthquake occurrence accurately or to control natural disasters artificially. However, it is possible to minimize damage by predicting areas vulnerable to earthquakes and potential damage, establishing policies suitable for such areas, and performing sustainable preparation in advance.

Seismic vulnerability assessment involves the comprehensive evaluation of factors that affect risks associated with earthquakes within predefined areas. Urban areas are at higher risk of seismic disasters than outlying areas due to their higher building and infrastructure density and larger population. Therefore, in assessing seismic vulnerability, it is essential to select suitable influential factors and methods for the area of interest. Several methodologies have been applied for seismic vulnerability assessment and mapping during the past few decades.

Seismic vulnerability assessment studies commonly analyze case studies using a combination of multi-criteria decision-making (MCDM) and geographic information system (GIS) approaches [5–7]. Among these, the analytical hierarchy process (AHP) is one of the most widely known MCDM methodologies; it stratifies and quantifies the importance of each applied influential factor to determine its relative importance, and assesses vulnerability by applying weights to all factors [8–12]. However, this method can be subjective because the opinion of the researcher can affect the weight assignment process; therefore, it is somewhat unsuitable for objective assessment. To address this problem, recent studies have applied hybrid models that combine various methodologies [13–17]. Lee et al. (2019) [16] developed the GIS-based Seismic-Related Vulnerability Calculation Software (SEVUCAS) for seismic vulnerability assessment, which includes a stepwise weight assessment ratio analysis (SWARA), radial basis function (RBF), and teaching–learning-based optimization (TLBO) methods. SEVUCAS provided reliable results by assigning the weights of main indicators and sub-indicators using SWARA and interpolation methods based on RBF and TLBO to reduce the effects of weights with significant variation at the boundary of each class for each factor. Yariyan et al. (2020) [17] constructed a hybrid model by integrating different decision support systems to increase the accuracy of seismic vulnerability mapping. Using this model, seismic vulnerability maps were created based on multiple-criteria decision analysis–multi-criteria evaluation (MCDA–MCE) and MCDA–fuzzy models to construct training datasets, and training points were randomly selected. The MCDA–MCE and MCDA–fuzzy models were found to have 0.85 and 0.80 model accuracy, respectively. Based on two training datasets, MCE–logistic regression (LR) (0.90) and fuzzy–LR (0.85) hybrid models were constructed. The accuracy of the resulting seismic vulnerability maps was found to be directly related to that of the training datasets.

Many recent studies related to seismic vulnerability assessment and mapping have been conducted using machine learning techniques [12,18–21]. For example, Han et al. (2019) [20] used a logistic regression (LR) model and applied the support vector machine (SVM) methodology to four kernel models (linear, polynomial, radial basis function, and sigmoid) to derive a suitable model for seismic vulnerability assessment; this study was notable in that the results of several seismic vulnerability models were compared analytically; such analyses are rarely conducted in this field, despite the broad application of machine learning techniques in recent years.

Vulnerability assessments have been conducted for natural disasters other than earthquakes, including floods [22–26], landslides [27–34], gully erosion [35–39], and groundwater contamination [40–44]. Some studies have compared the performance of various methodologies, including probabilistic techniques such as frequency ratio (FR) models [22,27,43], statistical techniques such as LR-based models [22,27,28,32,34,38,40–43], and machine learning algorithms such as decision tree (DT) [24,26,28,29,31,34,38,39,42–44], random forest (RF) [23,26,29–33,35,36,40,42,43], rotation forest (RoF) [23,31,33,42,44], adaptive boosting (AdaBoost) [23,39,42,44], random subspace (RS) [33,39,40,44], bagging [33,39,44], SVM [24,25,28,32,34,36–38,42,43], artificial neural network (ANN) [28,32,34,37], and naïve Bayes (NB) models [26,34,36,38,40].

Tree-based machine learning methodologies have mainly been applied in seismic vulnerability studies for parameter evaluation [45–47]. For other natural disasters, these methodologies have also been used to determine the relative influence of seismic parameters on the model results.

The objective of this study was to assess the seismic vulnerability of all buildings in Gyeongju, South Korea, and to create maps using these data. We applied FR, a probabilistic technique, and DT and RF, which are tree-based machine learning techniques, to construct models using 18 sub-indicators related to geotechnical, physical, structural, social, and capacity indicators as independent variables and building damage location data collected after the 2016 Gyeongju Earthquake as dependent variables. Model performance was verified using relative operating characteristic (ROC) curves. The results were compared and analyzed to identify models suitable for seismic vulnerability assessment and mapping and to evaluate the importance of each factor for each methodology. Finally, dangerous and safe areas were identified in each of 23 administrative districts by creating maps based on the model with the highest accuracy for each methodology, and the results were assessed (Figure 1).

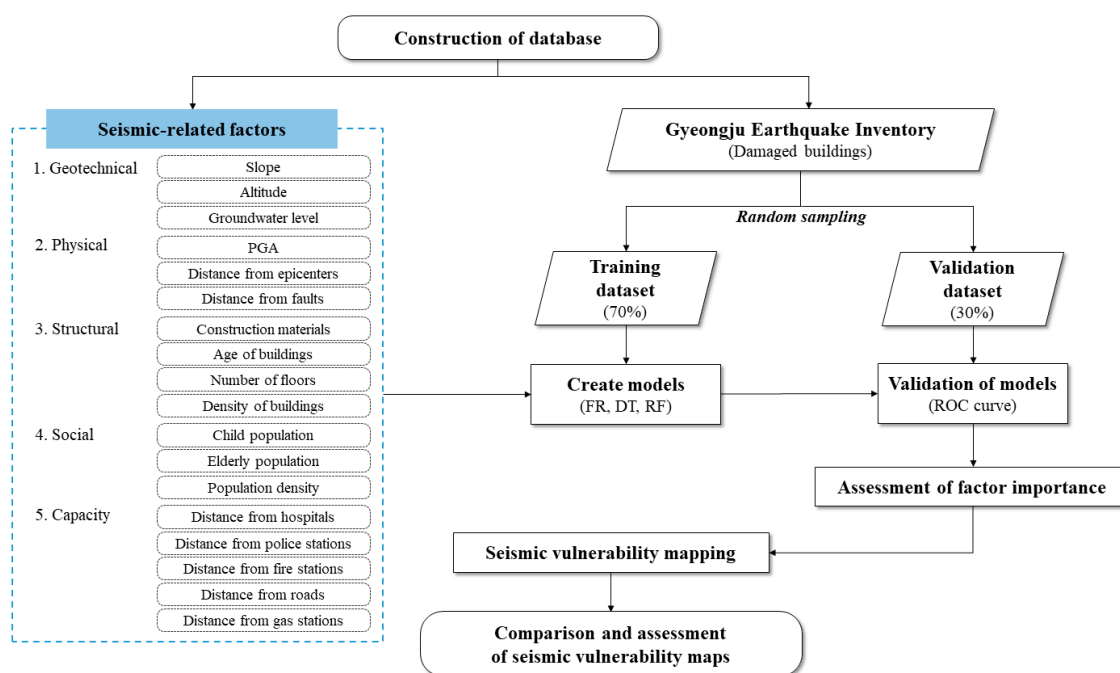


Figure 1. Flowchart of this study.

2. Study Area and Data

2.1. Study Area

The target area of this study was the city of Gyeongju, Gyeongsangbuk-do, South Korea ($35^{\circ}39'–36^{\circ}04' N$, $128^{\circ}58'–129^{\circ}31' E$). Gyeongju is in the southeastern part of the Korean Peninsula; it has a population of 254,853 and an area of 1324.82 km², and consists of 23 administrative districts (Figure 2). Within the total area, agriculture and forestry account for 42.36%, followed by green areas (31.04%) and other areas (26.6%) [48].

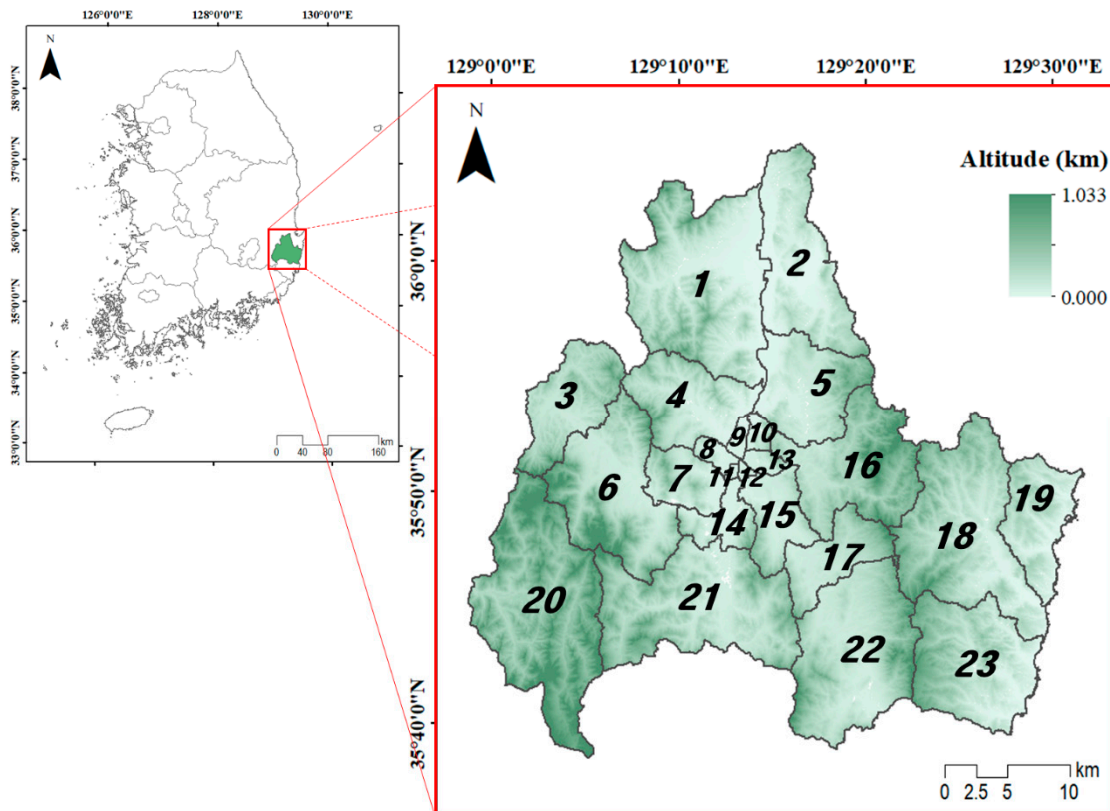


Figure 2. Study area in Gyeongju, South Korea: (1) Angang-eup, (2) Gangdong-myeon, (3) Seo-myeon, (4) Hyungok-myeon, (5) Cheonbuk-myeon, (6) Geoncheon-eup, (7) Seondo-dong, (8) Seonggun-dong, (9) Hwangseong-dong, (10) Yonggang-dong, (11) Jungbu-dong, (12) Hwangoh-dong, (13) Dongcheon-dong, (14) Hwangnam-dong, (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.

Several earthquakes of magnitude 3.0 or higher, which can be sensed by people, have occurred in Gyeongju. A magnitude 3.1 earthquake occurred 12 km northwest of the center of Gyeongju at around 08:13:23 AM on 24 December 1993, and a magnitude 4.2 earthquake occurred approximately 9 km east–southeast of the center of Gyeongju at around 03:50:22 AM on 26 June 1997. A magnitude 3.4 earthquake occurred 10 km to the northeast at around 06:12:23 PM on 2 June 1999, and a magnitude 3.2 earthquake occurred in the same area at 05:56:43 AM on 12 September 1999. A magnitude 3.0 earthquake occurred 9 km east–southeast of Gyeongju at 11:33:29 PM on 1 March 2003, and a magnitude 3.5 earthquake occurred approximately 18 km to the east–southeast at around 03:27:58 PM on 23 September 2014 [49].

Several faults in the study area, including Dongrae, Moryang, Miryang, Ulsan, and Yangsan, are distributed within the study area [50], and the Wolseong, Saoul, and Kori nuclear power plants are located along the nearby coastline to the southeast. Due to these regional and geographic characteristics,

the probability of earthquake occurrence in this region is considered to be relatively high, and secondary damage in the event of an earthquake with medium or higher magnitude constitutes an unusually high risk. In 2019, 957 earthquakes with magnitudes of less than 2.0 occurred in the Korean Peninsula; among these, 260 earthquakes (27.17%) occurred in the Gyeongsangbuk-do area (including Daegu) [49]. Among the 88 earthquakes of magnitude 2.0 or higher, 23 (26.17%) occurred in the same area. Since the 2016 Gyeongju Earthquake, large and small earthquakes have occurred continuously. Therefore, sustainable preparation and management planning for such events is required.

2.2. Data

We selected factors affecting seismic vulnerability based on the results of a previous study, taking into consideration applicability and practicality [51]. The main indicators were geotechnical, physical, structural, social, and capacity indicators; we selected a total of 18 sub-indicators corresponding to these categories. Geotechnical sub-indicators included slope, altitude, and groundwater level; physical sub-indicators included peak ground acceleration (PGA), epicenter distance, and fault distance; structural sub-indicators included building age, construction materials, building density, and number of floors; social sub-indicators indicators included elderly population (≥ 65 years), child population (< 15 years), and population density; and capacity sub-indicators included distances from hospitals, fire stations, police stations, roads, and gas stations. Sub-indicators were organized into a raster-based spatial database (10 m spatial resolution) and applied to all buildings in Gyeongju as independent variables (Figure 3).

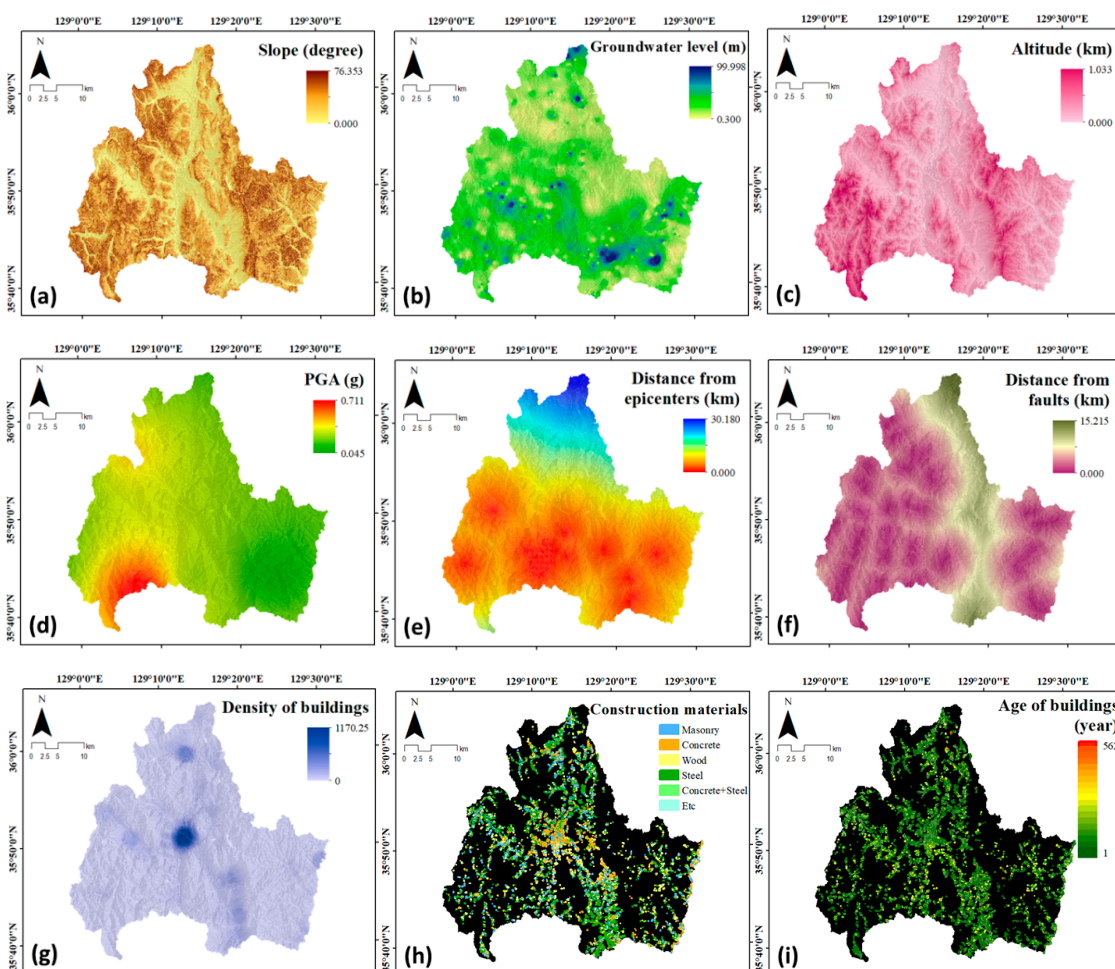


Figure 3. Cont.

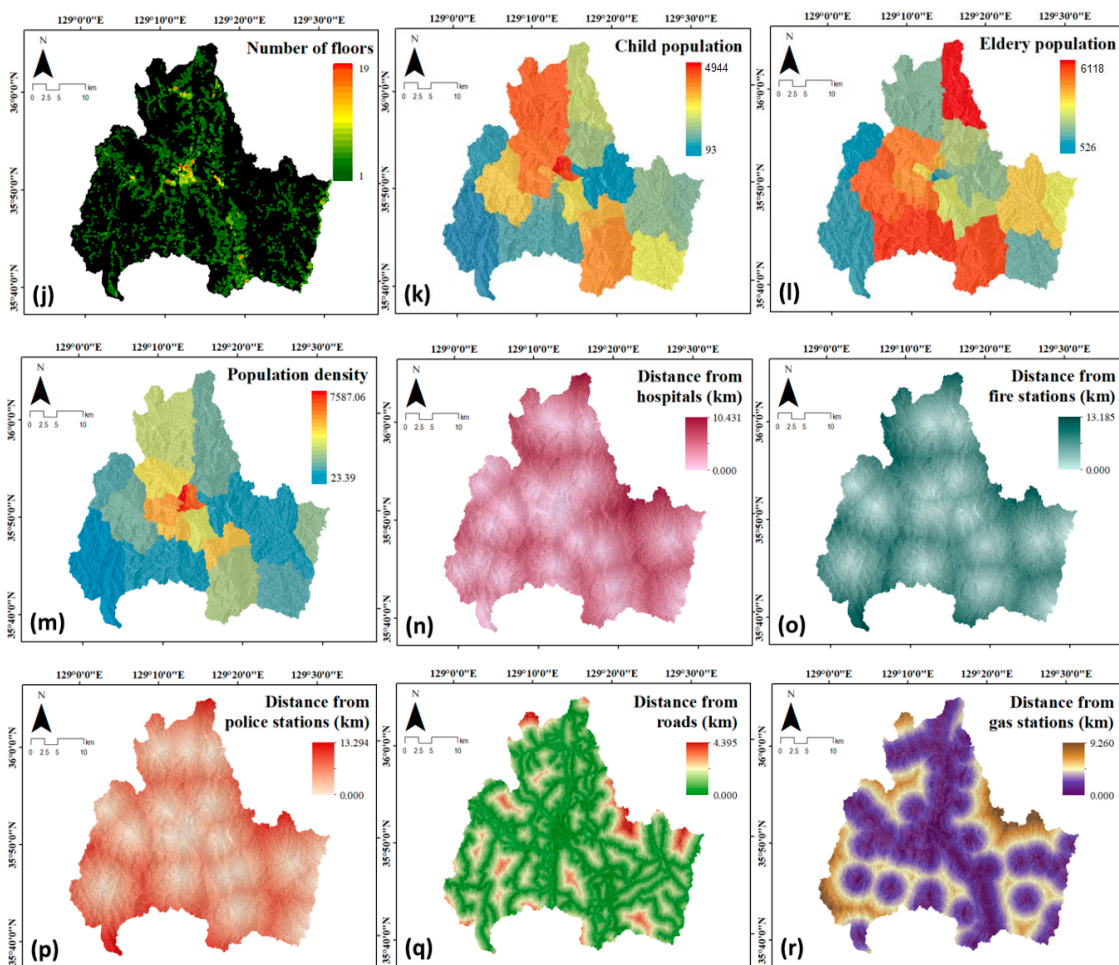


Figure 3. Sub-indicators related to seismic vulnerability: (a) slope, (b) groundwater level, (c) altitude, (d) peak ground acceleration (PGA), (e) distance from epicenters, (f) distance from faults, (g) density of buildings, (h) construction materials, (i) age of buildings, (j) number of floors, (k) child population, (l) elderly population, (m) population density, (n) distance from hospitals, (o) distance from fire stations, (p) distance from police stations, (q) distance from roads, and (r) distance from gas stations.

We used the 3896 buildings damaged during the 2016 Gyeongju Earthquake as dependent variables. The corresponding building polygons were converted into cells (10×10 m spatial resolution) for a total of 9847 cells. Among these cells, 70% (6893) were used as a training dataset to create the models and 30% (2954) were used as a test dataset. We extracted the same number of cells corresponding to undamaged buildings. All cells were randomly sampled, and the accuracy of each model was calculated based on the final training (13,786) and test datasets (5908).

3. Methodology

3.1. FR Model

The FR model is a probabilistic model used to determine the influence of each factor by analyzing the correlations between seismic vulnerability and earthquake-related factors. The FR model easily classifies the influence factors associated with the largest numbers of accidents during a disaster [52]. $FR > 1$ indicates strong correlation between seismic vulnerability and the factor class, whereas $FR < 1$ indicates weak correlation. FR is calculated as follows [53]:

$$FR = \frac{TGFC/WTG}{FC/WG} \quad (1)$$

TGFC: Training Grid of Factor Class

WTG: Whole Training Grid

FC: Factor Class Grid

WG: Whole Grid

In this study, *WTG* represents the number of cells corresponding to damaged buildings, *TGFC* is the number of cells corresponding to damaged buildings in the corresponding class, *WG* is the number of cells corresponding to all buildings, and *FC* is the number of cells corresponding to the buildings of the corresponding class. After *FR* values are calculated for each class of the 18 factors and applied to the grid format of each factor, they are superimposed to create the final seismic vulnerability maps.

3.2. DT Model

The DT model uses hierarchical structures to find structural patterns in data for the purpose of constructing decision-making rules to estimate the relationships between independent and dependent variables [54]. The DT model consists of three nodes: a root node (all data) located at the top, a set of internal nodes (splits), and a set of terminal nodes (leaves) located at the bottom. Pruning is performed from the top of the tree to its bottom until the terminal nodes are reached [55].

Four main algorithms are used to construct DTs: a classification and regression tree (CART), chi-square automatic interaction detector DT (CHAID), ID3, and C4.5 [56]. In this study, we constructed a regression tree model for seismic vulnerability assessment based on a CART algorithm developed by Breiman et al. (1984) [57]. CART is among the most widely known DT algorithms; it minimizes variance through binary recursive partitioning of the branches of a regression tree [58,59]. In this process, CART repeatedly creates two sub-nodes by partitioning a subset of the data using all predictors; its final goal is to create an optimal tree among several candidate trees [60].

In this study, we applied the DT model using the *rpart* package of the RStudio software (ver. 3.6.0), which creates an optimal model by adjusting the values of representative parameters, i.e., *minsplit*, *minbucket*, *maxdepth*, and *cp*. *Minsplit* is the minimum number of observations available at the node for splitting attempts, and *minbucket* is the minimum number of observations at all terminal nodes. As the *minbucket* value decreases, the tree becomes more detailed, thereby increasing the complexity of the model and increasing the prediction rate. *Maxdepth* is the maximum depth of the tree; if its value is 1, then a redundant column is not used as a node, whereas if it is 2 or greater, then redundant column nodes are allowed, increasing the complexity of the model. The *cp* value is a complexity parameter, and has values between zero and 1. As the *cp* value decreases, the size of the tree increases [31].

3.3. RF Model

RF [61] is a powerful ensemble algorithm that exhibits excellent performance; it has a wide variety of applications, including classification, regression, and unsupervised learning [60,62]. RF creates a binary tree by randomly selecting the training data of variables selected at each node based on a bootstrap sample, and constructs a DT for final prediction [63]. The tree inducer then selects the optimal data by randomly sampling an attribute subset instead of performing optimal partitioning; this process is an improved version of the bagging method, which forms a random DT at each iteration [64].

The regression algorithm of RF, which was used in this study, calculates estimates of the dependent variables using the average of the results. RF is suitable for analyzing hierarchical interactions and nonlinearity among large datasets because it does not require assumptions about the relationships between explanatory and response variables [65].

In this study, we applied the RF model using the *randomForest* package of the RStudio software (ver. 3.6.0). To create the model, we defined four parameters: the number of trees (*ntree*), the number

of variables to be used at each node (*mtry*), the maximum number of terminal nodes (*maxnodes*), and the depth setting of the tree (*nodesize*). Although increasing *ntree* does not guarantee an increase in model accuracy, several *ntree* values must be tested before finding a sufficiently high value to allow the error to converge [66]. If *maxnodes* is not given, the tree grows to its maximum; *nodesize* is a minimum number of nodes, and a small value creates a deep tree.

3.4. Assessment of Model Performance

Based on training and test datasets, the three models were verified for performance using statistical measures. These are classified into four categories, depending on how well they predicted the actual damaged building—true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*). The *TN* and *TP* are pixels that accurately classified as damaged and undamaged buildings, and *FP* and *FN* are pixels that are classified as opposed to actually damaged. These are used to calculate the following statistical metrics: Sensitivity (also referred to as recall) is the proportion of the damaged building pixels correctly classified, Specificity is the proportion of undamaged building pixels correctly classified, Precision is a positive predictive value that represents the proportion of actual damaged building pixels to those classified as damaged buildings by the model, Accuracy is the proportion of the correctly classified damaged and undamaged pixels, and F1-score means the harmonic mean of precision and sensitivity [23,31,42]. Statistical indices are calculated as follows.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (5)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (6)$$

In addition, the models created based on the three methodologies were verified using the receiver operating characteristics (ROC) curve method. This method evaluates overall model performance by calculating the area under the ROC curve (AUC) values. The AUC can be classified as follows: excellent (0.9–1), very good (0.8–0.9), good (0.7–0.8), average (0.6–0.7), and poor (0.5–0.6) [28]. The *y*-axis of the ROC curve graph represents sensitivity, or the true positive rate. The *x*-axis represents 1—specificity, or the false positive rate.

4. Results

4.1. Model Validation and Comparison

Based on the results of the statistical metrics, the performances of the models were compared (Table 1). The RF model showed a value of 1.000 with respect to all statistical indices, and generally showed excellent performance among the three models. The DT and FR models showed that the value of DT was higher than that of FR in most statistical indices. Its specificity was DT (0.842) and FR (0.415), and precision was DT (0.838) and FR (0.583). The accuracy was shown as DT (0.828) and FR (0.616), and F1-score as DT (0.826) and FR (0.584), whereas, for sensitivity, FR (0.816) was slightly higher than DT (0.814).

Table 1. Performance results of three models using training and test datasets.

	Training Dataset			Test Dataset		
	FR	DT	RF	FR	DT	RF
TP	5628	5614	6890	2573	2257	2545
TN	2862	5804	6890	939	2394	2609
FP	4031	1089	3	2015	560	345
FN	1265	1279	3	381	697	409
Sensitivity	0.816	0.814	1.000	0.871	0.764	0.862
Specificity	0.415	0.842	1.000	0.318	0.810	0.883
Precision	0.583	0.838	1.000	0.561	0.801	0.881
Accuracy	0.616	0.828	1.000	0.594	0.787	0.872
F1-score	0.584	0.826	1.000	0.616	0.782	0.881
AUC	0.661	0.899	1.000	0.655	0.851	0.949

The three models were verified for prediction ability using a test dataset. Most statistical indices showed that the highest value for RF, followed by DT and FR. The specificity was shown as RF (0.883), DT (0.810), and FR (0.318), and for FR, undamaged buildings were not the best classified. The precision was shown to be RF (0.881), DT (0.801), and FR (0.561), which best matched the positive predicted values of the model with the actual damaged buildings. Its accuracy was shown as RF (0.872), DT (0.787), and FR (0.594); the RF model was the best in classifying the damaged and undamaged buildings. The F1-score, which considered precision and sensitivity, was shown as RF (0.881), DT (0.782), and FR (0.616). For sensitivity, FR (0.871) was the highest, which was the best at classifying actual damaged buildings, followed by RF (0.862) and DT (0.764). These results confirmed that the RF model seems to be most suitable for predicting the damaged buildings.

The performance accuracy of the model was verified by calculating the success and prediction rates through ROC curves (Figures 4 and 5). The success rate is a measure of the training degree of the model based on the training data, and the prediction rate is a measure of how well the model predicts damage to the building based on the test data. We verified the accuracy of all models using the IBM SPSS software (ver. 25). The FR model exhibited a success rate of 0.661 and a prediction rate of 0.655. The DT model constructed the optimal model by adjusting the minsplit, minbucket, maxdepth, and cp values based on the training datasets. In this study, the optimal model was created at minsplit, minbucket, maxdepth, and cp values of 20, 7, 30, and 0.001, respectively. The success and prediction rates were 0.899 and 0.851, respectively. RF also created the optimal model by adjusting the ntree and mtry values based on training datasets. The highest accuracy was observed at an ntree of 8000 and mtry of 6. The RF model showed the highest performance among the three methodologies, with a success rate of 1.000 and prediction rate of 0.949. The validation based on the statistical indices and ROC curves confirmed that the RF model is the most suitable model for the training and test datasets.

4.2. Relative Importance of Factors

After deriving the optimal model for each methodology, we determined the relative importance of the 18 sub-indicators. First, for the FR model, each factor was divided into six classes using the natural breaks method to identify the class with the largest impact on seismic vulnerability. FR values were calculated for the sub-indicators in each class based on the number of pixels corresponding to undamaged and damaged buildings, respectively (Table 2). The classes with the greatest impact on seismic vulnerability were: altitude of 86.061–138.262 m (FR = 1.23), slope of 1.716–4.291° (1.11), groundwater level of 21.047–37.061 m (1.47), fault distance of 6.124–7.946 km (1.26), epicenter distance of 0.028–3.183 km (1.24), PGA of 0.244–0.288 g (1.17), building age of 33–59 years (1.38), 5–7 building

floors (1.53), construction materials consisting of concrete mixed with steel (1.44), building density of 949.480–1169.540 (1.53), child population of 1020–1279 (1.21), elderly population of 526 (1.94), population density of 201.358–586.957 (1.76), 0.000–1.205 km distance from police stations (1.18), 0.000–1.431 km distance from fire stations (1.16), 4.216–5.646 km distance from hospitals (1.16), 0.000–0.116 km distance from roads (1.05), and 0.000–0.680 distance from gas stations (1.16). Buildings corresponding to these classes are predicted to experience the highest degree of damage due to earthquakes.

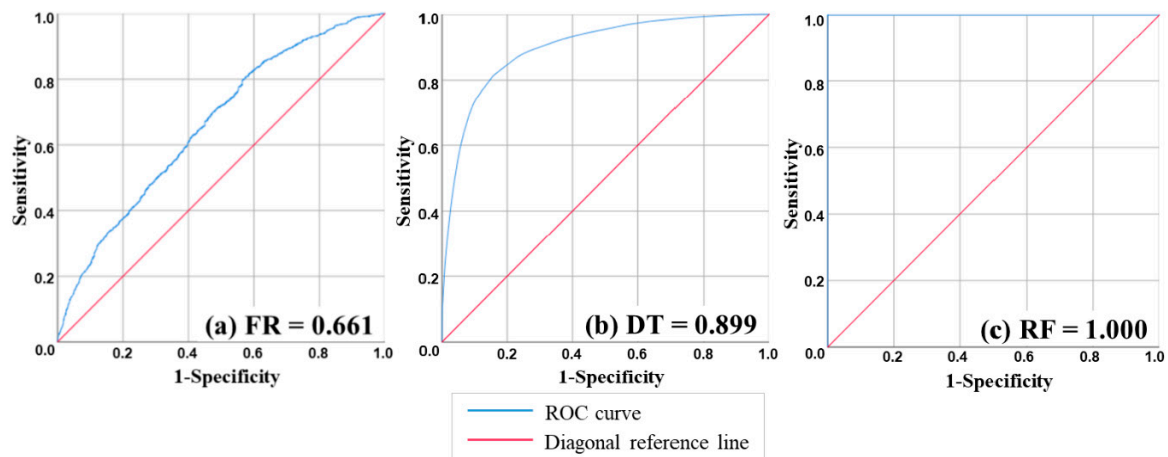


Figure 4. Success rates using the training dataset; (a) frequency ratio (FR), (b) decision tree (DT), and (c) random forest (RF) models.

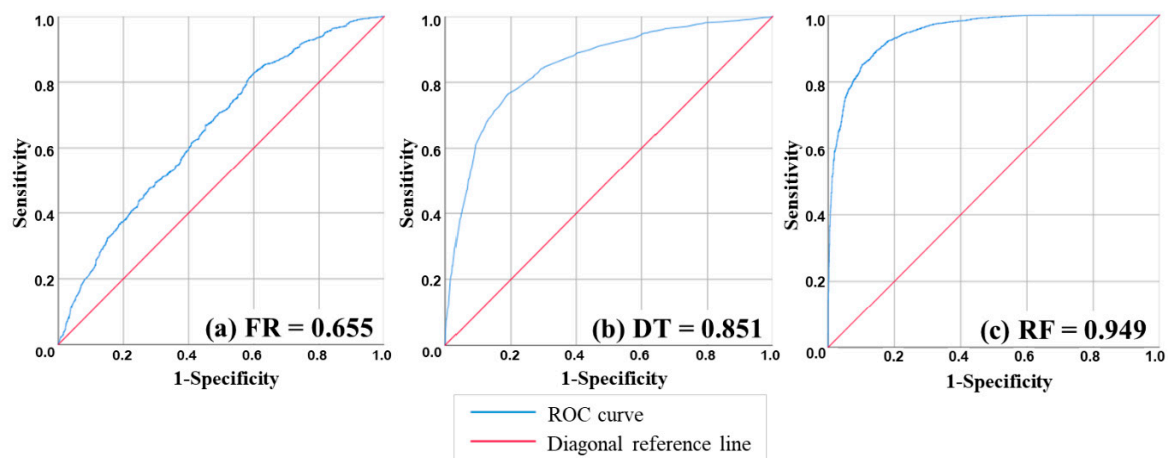


Figure 5. Prediction rates using test dataset: (a) FR, (b) DT, and (c) RF models.

Table 2. Frequency ratio of each factor.

	Class	No. of Pixels in Building	Building (%)	No. of Pixels in Damaged Building	Damaged Building (%)	Frequency Ratio
Altitude (m)	1.545–46.289	40,284	43.96	4221	42.87	0.98
	46.289–86.061	24,840	27.11	2399	24.36	0.90
	86.061–138.262	17,421	19.01	2308	23.44	1.23
	138.262–220.292	6468	7.06	746	7.58	1.07
	220.292–366.952	1787	1.95	164	1.67	0.85
	366.952–635.414	842	0.92	9	0.09	0.10

Table 2. Cont.

	Class	No. of Pixels in Building	Building (%)	No. of Pixels in Damaged Building	Damaged Building (%)	Frequency Ratio
Slope (degree)	0–1.716	47,128	51.43	5278	53.60	1.04
	1.716–4.291	23,371	25.50	2778	28.21	1.11
	4.291–7.725	13,189	14.39	1264	12.4	0.89
	7.725–12.016	5533	6.04	380	3.86	0.64
	12.016–18.597	1996	2.18	113	1.15	0.53
	18.597–72.959	425	0.46	34	0.35	0.74
Groundwater level (m)	0.346–7.377	30,754	33.56	2399	24.36	0.73
	7.377–12.845	39,133	42.70	4469	45.38	1.06
	12.845–21.047	15,209	16.60	2080	21.12	1.27
	21.047–37.061	5153	5.62	813	8.26	1.47
	37.061–83.346	1075	1.17	73	0.74	0.63
	83.346–99.947	318	0.35	13	0.13	0.38
Distance from faults (km)	0–1.973	25,199	27.50	2825	28.69	1.04
	1.973–3.947	29,228	31.89	3021	30.68	0.96
	3.947–6.124	15,147	16.53	1376	13.97	0.85
	6.124–7.946	10,904	11.90	1479	15.01	1.26
	7.946–9.768	6947	7.58	758	7.70	1.02
	9.768–12.906	4217	4.60	389	3.95	0.86
Distance from epicenters (km)	0.028–3.183	17,765	19.39	2368	24.05	1.24
	3.183–6.112	35,244	38.46	4529	45.99	1.20
	6.112–10.731	21,506	23.47	1931	19.61	0.84
	10.731–16.590	5767	6.29	686	6.97	1.11
	16.590–21.886	9184	10.02	326	3.31	0.33
	21.886–28.758	2176	2.37	7	0.07	0.03
PGA (g)	0.045–0.182	12,241	13.36	536	5.44	0.41
	0.182–0.244	23,945	26.13	2985	30.31	1.16
	0.244–0.288	38,745	42.28	4878	49.54	1.17
	0.288–0.371	14,222	15.52	1187	12.05	0.78
	0.371–0.510	1966	2.15	236	2.40	1.12
	0.510–0.705	523	0.57	25	0.25	0.44
Age of buildings (year)	1–17	36,688	40.03	3606	36.62	0.91
	18–32	34,320	37.45	3584	36.40	0.97
	33–59	13,275	14.49	1964	19.95	1.38
	60–98	6243	6.81	569	5.78	0.85
	99–172	1050	1.15	119	1.21	1.05
	173–562	66	0.07	5	0.05	0.71
Number of floors	1–2	72,680	79.31	7121	72.32	0.91
	3–4	12,901	14.08	1870	18.99	1.35
	5–7	3967	4.33	651	6.61	1.53
	8–12	1071	1.17	128	1.30	1.11
	13–16	786	0.86	62	0.63	0.73
	17–20	237	0.26	15	0.15	0.59

Table 2. Cont.

	Class	No. of Pixels in Building	Building (%)	No. of Pixels in Damaged Building	Damaged Building (%)	Frequency Ratio
Construction materials	Masonry	17,578	19.18	1642	16.68	0.7
	Concrete	27,048	29.51	3684	37.41	1.27
	Wood	11,096	12.11	1258	12.78	1.06
	Steel	35,262	38.48	3167	32.16	0.84
	Concrete + Steel	621	0.68	96	0.97	1.44
	Etc.	37	0.04	0	0.00	0.00
Density of buildings	0.476–156.351	65,002	70.93	6401	65.00	0.92
	156.351–376.410	10,396	11.34	914	9.28	0.82
	376.410–596.469	3409	3.39	412	4.18	1.23
	596.469–770.682	4205	4.59	675	6.85	1.49
	770.682–949.480	4586	5.00	732	7.43	1.49
	949.480–1169.540	4344	4.74	713	7.24	1.53
Child population (age < 15)	93–183	7735	8.44	868	8.81	1.04
	183–329	15,254	16.65	1823	18.51	1.11
	329–603	14,290	15.59	1431	14.53	0.93
	603–1020	8821	9.63	944	9.59	1.00
	1020–1279	20,238	22.08	2628	26.69	1.21
	1279–4944	25,304	27.61	2153	21.86	0.79
Elderly population (age ≥ 65)	526	2414	2.63	503	5.11	1.94
	526–1553	23,499	25.64	1441	14.63	0.57
	1553–2032	23,470	25.61	3026	30.73	1.20
	2032–2432	10,406	11.36	1354	13.75	1.21
	2432–3951	27,009	29.47	3410	34.63	1.17
	3951–6118	4844	5.29	113	1.15	0.22
Population density	23.390–82.713	16,580	18.09	1414	14.36	0.79
	82.713–201.358	41,645	45.44	3256	33.07	0.73
	201.358–586.957	11,329	12.36	2138	21.71	1.76
	586.957–2603.934	2674	2.92	40	4.09	1.40
	2603.934–5599.739	10,414	11.36	1065	10.82	0.95
	5599.739–7587.056	9000	9.82	1571	15.95	1.62
Distance from police stations (km)	0–1.205	36,638	39.98	4626	46.98	1.18
	1.205–2.458	19,864	21.68	2006	20.37	0.94
	2.458–3.807	17,359	18.96	1468	14.91	0.79
	3.807–5.350	10,379	11.33	1201	12.20	1.08
	5.350–8.145	6883	7.51	540	5.48	0.73
	8.145–12.291	519	0.57	6	0.06	0.11
Distance from fire stations (km)	0–1.431	34,930	38.12	4363	44.31	1.16
	1.431–2.766	22,245	24.27	2233	22.68	0.93
	2.766–4.102	14,318	15.62	1357	13.78	0.88
	4.102–5.533	12,231	13.35	1212	12.31	0.92
	5.533–8.204	7425	8.10	670	6.80	0.84
	8.204–12.164	493	0.54	12	0.12	0.23

Table 2. Cont.

Class	No. of Pixels in Building	Building (%)	No. of Pixels in Damaged Building	Damaged Building (%)	Frequency Ratio	
Distance from hospitals (km)	0–0.828	35,977	39.26	4115	41.79	1.06
	0.828–1.919	20,364	22.22	1976	20.07	0.90
	1.919–3.011	15,086	16.46	1637	16.62	1.01
	3.011–4.216	10,804	11.79	1044	10.60	0.90
	4.216–5.646	6744	7.36	840	8.53	1.16
	5.646–9.599	2667	2.91	235	2.39	0.82
Distance from roads (km)	0–0.116	54,351	59.31	6110	62.05	1.05
	0.116–0.311	22,932	25.02	2371	24.08	0.96
	0.311–0.610	9430	10.29	993	10.08	0.98
	0.610–1.025	2706	2.95	258	2.62	0.89
	1.025–1.609	1674	1.83	103	1.05	0.57
	1.609–3.310	549	0.60	12	0.12	0.20
Distance from gas stations (km)	0–0.680	47,099	51.39	5860	59.51	1.16
	0.680–1.391	19,988	21.81	2006	20.37	0.93
	1.391–2.195	13,483	14.71	1158	11.76	0.80
	2.195–3.091	6706	7.32	571	5.80	0.79
	3.091–4.390	3363	3.67	216	2.19	0.60
	4.390–7.884	1003	1.09	36	0.37	0.33

Importance scores for the various factors considered in the DT model are shown in Table 3. PGA was found to have the largest impact on building damage due to earthquakes (importance = 434.591), followed by epicenter distance (404.310) and distance from fire stations (307.873). Factors with the smallest impact on seismic vulnerability were related to construction materials (masonry, concrete, wood, steel, and concrete/steel mixture) and slope.

In the RF model, the percent mean square error (%IncMSE) and node purity (IncNodePurity) were determined as measures of factor importance in regression tree analysis. Maximum %IncMSE is reached when the variable with the highest value is removed from the model. An increase in IncNodePurity indicates a decrease in the Gini coefficient and includes a reduction in the residual sum of squares of the model. The Gini coefficient is a measure of tree node homogeneity; high Gini coefficient values indicate greater importance of the corresponding variable [67]. Epicenter distance (337.065) exhibited the highest %IncMSE, followed by distance from fire stations (325.576) and PGA (313.262) (Table 3). Epicenter distance (287.309) was found to be the most important factor based on IncNodePurity, followed by PGA (271.752) and altitude (254.792). Thus, epicenter distance and PGA have the greatest impact on seismic vulnerability according to the RF model, whereas factors related to construction materials are of low importance.

4.3. Seismic Vulnerability Mapping

Three seismic vulnerability maps were created based on data for all 71,888 buildings in Gyeongju. In the FR map, FR values were applied to each of the six classes for each sub-indicator. The final seismic vulnerability map was created by superimposing the resulting 18 sub-indicators. Seismic vulnerability maps based on the DT and RF models were created based on the prediction values of the models. In all three seismic vulnerability maps, indicator values were normalized to between 0 and 1, and then divided at equal intervals into five risk classes: safe, low risk, moderate risk, high risk, and very high risk.

Table 3. Importance variables of the DT and RF models.

Sub-indicators	Decision Tree	Random Forest	
	Importance	%IncMSE	IncNodePurity
Altitude	279.939	287.574	254.792
Slope	54.361	274.164	158.876
Groundwater level	202.317	233.859	243.677
Distance from faults	277.124	286.898	228.597
Distance from epicenters	404.310	337.065	287.309
PGA	434.591	313.262	271.752
Age of buildings	152.917	298.006	222.635
Number of floors	93.618	166.177	96.625
Construction materials			
Materials1 (masonry)	23.931	117.912	21.721
Materials2 (concrete)	43.296	68.501	20.897
Materials3 (wood)	0.000	84.983	13.467
Materials4 (steel)	72.115	122.198	47.492
Materials5 (concrete + steel)	0.000	39.827	2.189
Materials6 (etc.)	0.000	0.000	0.051
Density of buildings	240.093	287.399	202.289
Child population	169.186	124.942	62.086
Elderly population	273.094	114.642	81.323
Population density	192.077	168.966	115.059
Distance from police stations	284.950	308.095	201.307
Distance from fire stations	307.873	325.576	206.928
Distance from hospitals	211.459	290.069	204.312
Distance from roads	86.381	286.063	157.197
Distance from gas stations	251.988	302.629	198.339

Based on the resulting maps, the distribution of Gyeongju buildings' risk classes was compared among administrative districts. In the FR map, 589 buildings (0.82%) were classified as safe, 9999 (13.91%) as low risk, 36,172 (50.32%) as moderate risk, 21,299 (29.63%) as high risk, and 3829 (5.33%) as very high risk. Areas that are more vulnerable to earthquakes were then identified based on the sum of the proportions of buildings corresponding to high and very high risk. District 11 was found to be the most vulnerable district to earthquake damage, followed by districts 12, 9, 8, and 15. Among areas classified as safe and low risk, district 2 was found to be the safest, followed by districts 23, 1, 20, and 18 (Figure 6). In the DT map, 33,890 buildings (47.14%) were classified as safe, 13,621 (18.95%) as low risk, 9305 (12.94%) as moderate risk, 7593 (10.56%) as high risk, and 7479 (10.40%) as very high risk. The most vulnerable areas were districts 14, 7, 17, 15, and 8, whereas the safest areas were districts 19, 18, 23, 2, and 1 (Figure 7). In the RF map, 23,803 buildings (33.11%) were classified as safe, 26,429 (36.76%) as low risk, 13,669 (19.01%) as moderate risk, 6548 (9.11%) as high risk, and 1439 (2.00%) as very high risk. The most vulnerable areas were districts 14, 7, 15, 17, and 12, whereas the safest areas were districts 2, 18, 19, 23, and 5 (Figure 8). Figure 9 shows the building distribution by risk class.

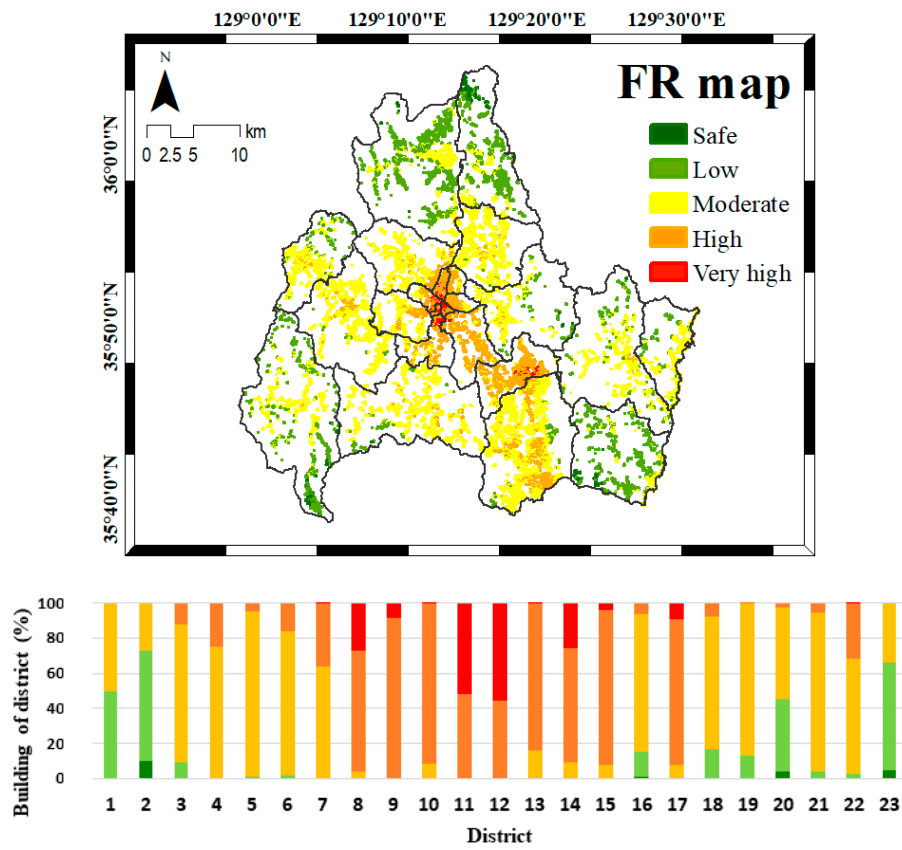


Figure 6. Seismic vulnerability map based on the FR model.

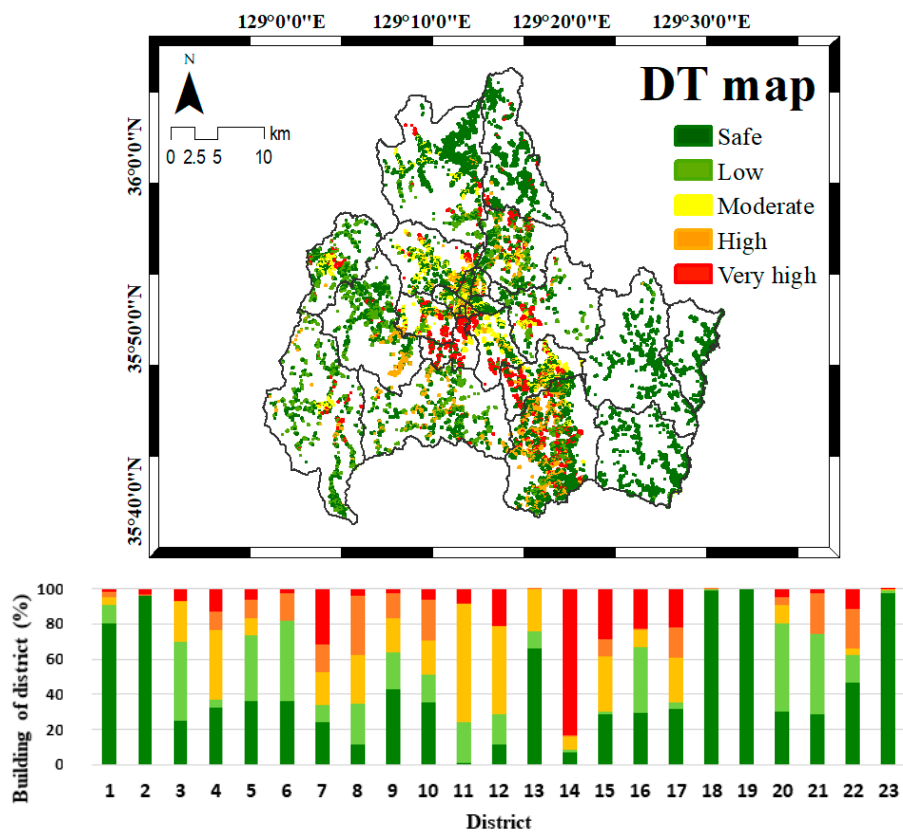


Figure 7. Seismic vulnerability map based on the DT model.

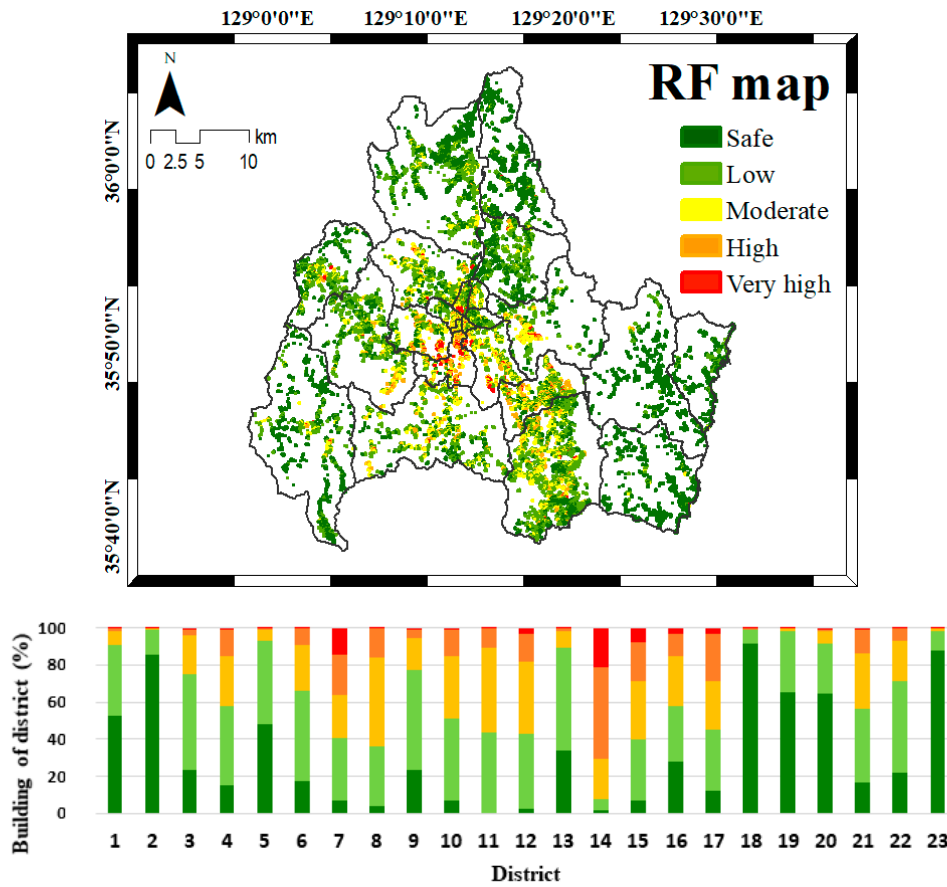


Figure 8. Seismic vulnerability map based on the RF model.

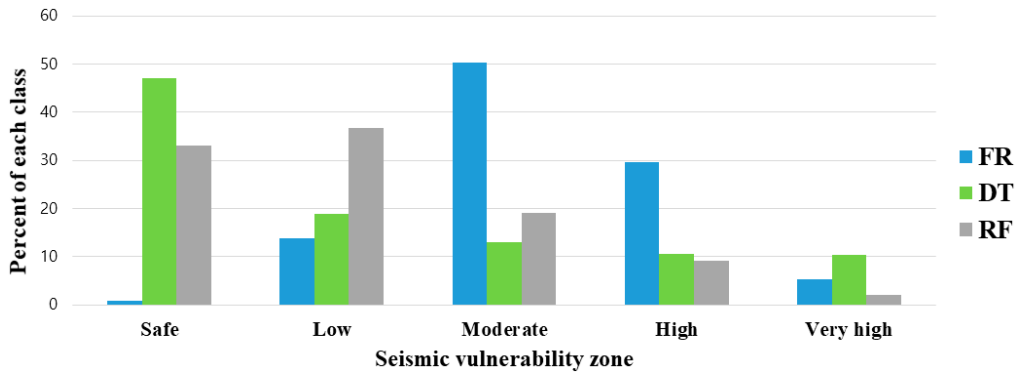


Figure 9. Percentages of different seismic vulnerability classes for the FR, DT, and RF models.

5. Discussion

In this study, three seismic vulnerability maps were created based on FR, DT, and RF methodologies, and their results were compared. First, we analyzed the importance of sub-indicators according to each methodology. Epicenter distance and PGA exhibited high importance in both the DT and RF models. Among all 9847 building cells, 5386 (54.70%) and 8756 (88.92%) corresponded to damaged buildings within 5 and 10 km of an epicenter, respectively. These results confirmed that most buildings close to epicenters were damaged; accordingly, this factor had a large influence on model construction. According to the seismic design criteria of South Korea, for an earthquake with a return period of 1000 years, the design ground acceleration of ground with normal rock (SB) is 0.154 g, whereas that of very dense ground (SC) is 0.18 g [3]. Based on these criteria, 9356 of cells corresponding to damaged

buildings (95.01%) were found in areas where PGA exceeded 0.18 g. Thus, most cells exhibited values higher than the design ground acceleration, which indicates that ground acceleration caused building damage during the 2016 Gyeongju Earthquake, and that PGA exerted a large influence on seismic vulnerability model construction. Factors exhibiting low importance included construction materials (masonry, concrete, wood, and steel/concrete mixture).

Among all damaged building cells, 3083 (31.31%) corresponded to buildings made of masonry or wood, which are relatively vulnerable construction materials. A much larger proportion of damaged building cells corresponded to concrete and steel, which are relatively strong construction materials. In addition, Gyeongju City, as a historic site, has many old buildings that correspond to the relatively weak wood and masonry. However, the corresponding buildings continue to be renovated to preserve historical values. Finally, it can be seen that most affected buildings are small buildings excluded from the seismic design targets (one- or two-story buildings with a floor area of less than 500 square meters) [68]. Therefore, construction materials are somewhat unsuitable for seismic vulnerability assessment.

Next, model success and prediction rates were analyzed to determine their functional differences. The RF model was found to be the most reliable among the three models, with the highest success (1.000) and prediction rates (0.949). The RF model complements the shortcomings of a single tree and operates well on large datasets; therefore, it performed best due to the relatively large number of datasets used in this study. The DT model showed the next best performance, with success and prediction rates of 0.899 and 0.851, respectively. The FR model showed success and prediction rates of 0.661 and 0.655, respectively, indicating low accuracy and underfitting, which prevents the reflection of important trends due to oversimplicity [69]. Therefore, the FR model is somewhat unsuitable for seismic vulnerability assessment.

Several studies of disaster susceptibility have also compared model performance among methodologies similar to those used in the present study. Xiao et al. (2019) [70] produced landslide susceptibility maps of Wanzhou County, China, using FR, certainty factor (CF), index of entropy (IOE), and RF methodologies. The prediction accuracy values of the models descended in the following order: RF (0.801), IOE (0.738), CF (0.732), and FR (0.728). Among the three statistical and probabilistic models, RF, based on machine learning, showed the highest accuracy, and IOE, based on weighted coefficients, showed the highest performance. Chen et al. (2017) [60] compared three tree-based data-mining techniques for the spatial prediction of landslide susceptibility: RF, CART, and logistic model tree (LMT). The prediction accuracy of the models descended in the following order: RF (0.781), LMT (0.752), and CART (0.742), with somewhat low overall accuracy. In a similar study, Pham et al. (2017) [71] created models based on four tree-based machine learning methods (RF, CART, LMT, and best first DTs (BFDT)) and compared their performance for landslide susceptibility assessment and mapping. The RF model exhibited the highest prediction accuracy (98.5%), followed by LMT (0.945), BFDT (0.934), and CART (0.933). Thus, several studies have found that tree-based machine learning models exhibited higher performance than statistical models, and RF models exhibited high performance in most studies, confirming their suitability for vulnerability analysis. In a previous study, Han et al. (2019) [20] used 15 factors except for social indicators to build LR and SVM kernel models to compare and analyze their performance. The results showed that the performance of the model based on the radial basis function (RBF) kernel (0.998) of SVM was the best, followed by polynomial (0.842), linear (0.649), LR (0.649), and sigmoid (0.630). The prediction rates were shown for RBF (0.919), polynomial (0.804), LR (0.655), linear (0.651), and sigmoid (0.629). The results showed with the prediction rates that the RF model was about 3% more accurate than the RBF kernel-based model.

Finally, we compared the seismic vulnerability maps created in this study. In all three maps, district 15 (Wolseong) was found to be the most dangerous area, whereas districts 2 (Gangdong), 18 (Yangbuk), and 23 (Yangnam) were identified as safe. Therefore, we mainly focused our sub-indicator characterization and comparison analyses on these districts. District 15 is located in central Gyeongju, with an epicenter distance of 2.798 km, fault distance of 4.269 km, PGA of 0.262 g, altitude of 62.825 m, and groundwater level of 15.078 m. In terms of its structural indicators, district 15 has a building density

and age of 322.834 and 43 years, respectively. Among all of the buildings in this district, 1521 (68.36%) are less than 50 years in age. District 15 has a population density of 237.167. In contrast, districts 2, 18, and 23 are located near the northern and southeastern coasts of Gyeongju, with epicenter and fault distances of 11.211 and 4.407 km, respectively, which are further than those of district 15. These districts have a PGA of 0.159 g, altitude of 54.165 m, and groundwater level of 10.292 m, which are lower than those of district 15, as well as a building density and age of 85.869 (ca. 3.76-fold lower than that of district 15) and 32 years, respectively. Among all buildings, 6469 (77.90%) are less than 50 years in age. These districts have a population density of 68.997, which is ca. 3.44-fold lower than that of district 15. There was no significant difference in the average values of five capacity-related factors (distance from hospitals, police stations, fire stations, roads, and gas stations) between dangerous and safe areas.

This study is meaningful in evaluating seismic vulnerability by comprehensively considering 18 factors related to geotechnical, physical, social, and capacity indicators, along with structural characteristics. It was also intended to derive a model suitable for the assessment of seismic vulnerability in Gyeongju by establishing models corresponding to various methodologies. Based on the results of the study, the seismic vulnerability assessment data provided in this study may be used as reference data for selecting parameters for seismic vulnerability assessments in other regions using more or fewer influence factors. The proposed method is also expected to contribute to improving seismic vulnerability assessment and mapping in domestic areas other than Gyeongju.

6. Conclusions

In this study, seismic vulnerability maps were created and seismic vulnerability assessment was performed for buildings in Gyeongju, South Korea using the probabilistic FR model and machine-learning-based DT and RF models. Models were created for each methodology using 18 factors affecting seismic vulnerability (slope, altitude, groundwater level, PGA, epicenter distance, fault distance, building age, construction materials, building density, number of floors, elderly population, child population, population density, and distances from hospitals, fire stations, police stations, roads, and gas stations) as independent variables and buildings damaged in the 2016 Gyeongju Earthquake as dependent variables. Epicenter distance and PGA were found to be the most important factors in the DT and RF models, and factors related to construction materials were the least important. These results may be used as reference data for models based on other methodologies. Model accuracy (success and prediction rates) was verified using ROC curves; the RF and FR models exhibited the highest and lowest performance, respectively, indicating that the machine-learning-based model is more suitable for seismic vulnerability assessment. Dangerous and safe areas were identified based on the seismic vulnerability maps created using the three models; in all three maps, district 15 was found to be the most dangerous area and districts 2, 18, and 23 were the safest areas. Therefore, district 15 must be managed first in preparation for future earthquakes. The seismic vulnerability maps created in this study facilitate intuitive identification of dangerous districts within the target area, which will prevent greater damage in future earthquakes through the establishment of evacuation routes for residents. As reference data, our findings may be used for developing earthquake-related policies and determining suitable locations for vulnerable infrastructure (e.g., pipelines or high-voltage facilities), as well as important national facilities (e.g., airports, military facilities, and nuclear power plants).

Author Contributions: Conceptualization, S.P.; data curation, J.H.; funding acquisition, J.K.; methodology, S.P.; project administration, J.K.; software, S.S.; supervision, J.K.; validation, J.H.; visualization, J.H. and M.R.; writing—original draft, J.H.; writing—review & editing, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry, and Energy (MOTIE) of the Republic of Korea (No. 201715101960).

Acknowledgments: This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry, and Energy (MOTIE) of the Republic of Korea (No. 201715101960).

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

References

1. 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.* **2016**, *20*, 747–752. [[CrossRef](#)]
2. 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.* **2016**, *20*, 753–757. [[CrossRef](#)]
3. Ministry of Public Safety and Security (MPSS). *Report on the 9.12 Earthquake and Countermeasures*; MPSS: Seoul, Korea, 2017.
4. Wallemacq, P. *Economic Losses, Poverty & Disasters: 1998–2017*; Centre for Research on the Epidemiology of Disasters: Brussels, Belgium, 2018.
5. Armaş, I. Multi-Criteria Vulnerability Analysis to Earthquake Hazard of Bucharest, Romania. *Nat. Hazards* **2012**, *63*, 1129–1156. [[CrossRef](#)]
6. 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. [[CrossRef](#)]
7. Sadrykia, M.; Delavar, M.R.; Zare, M. A GIS-Based Decision Making Model using Fuzzy Sets and Theory of Evidence for Seismic Vulnerability Assessment Under Uncertainty (Case Study: Tabriz). *J. Intell. Fuzzy Syst.* **2017**, *33*, 1969–1981. [[CrossRef](#)]
8. 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. [[CrossRef](#)]
9. Nath, S.; Adhikari, M.; Devaraj, N.; Maiti, S. Seismic Vulnerability and Risk Assessment of Kolkata City, India. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 1103. [[CrossRef](#)]
10. 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. [[CrossRef](#)]
11. 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. [[CrossRef](#)]
12. Alizadeh, M.; Alizadeh, E.; Asadollahpour Kotenaee, 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. [[CrossRef](#)]
13. Moradi, M.; Delavar, M.R.; Moshiri, B. A GIS-Based Multi-Criteria Decision-Making Approach for Seismic Vulnerability Assessment using Quantifier-Guided OWA Operator: A Case Study of Tehran, Iran. *Ann. GIS* **2015**, *21*, 209–222. [[CrossRef](#)]
14. Nyimbili, P.H.; Erden, T.; Karaman, H. Integration of GIS, AHP and TOPSIS for Earthquake Hazard Analysis. *Nat. Hazards* **2018**, *92*, 1523–1546. [[CrossRef](#)]
15. Alam, M.S.; Haque, S.M. Assessment of Urban Physical Seismic Vulnerability using the Combination of AHP and TOPSIS Models: A Case Study of Residential Neighborhoods of Mymensingh City, Bangladesh. *J. Geosci. Environ. Prot.* **2018**, *6*, 165. [[CrossRef](#)]
16. Lee, S.; Panahi, M.; Pourghasemi, H.R.; Shahabi, H.; Alizadeh, M.; Shirzadi, A.; Khosravi, K.; Melesse, A.M.; Yekrangnia, M.; Rezaie, F. Sevucas: A Novel Gis-Based Machine Learning Software for Seismic Vulnerability Assessment. *Appl. Sci.* **2019**, *9*, 3495. [[CrossRef](#)]
17. Yariyan, P.; Avand, M.; Soltani, F.; Ghorbanzadeh, O.; Blaschke, T. Earthquake Vulnerability Mapping using Different Hybrid Models. *Symmetry* **2020**, *12*, 405. [[CrossRef](#)]
18. 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. [[CrossRef](#)]
19. 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. [[CrossRef](#)]

20. Han, J.; Park, S.; Kim, S.; Son, S.; Lee, S.; Kim, J. 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. *Sustainability* **2019**, *11*, 7038. [[CrossRef](#)]
21. 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). *Geomat. Nat. Hazards Risk* **2019**, *10*, 958–985. [[CrossRef](#)]
22. 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. [[CrossRef](#)]
23. Al-Abadi, A.M. Mapping Flood Susceptibility in an Arid Region of Southern Iraq using Ensemble Machine Learning Classifiers: A Comparative Study. *Arab. J. Geosci.* **2018**, *11*, 218. [[CrossRef](#)]
24. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Sajedi-Hosseini, F.; Mosavi, A. An Ensemble Prediction of Flood Susceptibility using Multivariate Discriminant Analysis, Classification and Regression Trees, and Support Vector Machines. *Sci. Total Environ.* **2019**, *651*, 2087–2096. [[CrossRef](#)]
25. Tehrany, M.S.; Kumar, L.; Shabani, F. A Novel GIS-Based Ensemble Technique for Flood Susceptibility Mapping using Evidential Belief Function and Support Vector Machine: Brisbane, Australia. *PeerJ* **2019**, *7*, e7653. [[CrossRef](#)] [[PubMed](#)]
26. Chen, W.; Li, Y.; Xue, W.; Shahabi, H.; Li, S.; Hong, H.; Wang, X.; Bian, H.; Zhang, S.; Pradhan, B. Modeling Flood Susceptibility using Data-Driven Approaches of Naïve Bayes Tree, Alternating Decision Tree, and Random Forest Methods. *Sci. Total Environ.* **2020**, *701*, 134979. [[CrossRef](#)] [[PubMed](#)]
27. 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. [[CrossRef](#)]
28. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial Prediction Models for Shallow Landslide Hazards: A Comparative Assessment of the Efficacy of Support Vector Machines, Artificial Neural Networks, Kernel Logistic Regression, and Logistic Model Tree. *Landslides* **2016**, *13*, 361–378.
29. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide Susceptibility Mapping using Random Forest, Boosted Regression Tree, Classification and Regression Tree, and General Linear Models and Comparison of their Performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* **2016**, *13*, 839–856. [[CrossRef](#)]
30. 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. [[CrossRef](#)]
31. Park, S.; Hamm, S.; Kim, J. Performance Evaluation of the Gis-Based Data-Mining Techniques Decision Tree, Random Forest, and Rotation Forest for Landslide Susceptibility Modeling. *Sustainability* **2019**, *11*, 5659. [[CrossRef](#)]
32. Wang, Y.; Wu, X.; Chen, Z.; Ren, F.; Feng, L.; Du, Q. Optimizing the Predictive Ability of Machine Learning Methods for Landslide Susceptibility Mapping using Smote for Lishui City in Zhejiang Province, China. *Int. J. Environ. Res. Public Health* **2019**, *16*, 368. [[CrossRef](#)]
33. Nhu, V.; Shirzadi, A.; Shahabi, H.; Chen, W.; Clague, J.J.; Geertsema, M.; Jaafari, A.; Avand, M.; Miraki, S.; Talebpour Asl, D. Shallow Landslide Susceptibility Mapping by Random Forest Base Classifier and its Ensembles in a Semi-Arid Region of Iran. *Forests* **2020**, *11*, 421. [[CrossRef](#)]
34. Nhu, V.; Shirzadi, A.; Shahabi, H.; Singh, S.K.; Al-Ansari, N.; Clague, J.J.; Jaafari, A.; Chen, W.; Miraki, S.; Dou, J. Shallow Landslide Susceptibility Mapping: A Comparison between Logistic Model Tree, Logistic Regression, Naïve Bayes Tree, Artificial Neural Network, and Support Vector Machine Algorithms. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2749. [[CrossRef](#)]
35. Avand, M.; Janizadeh, S.; Naghibi, S.A.; Pourghasemi, H.R.; Khosrobeigi Bozchaloei, S.; Blaschke, T. A Comparative Assessment of Random Forest and k-Nearest Neighbor Classifiers for Gully Erosion Susceptibility Mapping. *Water* **2019**, *11*, 2076. [[CrossRef](#)]
36. Garosi, Y.; Sheklabadi, M.; Conoscenti, C.; Pourghasemi, H.R.; Van Oost, K. Assessing the Performance of GIS-Based Machine Learning Models with Different Accuracy Measures for Determining Susceptibility to Gully Erosion. *Sci. Total Environ.* **2019**, *664*, 1117–1132. [[CrossRef](#)]
37. Pourghasemi, H.R.; Yousefi, S.; Kornejady, A.; Cerdà, A. Performance Assessment of Individual and Ensemble Data-Mining Techniques for Gully Erosion Modeling. *Sci. Total Environ.* **2017**, *609*, 764–775. [[CrossRef](#)] [[PubMed](#)]

38. Tien Bui, D.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Omidav, E.; Pham, B.T.; Talebpour Asl, D.; Khaledian, H.; Pradhan, B.; Panahi, M. A Novel Ensemble Artificial Intelligence Approach for Gully Erosion Mapping in a Semi-Arid Watershed (Iran). *Sensors* **2019**, *19*, 2444. [[CrossRef](#)] [[PubMed](#)]
39. Nhu, V.; Janizadeh, S.; Avand, M.; Chen, W.; Farzin, M.; Omidvar, E.; Shirzadi, A.; Shahabi, H.; Clague, J.J.; Jaafari, A. Gis-Based Gully Erosion Susceptibility Mapping: A Comparison of Computational Ensemble Data Mining Models. *Appl. Sci.* **2020**, *10*, 2039. [[CrossRef](#)]
40. Miraki, S.; Zanganeh, S.H.; Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Pham, B.T. Mapping Groundwater Potential using a Novel Hybrid Intelligence Approach. *Water Resour. Manag.* **2019**, *33*, 281–302. [[CrossRef](#)]
41. Nhu, V.; Rahmati, O.; Falah, F.; Shojaei, S.; Al-Ansari, N.; Shahabi, H.; Shirzadi, A.; Górski, K.; Nguyen, H.; Ahmad, B.B. Mapping of Groundwater Spring Potential in Karst Aquifer System using Novel Ensemble Bivariate and Multivariate Models. *Water* **2020**, *12*, 985. [[CrossRef](#)]
42. Tien Bui, D.; Shirzadi, A.; Chapi, K.; Shahabi, H.; Pradhan, B.; Pham, B.T.; Singh, V.P.; Chen, W.; Khosravi, K.; Bin Ahmad, B. A Hybrid Computational Intelligence Approach to Groundwater Spring Potential Mapping. *Water* **2019**, *11*, 2013. [[CrossRef](#)]
43. Chen, W.; Li, Y.; Tsangaratos, P.; Shahabi, H.; Ilia, I.; Xue, W.; Bian, H. Groundwater Spring Potential Mapping using Artificial Intelligence Approach Based on Kernel Logistic Regression, Random Forest, and Alternating Decision Tree Models. *Appl. Sci.* **2020**, *10*, 425. [[CrossRef](#)]
44. Chen, W.; Zhao, X.; Tsangaratos, P.; Shahabi, H.; Ilia, I.; Xue, W.; Wang, X.; Ahmad, B.B. Evaluating the Usage of Tree-Based Ensemble Methods in Groundwater Spring Potential Mapping. *J. Hydrol.* **2020**, *583*, 124602. [[CrossRef](#)]
45. Şengezer, B.; Ansal, A.; Bilen, Ö. Evaluation of Parameters Affecting Earthquake Damage by Decision Tree Techniques. *Nat. Hazards* **2008**, *47*, 547–568. [[CrossRef](#)]
46. 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. [[CrossRef](#)]
47. Ahmed, M.; Morita, H. An Analysis of Housing Structures' Earthquake Vulnerability in Two Parts of Dhaka City. *Sustainability* **2018**, *10*, 1106. [[CrossRef](#)]
48. Gyeongju City Hall. Available online: <http://www.gyeongju.go.kr/> (accessed on 10 March 2020).
49. Korea Meteorological Administration. Available online: <http://www.weather.go.kr/> (accessed on 17 March 2020).
50. 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. [[CrossRef](#)]
51. 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.
52. Lee, M.; Kang, J. Predictive Flooded Area Susceptibility and Verification using GIS and Frequency Ratio. *J. Korean Assoc. Geogr. Inf. Stud.* **2012**, *15*, 86–102. [[CrossRef](#)]
53. Son, J. Susceptibility Assessment of Landslide and Land Subsidence Applying the Radius of Influence to Frequency Ratio Model. Ph.D. Thesis, Graduate School of Seoul National University, Seoul, Korea, 2017.
54. 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. [[CrossRef](#)]
55. Saito, H.; Nakayama, D.; Matsuyama, H. Comparison of Landslide Susceptibility Based on a Decision-Tree Model and Actual Landslide Occurrence: The Akaishi Mountains, Japan. *Geomorphology* **2009**, *109*, 108–121. [[CrossRef](#)]
56. 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. [[CrossRef](#)]
57. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.
58. Choi, J.; Seo, D. Application of Data Mining Decision Tree. *Res. Stat. Anal.* **1999**, *4*, 61–83.
59. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. An Assessment of Multivariate and Bivariate Approaches in Landslide Susceptibility Mapping: A Case Study of Duzkoy District. *Nat. Hazards* **2015**, *76*, 471–496. [[CrossRef](#)]

60. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; Duan, Z.; Ma, J. A Comparative Study of Logistic Model Tree, Random Forest, and Classification and Regression Tree Models for Spatial Prediction of Landslide Susceptibility. *Catena* **2017**, *151*, 147–160. [[CrossRef](#)]
61. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
62. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
63. Park, S.; Kim, J. Landslide Susceptibility Mapping Based on Random Forest and Boosted Regression Tree Models, and a Comparison of their Performance. *Appl. Sci.* **2019**, *9*, 942. [[CrossRef](#)]
64. Kavzoglu, T.; Colkesen, I.; Sahin, E.K. Machine learning techniques in landslide susceptibility mapping: A survey and a case study. In *Landslides: Theory, Practice and Modelling*; Springer: Cham, Switzerland, 2019; pp. 283–301.
65. 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. [[CrossRef](#)]
66. Taalab, K.; Cheng, T.; Zhang, Y. Mapping Landslide Susceptibility and Types using Random Forest. *Big Earth Data* **2018**, *2*, 159–178. [[CrossRef](#)]
67. Paul, S.S.; Li, J.; Li, Y.; Shen, L. Assessing Land use–land Cover Change and Soil Erosion Potential using a Combined Approach through Remote Sensing, RUSLE and Random Forest Algorithm. *Geocarto Int.* **2019**, 1–15. [[CrossRef](#)]
68. Kang, T.H.; Jeong, S.Y.; Kim, S.; Hong, S.; Choi, B.J. A Comparative Case Study of 2016 Gyeongju and 2011 Virginia Earthquakes. *J. Earthq. Eng. Soc. Korea* **2016**, *20*, 443–451. [[CrossRef](#)]
69. Arredondo Parra, Á. Application of Machine Learning Techniques for the Estimation of Seismic Vulnerability in the City of Port-au-Prince (Haiti). Master’s Thesis, Universidad Politécnica de Madrid, Madrid, Spain, 2019.
70. Xiao, T.; Yin, K.; Yao, T.; Liu, S. Spatial Prediction of Landslide Susceptibility using GIS-Based Statistical and Machine Learning Models in Wanzhou County, Three Gorges Reservoir, China. *Acta Geochim.* **2019**, *38*, 654–669. [[CrossRef](#)]
71. Pham, B.T.; Khosravi, K.; Prakash, I. Application and Comparison of Decision Tree-Based Machine Learning Methods in Landside Susceptibility Assessment at Pauri Garhwal Area, Uttarakhand, India. *Environ. Process.* **2017**, *4*, 711–730. [[CrossRef](#)]



© 2020 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/>).