



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

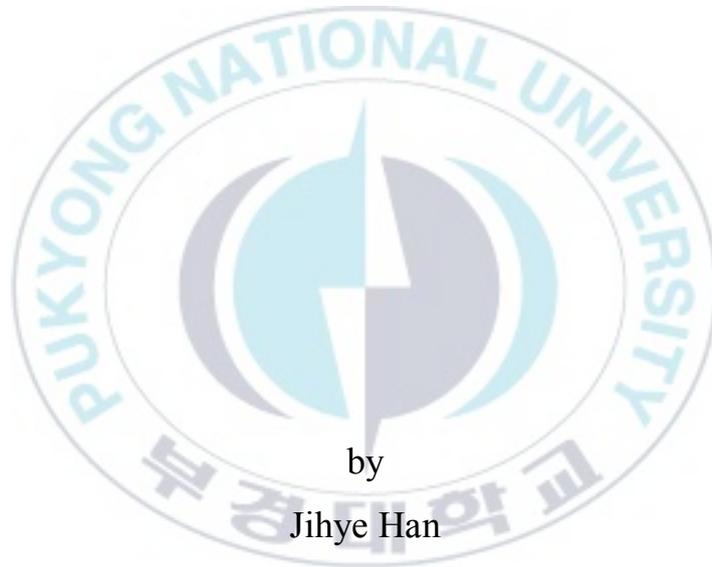
저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Thesis for the Degree of Master of Engineering

Seismic Vulnerability Assessment and Mapping: A Case Study of Gyeongju Earthquake on September 12, 2016



by

Jihye Han

Department of Spatial Information Engineering

The Graduate School

Pukyong National University

February 2020

Seismic Vulnerability Assessment and Mapping: A Case Study of Gyeongju Earthquake on September 12, 2016

(지진 취약성 평가 및 지도제작

: 2016년 9월 12일

경주 지진을 중심으로)

Advisor: Prof. Jinsoo Kim

by

Jihye Han

A thesis submitted in partial fulfillment of the requirements
for the degree of
Master of Engineering
in Department of Spatial Information Engineering, The Graduate School,
Pukyong National University

February 2020

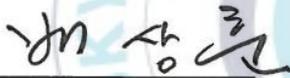
Seismic Vulnerability Assessment and Mapping
: A Case Study of Gyeongju Earthquake on September 12, 2016

A dissertation

by

Jihye Han

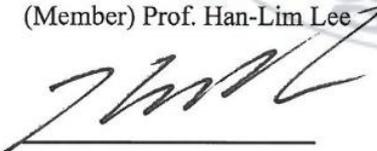
Approved by:



(Chairman) Prof. Sang-Hoon Bae



(Member) Prof. Han-Lim Lee



(Member) Prof. Jin-Soo Kim

February, 2020

CONTENTS

CONTENTS	I
LIST OF FIGURES	III
LIST OF TABLES	IV
1. Introduction	1
1.1. Background	1
1.2. Literature review	3
1.3. Objectives.....	8
2. Data	10
2.1. Study area.....	10
2.2. 9.12 Gyeongju Earthquake inventory	12
2.3. Spatial database preparation.....	13
2.3.1. Geotechnical indicator.....	17
2.3.2. Physical indicator	17
2.3.3. Structural indicator.....	18
2.3.4. Social indicator.....	19
2.3.5. Capacity indicator.....	19
3. Methodology.....	21
3.1. Frequency ratio.....	21
3.2. Logistic regression.....	22
3.3. Support vector machine	23
3.4. Random forest	25
4. Results	27
4.1. Validation of model	27
4.2. Importance factor.....	31
4.3. Seismic vulnerability map.....	41
4.4. Building class distribution by Gyeongju districts	46

5. Discussion	48
6. Conclusions	53
References	55

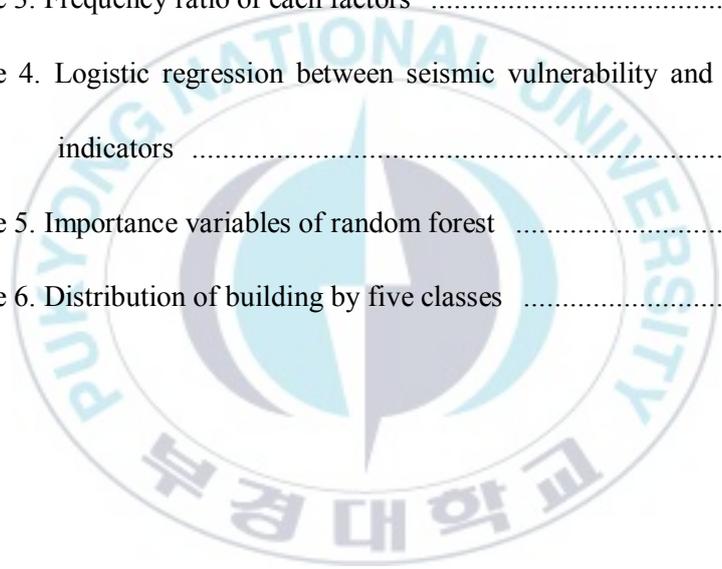


LIST OF FIGURES

Fig. 1. Flowchart of this study.	9
Fig. 2. The study area.	11
Fig. 3. Training and validation datasets.	12
Fig. 4. Sub-indicators related earthquake.	14
Fig. 5. ROC curve (Success rate).	29
Fig. 6. ROC curve (Prediction rate).	30
Fig. 7. Seismic vulnerability map.	45
Fig. 8. Ratio of vulnerable buildings of Gyeongju districts by each model.	47

LIST OF TABLES

Table 1. Independent variables for the seismic vulnerability assessment	16
Table 2. Optimal parameters of kernels	28
Table 3. Frequency ratio of each factors	33
Table 4. Logistic regression between seismic vulnerability and related indicators	38
Table 5. Importance variables of random forest	40
Table 6. Distribution of building by five classes	44



지진 취약성 평가 및 지도제작
: 2016년 9월 12일 경주 지진을 중심으로

한 지 혜

부 경 대 학 교 대 학 원 공 간 정 보 시 스템 공 학 과

요 약

2016년 9월 12일 대한민국 경주시 남남서쪽 8 km 지점에서 발생한 규모 5.8의 지진은 한반도 대부분의 지역에서 진동을 감지할 수 있을 정도의 대규모 지진으로, 이로 인해 수많은 인명 피해 및 경제적 손실 등을 초래하였다. 추후 9.12 경주 지진과 같은 대규모 지진의 발생 가능성을 배제할 수 없는 실정을 반영하여 본 연구에서는 경주시의 지진 취약성을 평가하고자 한다. 먼저, 지진 취약성에 영향을 미치는 5 가지 주요 지표 (geotechnical, physical, social, structural, capacity)를 선정하였으며, 이와 관련하여 18 개의 하위 지표 (elevation, slope, groundwater level, PGA, distance to faults, distance to epicenters, age of buildings, number of floors, construction materials, density of buildings, child population, elderly population, population density, distance to roads, distance to hospitals, distance to police stations, distance to gas stations, distance to fire stations)를 10 m 공간 해상도로 구축하여 독립변수로 사용하였다. 종속변수로는 9.12 경주 지진 발생 당시 실제 지진으로 인해 피해 입은 건물들의 위치 자료를 적용하였다. 분석에는 통계적 방법론인 logistic regression (LR), 확률적 방법론인 frequency ratio (FR), 기계학습 방법인 support vector machine (SVM)과 random forest (RF) 방법론을 사용하였다. 훈련 및 검증 데이터셋은 7:3 비율로 무작위 선별되었으며, 정확도 검증은 receiver operating characteristic (ROC) 곡선을 사용하여 최적 모델을 선별하였다. 비교 결과, RF 모델 (100%) 및 예측 (94.9%) 정확도가 가장 높게 산출되었으며, RBF-SVM, FR, LR 순서로 높은 성능을 나타냈다. 각 방법론 별 가장 정확도가 높은 모델을 기반으로 경주시 전체 건축물의 예측 값을 산출하였으며, 이를 기반으로 지진 취약성 지도를 제작하였다. 예측 값은 0~1 범위 값으로 정규화 하여 5 개의 동일 간격으로 분류하였으며, 이를 토대로 23 개의 행정동 별로 고위험 및 안전 지역을 도출하였다. 4 개의 지도를 검토한 결과 공통적으로 강동면이 가장 안전지역으로, 3 개의 지도에서 황남동 및 월성동이 취약지역으로 도출되었다. 본 연구에서 제작한 지진 취약성 지도는 환경, 토지, 시설물 관리 및 정책 수립 등 다양한 분야에서 지진으로 인한 피해를 저감하기 위한 기초 자료로 활용 가능할 것으로 기대된다.

1. Introduction

1.1. Background

Natural disasters cause physical damage, such as building damage, and human, environmental and economic losses, due to unforeseen changes in the environment such as earthquakes, landslides, and floods (Kavzoglu et al., 2014). The earthquake is a serious threat to human life and safety and, most of the country's most destructive power of highly classified as natural disasters (Alizadeh et al., 2018; Bahadori et al., 2017). According to a United Nations report, earthquake and volcanic-related disasters account for about 10 percent of natural disasters over a period of about 10 years (1998-2017) (Wallemacq and House, 2018). Among them, economic damage from earthquakes accounted for about 23 percent of the total damage caused by natural disasters, with deaths from earthquakes accounting for about 56 percent of the total deaths. This means that although earthquakes have a relatively low incidence rate compared to other natural disasters, once they occur, the damage is very large (Lee et al., 2018).

The Korean Peninsula belongs to the plate where earthquakes can occur due to the accumulation of local stress caused by the plate tectonic

movement, and it has the characteristics of an intraplate earthquake. Intraplate earthquake is infrequent and irregular, which the time and space distribution of earthquakes is irregular compared to interplate earthquake and cannot be easily predicted (Shin et al., 2016). Since the 20th century, there have been fewer earthquakes of medium or greater magnitude on the Korean Peninsula, and there has been no significant change in seismic activity (Lee, 2010).

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 (Kim et al., 2016a; 2016b). The 9.12 Gyeongju Earthquake was recorded as the largest earthquake since South Korea began measuring earthquakes in 1978 (Kim et al., 2016a; 2016b). It reported tremors were detected in most parts of the country, with 23 injured and 5,368 property damage reported (MPSS, 2017).

Cities can easily amplify damage in the event of an earthquake disaster, and because of the concentration of various facilities and populations, the resulting ripple effects can be seen in the long term, resulting in great economic damage (NEMA 2012). Natural hazards management should be promoted in order to mitigate losses on multiple sides (Pourghasemi et al.,

2019), in advance of sustainable management. Based on this, we can reduce the overall damage caused by the earthquake in case of a disaster (Moustafa et al., 2016; Walker et al., 2014).

1.2. Literature review

In the past several years, many studies of natural disasters have implemented various methods to assess vulnerability, resulting in the development of maps.

A study based on probabilistic and statistical methods, Yalcin et al. (2011) is a comparison of five risk studies that performed landslide susceptibility mapping in the Turkey Trabzon region using the analytical hierarchy process (AHP) and bivariate statistics, FR, and LR methodologies. Rahmati et al. (2015) used FR and weights-of-evidence (WoE) models for flood susceptibility mapping, with 76.47% and 74.74% prediction accuracy. Youssef et al. (2015) also produced FR, FR-LR ensemble models, with 91.3% and 89.6% of their predicted accuracy. In the Khosravi et al. (2016) study, the generation of FR, WoE, AHP, and FR-AHP ensemble models resulted in 96.57%, 95.96%, 94.92%, and 84.69% of each prediction accuracy. Based on these criteria, the level of risk was 5 and flooding usability mapping was performed. Compared to

the ensemble model, the predictive accuracy of the model with a single methodology was generally higher, and the probability-based FR model was found to be more accurate.

A study based on machine learning, Conforti et al. (2014) produced a map of landslide susceptibility in the Turbolo River Catchment, North Calabria, South Italy basin using the artificial neural networks (ANN) methodology, with a prediction accuracy of 87%. Kim et al. (2018) generated a boosted tree and RF model, a machine learning method affiliated with the tree, to perform a comparison verification based on the landslide susceptibility mapping of Pyeongchang, Korea. The predictive accuracy of the Boosted tree model was 84.87%, about 5% higher than that of RF.

We can also see studies based on probability, statistical methods, and machine learning methodologies, associated with landslide vulnerabilities.

Yilmaz (2010) maps and compares the probability-based conditional probability (CP) model with the statistics-based LR model, machine-learning-based ANN and SVM model. As a result of the validation, ANN (0.846) showed the highest accuracy, followed by SVM (0.841), LR (0.831), and CP (0.827). Wang et al. (2016) used the FR, LR, ANN, DT, and WoE methodologies, and produced three versions of training data to evaluate the accuracy of the model. Models created by randomly

classifying training data within the study area at 5:5 ratio showed generally high performance, with high predictability in the order LR, FR, WOE, DT, and ANN. Chen et al. (2019) created the WoE, WoE-LR and WoE-RF models, which hybrid models based on WoE., to compare and analyze their performance. The prediction accuracy was 69.5%, 76.3%, and 78.2%, indicating that the WoE-RF had the best performance, and demonstrated the superiority of the hybrid model. In the Kadavi et al. (2019) study, LR and DT models were used, and the DT model generated three models (CHAID, exhaustive CHAID, and QUEST) based on three algorithms. As a result, based on DT the exhaustive chaid algorithm (0.906) the highest prediction accuracy, and the DT CHAID (0.902), LR (0.901), and DT QUEST (0.843) are so accurate in order. In general, many studies were conducted with probability and statistics-based methodologies to compare with machine learning, and the accuracy of machine learning models was found to be relatively high. In the paper using the SVM, it was found that the model was created mainly based on the RBF kernel

Whereas the vulnerability study applied with earthquakes is relatively inadequate compared to natural disasters such as landslides and floods. Seismic vulnerabilities are multi-criteria decision problems that inherently sustainable development (Amiri et al., 2007), in which AHP and multi-criteria decision analysis (MCDA) methodologies based on the

geographies information system (GIS) were used in many studies to assess them. If multiple targets require evaluation, these are stratified and their importance quantified to determine the relative priorities of their criteria, using weighting of factor (Aliabadi et al., 2015; Armaş, 2012; Bahadori et al., 2017; Panahi et al., 2014; Rezaie and Panahi, 2015; Walker et al., 2014).

Recent studies have shown attempts to assess seismic vulnerability by applying machine learning, although not enough cases are available. Machine learning analyzes and predicts data based on automatic learning of statistical rules and patterns from large volumes of data (Kim and Yoon, 2018), and has proven applicable in a variety of areas (Lary et al., 2016). Şengezer et al. (2008) used the decision tree (DT) technique to evaluate parameters affecting earthquake damage, while Borfecchia et al. (2010) used DT and ANN each data mining method to estimate the urban vulnerability. Tesfamariam and Liu (2010) used support vector machine (SVM) and random force (RF) and six other classification techniques, while Guettiche et al. (2017) used association rule learning (ARL) techniques to perform seismic vulnerability assessment of buildings. Riedel et al. (2015) and Liu et al. (2019) presented methods for estimating seismic vulnerabilities of buildings from SVMs and ARLs based on building data. Alizadeh et al. (2018) Iran through the Seismic risk model

is created based on the ANN Tabriz city of social vulnerability assessment, and Ahmed and Morita (2018) analyzed seismic vulnerabilities in residential buildings in Dhaka City, Bangladesh, based on RF and DT.

Studies based on machine learning have assessed the seismic vulnerability of target areas using seismic factors, with a focus on buildings. Although the seismic vulnerability assessment applied to a single factor, such as lipids, buildings, and social factors, was conducted by several researchers, but the study considering the various factors comprehensively was insufficient. It was also possible to find that there were insufficient research cases in which models were created and compared based on different types of methodologies.

1.3. Objectives

In this study aimed to evaluate and mapping of seismic vulnerability targeting all buildings in Gyeongju city, South Korea. To performed this, we used four methodologies; probabilistic methodologies such as frequency ratio (FR), statistical methodology such as logistic regression (LR), and support vector machine (SVM) and random forest (RF) methodology, which are evaluated as the most robust of machine learning methodologies. The SVM has created a models with four kernels (linear, polynomial, radial function, and sigmoid) and applied 18 sub-indicators related to geotechnical, physical, structural, social, and capacity indicators to a total of seven models. The accuracy of each model was verified using the relative operating characteristic (ROC) curve, and a seismic vulnerability map was produced to evaluate the target regions according to administrative district. The flow chart of this study is shown in Fig. 1.

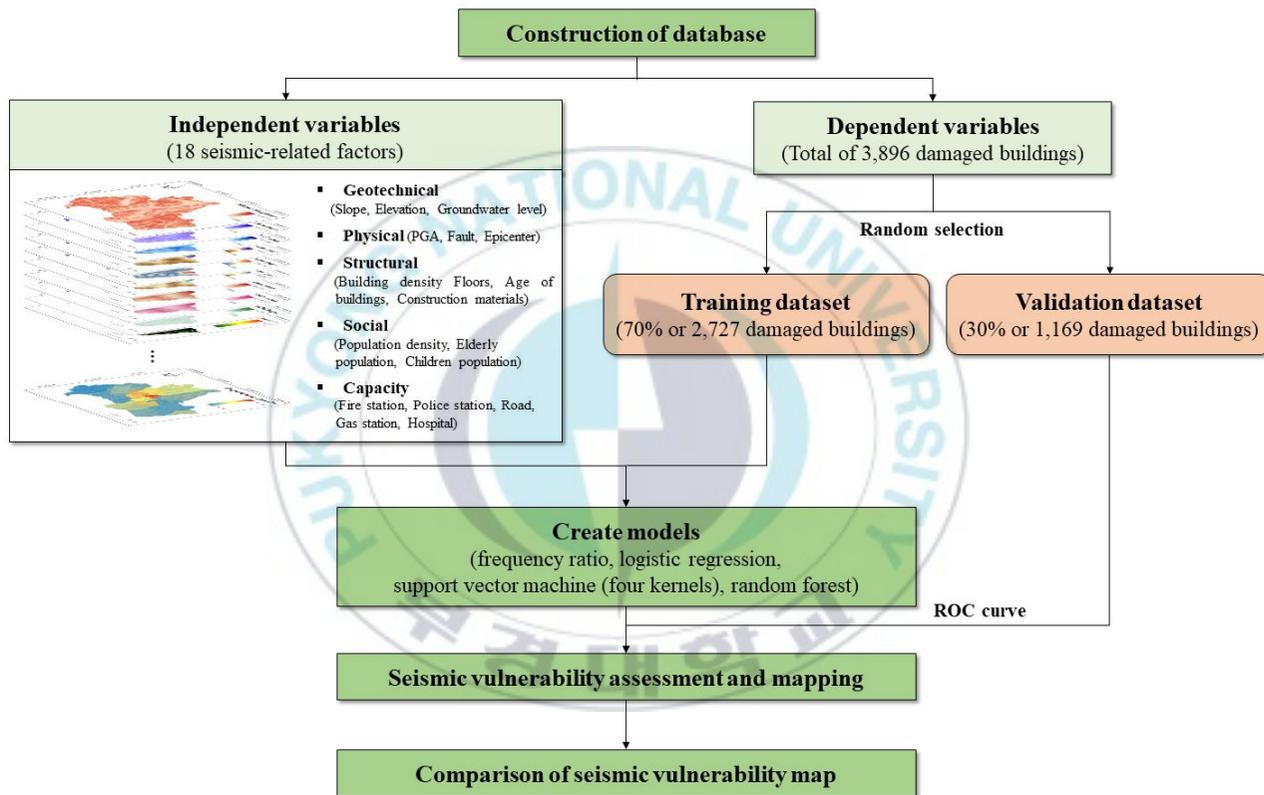


Fig. 1. Flow chart of this study.

2. Data

2.1. Study area

The target region of the present study was Gyeongju, Gyeongsangbuk-do, South Korea, is located between 35° 39'N and 36° 04'N in latitude and between 128° 58'E and 129° 31'E in longitude, which adjoining the East Sea in the east; Cheongdo County and Yeongcheon City, Gyeongsangbuk-do, in the west; Ulju County, Ulsan, in the south; and Pohang City, Gyeongsangbuk-do, in the north. Comprising four eups, eight myeons, and eleven administrative districts, Gyeongju has a size of 1,324.82 km² and population of 256,141 (Fig. 2).

Of the total area, Gyeongju has 67.9 percent of forest land, 16.8 percent of farmland and 15.3 percent of other land areas, with a large proportion of forest areas and low cultivation rates. Several faults are distributed, including the Ulsan and Yangsan faults, and quaternary fault movement has been reported along the Dongrae, Moryang, Miryang, and Ilkwang faults (Kim et al., 2017). These geographical features can contribute to higher probability of earthquake occurrence in the future, and secondary natural disasters are also expected.

According to the historical records, of the total earthquake occurrences in the south-eastern Korean Peninsula, there have been 75 earthquakes in Gyeongju, and before the 9.12 Gyeongju Earthquake, there had been 21 instrumental earthquakes (MPSS, 2017). In total, it is necessary to minimize and prevent the spread of secondary damage in case of an earthquake by conducting a preliminary preparation based on the vulnerability analysis of earthquake disaster in Gyeongju.

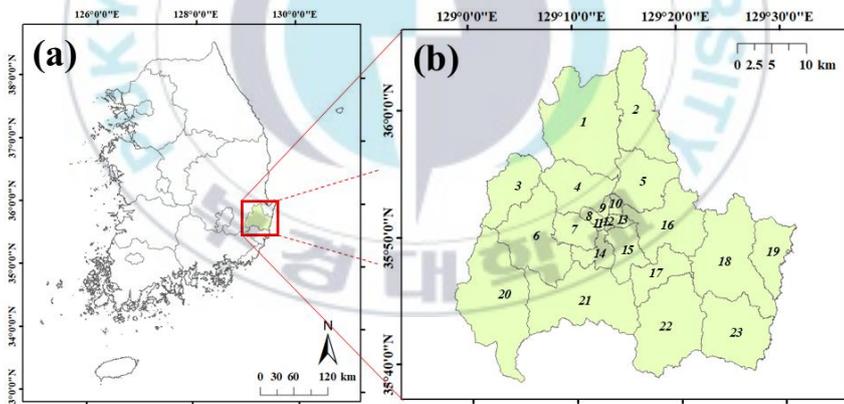


Fig. 2. The study area (a) Gyeongju, South Korea; (b) Administrative districts within Gyeongju (1: Angang-eup, 2: Gangdong-myeon, 3: Seo-myeon, 4: Hyungok-myeon, 5: Cheonbuk-myeon, 6: Geoncheon-eup, 7: Seondo-dong, 8: Seonggun-dong, 9: Hwangseong-dong, 10: Yonggang-dong, 11: Jungbu-dong, 12: Hwangoh-dong, 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, 23: Yangnam-myeon).

2.2. 9.12 Gyeongju Earthquake inventory

The dependent variable used in this study comprised a dataset of the 3,896 buildings damaged by the 9.12 Gyeongju Earthquake. These buildings were converted to 9,847 cells at a spatial resolution of 10 m; 70% of the data (6,893) was used as a training dataset to create the model, and 30% (2,954) was used to test model accuracy. The data were randomly sampled; we used the same numbers of undamaged buildings (Fig. 3).

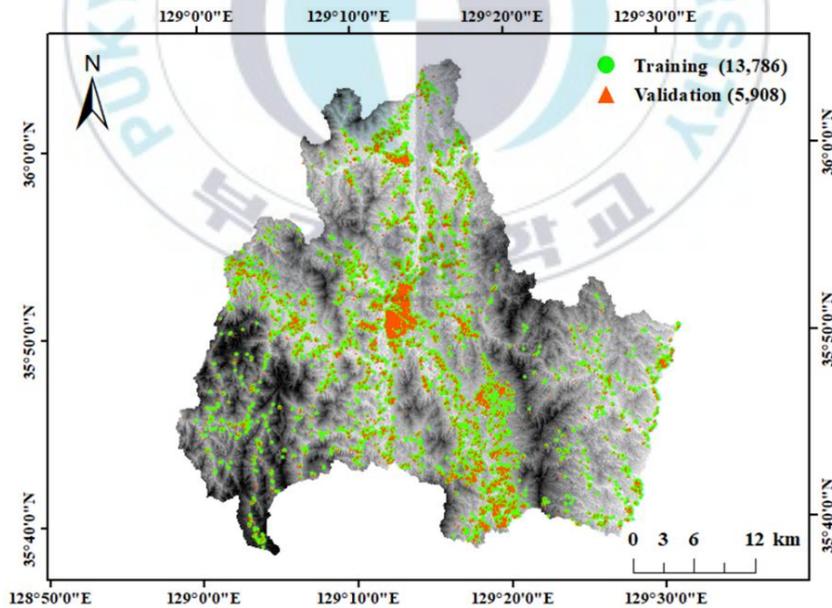
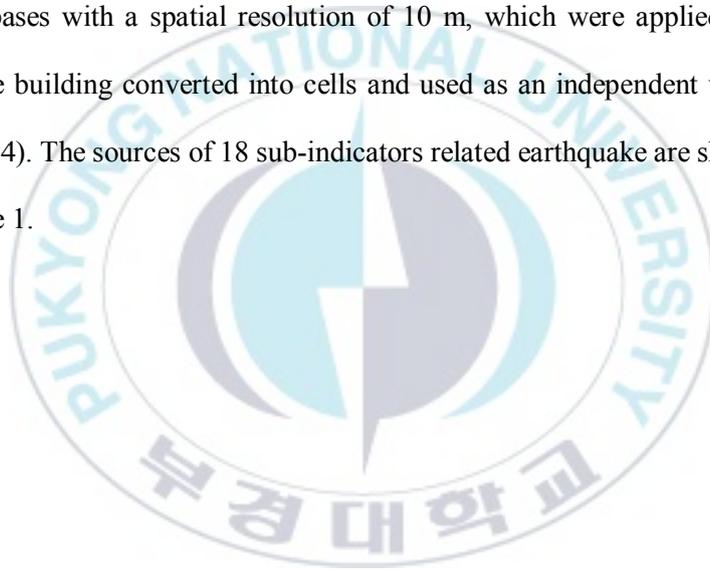
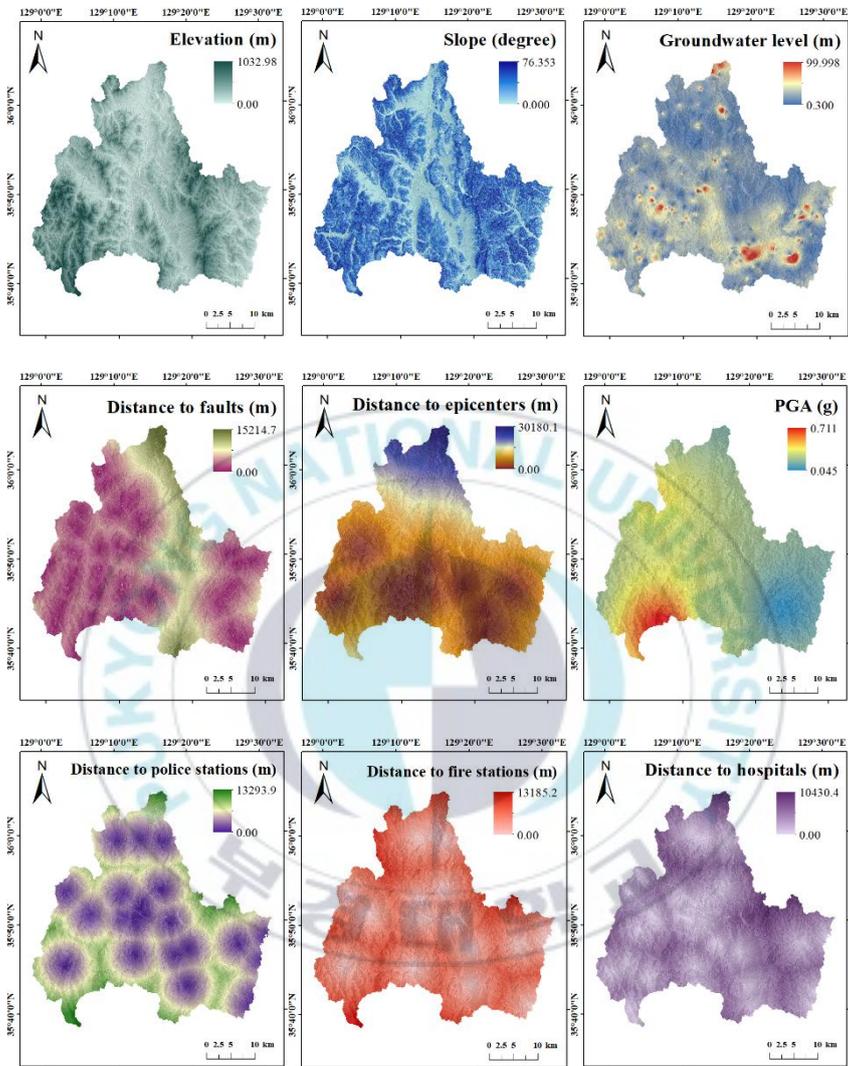


Fig. 3. Training and validation datasets.

2.3. Spatial database preparation

It was considered that various factors affecting earthquake to assess seismic vulnerability. First, five main-indicators were set up, such as geotechnical, physical, structural, social, and capacity factors. Next, 18 sub-indicators were selected, all of which were built as raster-type spatial databases with a spatial resolution of 10 m, which were applied to the entire building converted into cells and used as an independent variable (Fig. 4). The sources of 18 sub-indicators related earthquake are shown in Table 1.





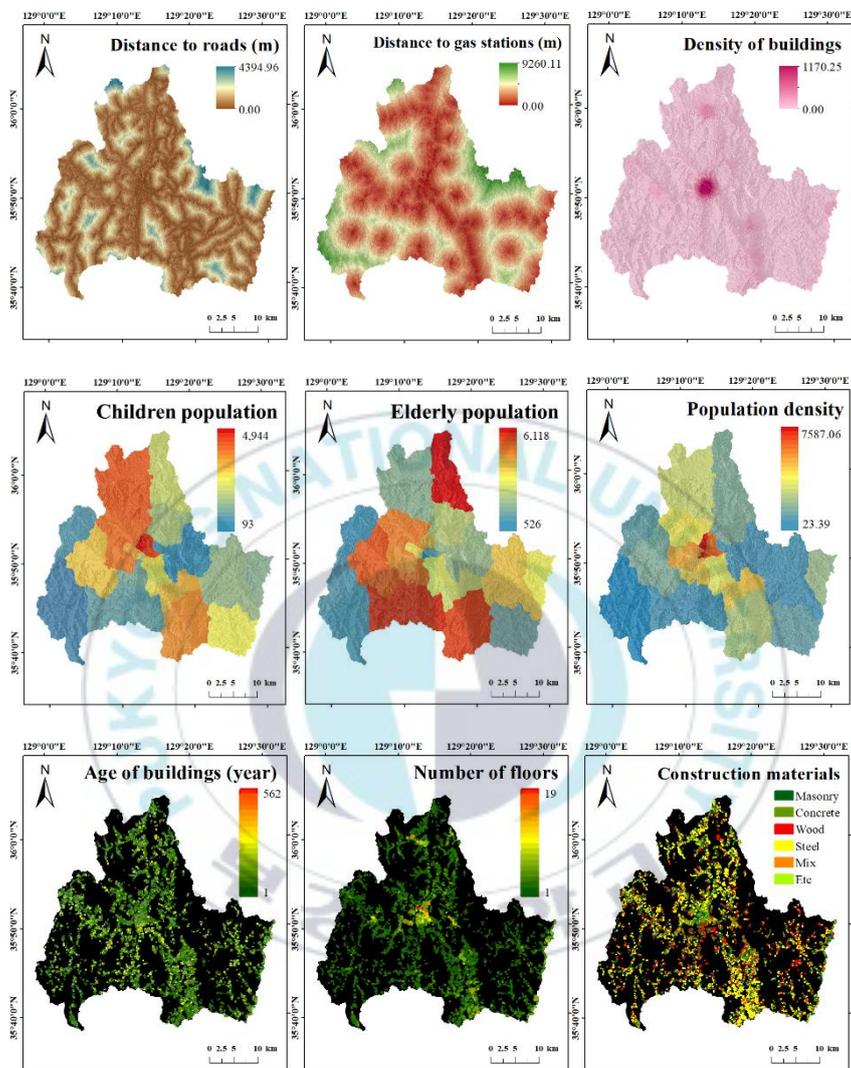


Fig. 4. Sub-indicators related earthquake.

Table 1. Independent variables for the seismic vulnerability assessment

Main-indicators	Sub-indicators	Source
Geotechnical	Elevation	NGII ¹⁾
	Slope	
	Groundwater level	NGIC ²⁾
Physical	Distance to fault	KMA ³⁾
	Distance to epicenter	KIGAM ⁴⁾
	PGA	Kim et al. (2016a, b)
Structural	Age of buildings	NSCI ⁵⁾
	Number of floors	
	Construction materials	
	Density of buildings	
Social	Child population	KOSIS ⁶⁾
	Elderly population	
	Population density	
Capacity	Distance to police stations	NGII
	Distance to fire stations	
	Distance to hospitals	
	Distance to roads	
	Distance to gas stations	ITS ⁷⁾

¹⁾ National Geographic Information Institute

²⁾ National Groundwater Information Center

³⁾ Korea Meteorological Administration

⁴⁾ Korea Institute of Geoscience and Mineral Resources

⁵⁾ National Spatial Data Infrastructure Portal

⁶⁾ Korea Statistical Information Service

⁷⁾ Intelligent Transport System Standard Node Link

2.3.1. Geotechnical indicator

Geotechnical indicators are the most influential factors that increase urban vulnerability in the event of an earthquake, which must be considered in the seismic vulnerability assessment. We considered three indicators (slope, elevation, and groundwater level), among which the slope and elevation are factors that could increase the probability of falling structures, and ground collapse of rocks along steep slopes and high altitudes, resulting in additional disasters (Moustafa et al., 2016). Groundwater level is an element that influences seismic response in the event of a large scale earthquake. Groundwater level data were collected by location of tubular well, which was interpolated across the entire region in Gyeongju (Thaker et al., 2018).

2.3.2. Physical indicator

The epicenter, or the point on the direct uppermost surface of the epicenter, is the most important indicator related to earthquake occurrence; 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. Next, the peak ground acceleration (PGA) is the degree of the ground shakes at the surface (MPSS, 2017), which is typically associated with the activity of the fault,

and is most important in assessing seismic vulnerabilities (Rezaie and Panahi, 2015). In this study, raw data measured at each National Weather Services observatory in South Korea were converted to acceleration data and interpolated throughout Gyeongju (Kim et al., 2016a; 2016b). Finally, the distance from each fault was also reflected in the assessment, as the degree of damage may vary depending on the structure type of the fault plane.

2.3.3. Structural indicator

The 9.12 Gyeongju Earthquake caused 5,368 property damage and later highlighted the importance of earthquake-resistant design of buildings. Since Korea introduced earthquake-resistant designs in 1988, only buildings with three or more stories or more are required to be designed to withstand earthquakes, and as of November 2016, only 29.9 percent of buildings in Seoul and 23.7 percent of non-residential buildings are earthquake-resistant (Kang and Kim, 2017). 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 structural vulnerability, we identified four structural indicators of seismic vulnerability: age of buildings, number of floors, construction materials, and density of buildings. Construction materials

included masonry, wood, concrete, steel, and a mixture of concrete and steel.

2.3.4. Social indicator

Social vulnerabilities affect people's ability to prepare for disasters in a given environment and to rebuild after a disaster (Chen et al., 2013). Based on social factors, the population structure of the city can be identified and directly assist in the identification of casualties and relief activities after the earthquake. Therefore, in the present study, three indicators were selected related to the population affecting social vulnerabilities.

First, the relatively weak age group of children under the age of 15 and the number of elderly people over 65 were used as factors for evaluation. The population density of Gyeongju City was also considered, judging that there is a higher probability of further damage in densely populated areas.

2.3.5. Capacity indicator

Since disaster accommodation facilities in urban are not only irregularly distributed, not all people have equivalent access, which is difficult to predict the scale of damage that may be caused by a disaster that results in considerable economic losses. Therefore, the approach is to be identified

based on the location of the infrastructure that can help in the event of an earthquake and the hazardous facility that can cause significant damage. In this study, the degree of accessibility in the event of a disaster was assessed by considering the physical distance of a total of five factors, including infrastructure, hospitals, fire stations, police stations, road networks and gas stations.



3. Methodology

3.1. Frequency ratio

The frequency ratio (FR) model is a probabilistic model, which analyzes the impact of each factor by class through correlative analysis between seismic vulnerabilities and earthquake-related factors. The FR can easily calculate which class of each impact factor had the greatest number of events when the event occurred, thereby quantitatively calculating the influence of the factors by class in the area where the event occurred (Lee and Kang, 2012).

FR values are greater than 1, it can be considered that the seismic vulnerability is highly correlated with the corresponding factors, and if it is less than 1 it can be estimated to have a lower correlation. The FR can be calculated in the following equation (Son, 2017).

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

TGFC: Training Grid of Factor Class

WTG: Whole Training Grid

FC: Factor Class Grid

WG: Whole Grid

The 'WTG' is the number of cells corresponding to the affected building, 'TGFC' refers to the number of cells corresponding to the affected building in the class, 'WG' refers to the number of cells corresponding to the entire building, and 'FC' refers to the number of cells corresponding to the building in the class.

Apply the frequency ratio values calculated by the grade of 18 indicators to the grid format of each factor, and then overlay them all to prepare the final seismic vulnerability map.

3.2. Logistic regression

The logistic regression (LR) model which was developed by McFadden (1973), is a multivariate regression analysis model that describes the relationship between a bivariate dependent variable and several independent variables through estimation of an optimal model. The LR model estimates the best model to define the relationship between dependencies and independent variables. The addition of a link function suitable for a general linear regression model allows the parameter type to be continuous, discrete, or mixed, which has the advantage of not necessarily having a normal distribution. (Colkesen et al., 2016; Lee, 2005).

The LR model based on a general linear model can be derived from the following equation:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n \quad (2)$$

$$P = \frac{e^y}{1 + e^y} \quad (3)$$

where y is the linear logistic model, b_0 is y-intercept of the model, b_n is the weight of each factor, n is the number of seismic-controlling factors, x is the earthquake conditioning factor, and P indicates the probability of damage (ranging from 0 to 1) in the event of an earthquake (Wang et al., 2016).

3.3. Support vector machine

Support vector machine (SVM) is a supervised-machine-learning method performed to solve complex classification and regression problems, based on statistical learning theory and the principle of minimizing structural risks (Vapnik, 1995; Vapnik, 1998).

SVM focuses on finding the optimum hyperplane, which is divided into two classes, using training dataset. The hyperplane with maximum margin between two classes is called optimal hyperplane; the closest point to it is called a support vector (Feizizadeh et al., 2017).

Vapnik (1995) 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). The formula for each kernel is as follows.

$$\begin{aligned}
 \text{Linear: } K(x_i, y_i) &= X_i^T X_j, \\
 \text{Polynomial: } K(x_i, y_i) &= (\gamma X_i^T X_j + r)^d, \gamma > 0, \\
 \text{Radial basis function: } K(x_i, y_i) &= e^{-\gamma(x_i - x_j)^2}, \gamma > 0, \\
 \text{Sigmoid: } K(x_i, y_i) &= \tanh(\gamma X_i^T X_j + r) \quad (4)
 \end{aligned}$$

Cost (C) is the reciprocal of the normalization parameter λ , which is a common parameter applied to all functions. For each controlling factor, higher C values correspond to less influence. The γ term controls the width of the Gaussian kernel, and is present in all functions except the linear function, 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.

3.4. Random forest

Random forest (RF) is a powerful ensemble technique proposed by Breiman (2001), which can be applied to classification, regression and non-supervised learning, showing good performance in various researches (Liaw and Wiener, 2002).

The RF method generates a random binary tree based on a bootstrap sample, using a random subset of variables selected on each node through the classification and regression trees (CART) procedure. In this study, a regression algorithm for RF was used. This produces an estimate of the dependent variable based on the average of the results. The RF model is an appropriate way to analyze the hierarchical interactions and nonlinearities of large data sets, since it does not require assumptions about the relationship between the description and response variables (Kim et al., 2018).

To generate the RF model, two parameters should be defined: number of trees (ntree) and number of variables (mtry) to be used on each node. Although increasing the value of ntree does not necessarily increase the

accuracy of the model, it is necessary to test the ntrees in a number of cases and set them high enough to collect errors. (Taalab et al., 2018).



4. Results

4.1. Validation of model

The seven models produced based on the four methodologies were verified using the relative operating characteristic (ROC) curve methods. The ROC curve evaluates the accuracy of the model by calculating the area under the receiver operator mechanical curve (AUC) values, which the AUC's value is accuracy, and the closer it is to 1, the better the model is (Pradhan, 2013).

Through this process, the accuracy of the model (the success rate), and the predictive accuracy rate were calculated. The y-axis of the ROC curve graph represents the positive rate correctly classified as sensitivities, and the x-axis, 1-Specificity, means the correct classified negative rate (Fig. 5, 6).

We then used IBM SPSS Statistics software (version 25) to create models for FR and LR models, and SVM and RF models created using RStudio software (version 3.6.0). All of models verified using SPSS software.

The success rate of FR was 0.661 and the prediction rate was 0.655, and the success rate of LR was 0.634 and the prediction rate was 0.648. Next,

the SVM adjusted the C and gamma values for all kernel models, while the degree and r applied default values. The optimal parameters derived for each model are as shown in Table 2.

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 0.633, 0.862, 0.980, and 0.617, respectively. The prediction rates using test dataset for the LN-SVM, PL-SVM, RBF-SVM, and SIG-SVM models were 0.642, 0.825, 0.890, and 0.621, respectively.

Next, the RF model also created an optimal model by adjusting the ntree and mtry values based on the training database. The highest accuracy was shown when the ntree was 8000 and the mtry was 6 and the success rate was 1.000 and the prediction rate was 0.949 for the highest performance.

Table 2. Optimal parameters of kernels

Model	C	gamma
LN-SVM	2^{-10}	-
PL-SVM	0.5	1
RBF-SVM	10	1
SIG-SVM	0.125	2^{-10}

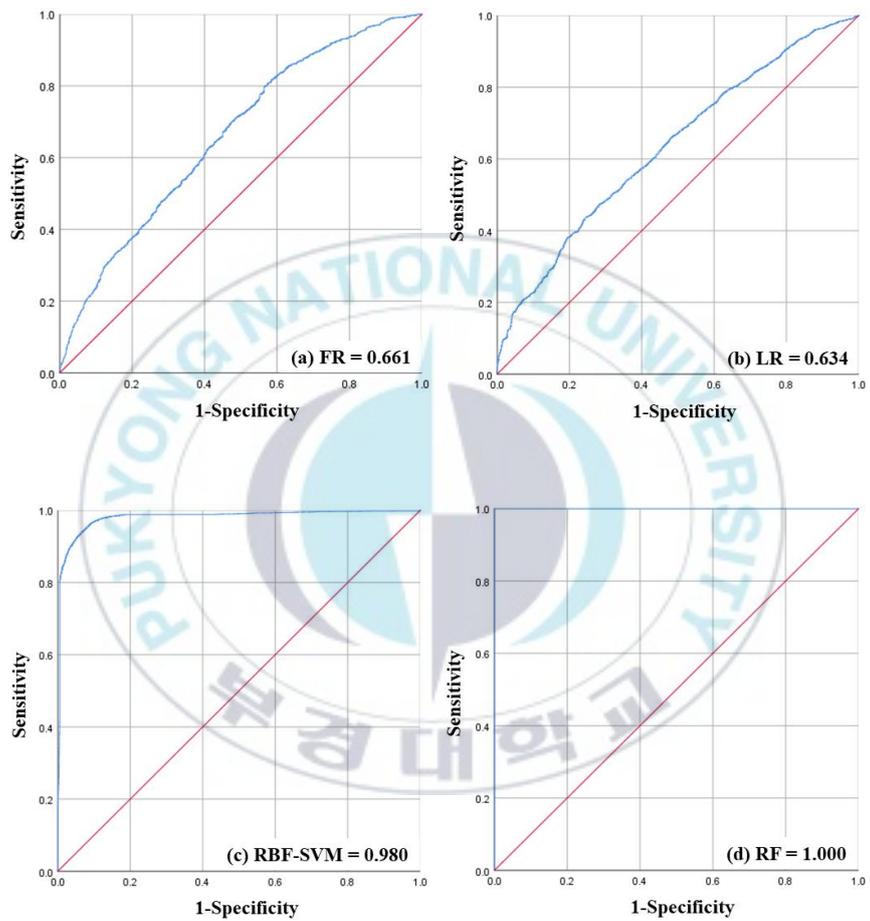


Fig. 5. ROC curve (Success rate)

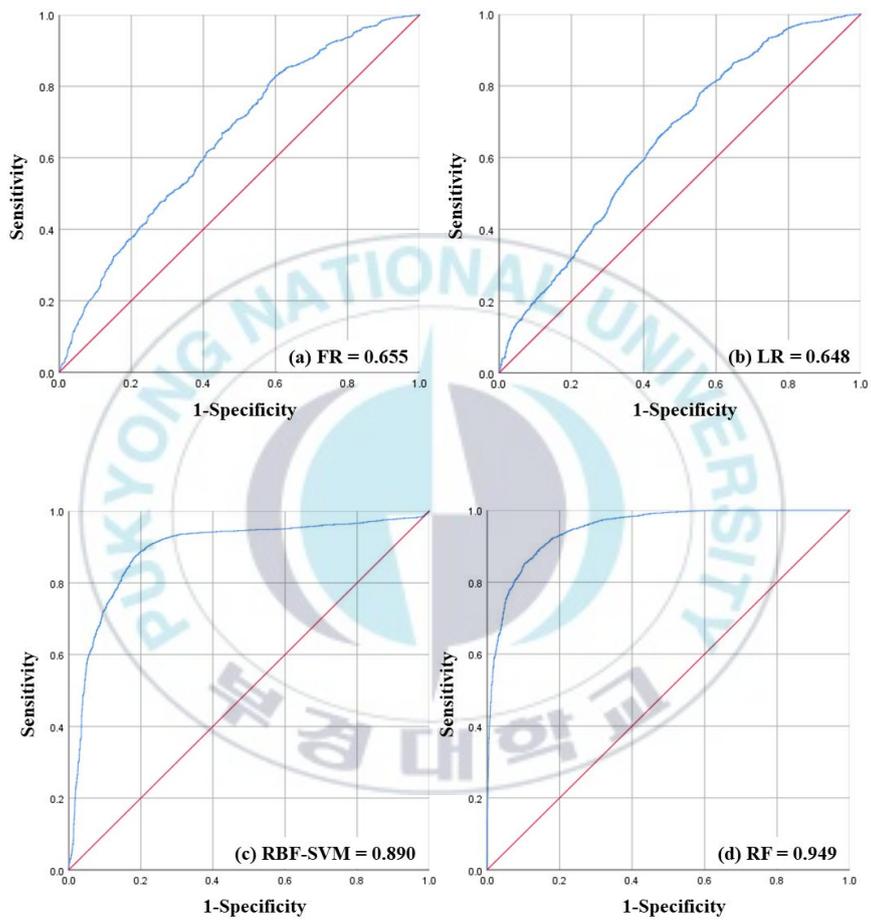


Fig. 6. ROC curve (Prediction rate)

4.2. Importance factor

In order to analyze using the frequency ratio (FR) model, 18 factors were divided into six classes using natural breaks method. Sub-indicators classified into six classes calculated FR values based on the number of pixels corresponding to the general building in each section and the number of pixels corresponding to the damaged building (Table 3).

First, in sub-indicators of geotechnical factor, the ratio of elevation was 1.23 at 86.061-138.262 (m), the ratio of slope was 1.11 at 1.716-4.291 (degree), and the ratio of groundwater level was 1.47 at 21.047-37.061 (m). Next, In the case of relationship between seismic vulnerability and physical factor, the ratio was 1.26 for distance to fault in the range 6.124-7.946 (km), the ratio of distance to epicenter was 1.24 at 0.028-3.183 (km), and the ratio of PGA was 1.24 at 0.244-0.288(g). The relationships between seismic vulnerability and structural factor, the ratio was 1.38 for age of buildings in the range 33-59 (years), the ratio was 1.53 for building density in the range 949.480-1,169.540, the ratio was highest (1.53) when number of floors was 5-7 (floors), and the ratio of construction materials was mixed concrete and steel type (1.44). The relationships between seismic vulnerability and social factor, the FR value was 1.21 for people under 15 years old at 1,020-1,279 (people), 1.94 for people over 65 at 526

(people), and 1.76 for population density at 201.359-586.957. Finally, the distance from the police stations of capacity factor was 1.18 in the range 0.000-1.205 (km), distance to fire stations was 1.16 in the range 0.000-1.43 (km). The ratio of distance to hospitals (4.216-5.646 (km)) and gas station (0.000-0.680 (km)) was 1.16, and distance of roads was 1.05 at 0.000-0.116 (km).

Next, LR can check the importance of the factor based on the significance coefficient values (Table 4). Significance coefficient below 0.05 can be considered to have less impact on seismic vulnerabilities.

Factors with a significance coefficient over 0.05 are elevation, slope, age of buildings, children population, elderly population, population density, and distance to roads, which are 0.458, 0.362, 0.568, 0.508, 0.924, 0.359, and 0.080. The factors were found to be inadequate for the analysis of seismic vulnerabilities. Also, construction materials are deemed to be inadequate as the p values are over 0.05 in all categories.

Table 3. Frequency ratio of each factors

	Class	No. of pixel in building	Building (%)	No. of pixel in damage building	Damage building (%)	Frequency ratio
Elevation (m)	1545-46289	40,284	43.96	4,221	42.87	0.98
	46289-86061	24,840	27.11	2,399	24.36	0.90
	86061-138262	17,421	19.01	2,308	23.44	1.23
	138262-220292	6,468	7.06	746	7.58	1.07
	220292-366952	1,787	1.95	164	1.67	0.85
	366952-635414	842	0.92	9	0.09	0.10
Slope (degree)	0-1716	47,128	51.43	5,278	53.60	1.04
	1716-4291	23,371	25.50	2,778	28.21	1.11
	4291-7725	13,189	14.39	1,264	12.4	0.89
	7725-12016	5,533	6.04	380	3.86	0.64
	12016-18597	1,996	2.18	113	1.15	0.53
	18597-72959	425	0.46	34	0.35	0.74
Groundwater level (m)	0346-7377	30,754	33.56	2,399	24.36	0.73
	7377-12845	39,133	42.70	4,469	45.38	1.06
	12845-21047	15,209	16.60	2,080	21.12	1.27
	21047-37061	5,153	5.62	813	8.26	1.47
	37061-83346	1,075	1.17	73	0.74	0.63
	83346-99947	318	0.35	13	0.13	0.38
Distance of faults (km)	0-1973	25,199	27.50	2,825	28.69	1.04
	1973-3947	29,228	31.89	3,021	30.68	0.96
	3947-6124	15,147	16.53	1,376	13.97	0.85
	6124-7946	10,904	11.90	1,479	15.01	1.26
	7946-9768	6,947	7.58	758	7.70	1.02

	9.768–12.906	4,217	4.60	389	3.95	0.86
Distance of epicenters (km)	0.028–3.183	17,765	19.39	2,368	24.05	1.24
	3.183–6.112	35,244	38.46	4,529	45.99	1.20
	6.112–10.731	21,506	23.47	1,931	19.61	0.84
	10.731–16.590	5,767	6.29	686	6.97	1.11
	16.590–21.886	9,184	10.02	326	3.31	0.33
	21.886–28.758	2,176	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	2,985	30.31	1.16
	0.244–0.288	38,745	42.28	4,878	49.54	1.17
	0.288–0.371	14,222	15.52	1,187	12.05	0.78
	0.371–0.510	1,966	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	3,606	36.62	0.91
	18–32	34,320	37.45	3,584	36.40	0.97
	33–59	13,275	14.49	1,964	19.95	1.38
	60–98	6,243	6.81	569	5.78	0.85
	99–172	1,050	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	7,121	72.32	0.91
	3–4	12,901	14.08	1,870	18.99	1.35
	5–7	3,967	4.33	651	6.61	1.53
	8–12	1,071	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
	Masonry	17,578	19.18	1,642	16.68	0.7

Construction materials	Concrete	27,048	29.51	3,684	37.41	1.27
	Wood	11,096	12.11	1,258	12.78	1.06
	Steel	35,262	38.48	3,167	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	0476-156351	65,002	70.93	6,401	65.00	0.92
	156351-376410	10,396	11.34	914	9.28	0.82
	376410-596469	3,409	3.39	412	4.18	1.23
	596469-770682	4,205	4.59	675	6.85	1.49
	770682-949480	4,586	5.00	732	7.43	1.49
	949480-1,169,540	4,344	4.74	713	7.24	1.53
Child population (age<15)	93-183	7,735	8.44	868	8.81	1.04
	183-329	15,254	16.65	1,823	18.51	1.11
	329-603	14,290	15.59	1,431	14.53	0.93
	603-1,020	8,821	9.63	944	9.59	1.00
	1,020-1,279	20,238	22.08	2,628	26.69	1.21
	1,279-4,944	25,304	27.61	2,153	21.86	0.79
Elderly population (age>65)	526	2,414	2.63	503	5.11	1.94
	526-1,553	23,499	25.64	1,441	14.63	0.57
	1,553-2,032	23,470	25.61	3,026	30.73	1.20
	2,032-2,432	10,406	11.36	1,354	13.75	1.21
	2,432-3,951	27,009	29.47	3,410	34.63	1.17
	3,951-6,118	4,844	5.29	113	1.15	0.22
Population density	23,390-82,713	16,580	18.09	1,414	14.36	0.79
	82,713-201,358	41,645	45.44	3,256	33.07	0.73
	201,358-586,957	11,329	12.36	2,138	21.71	1.76

	586957-2608934	2,674	2.92	40	4.09	1.40
	260893-559974	10,414	11.36	1,065	10.82	0.95
	559974-758706	9,000	9.82	1,571	15.95	1.62
Distanceto police stations (km)	0-1205	36,638	39.98	4,626	46.98	1.18
	1205-2458	19,864	21.68	2,006	20.37	0.94
	2458-3807	17,359	18.96	1,468	14.91	0.79
	3807-5350	10,379	11.33	1,201	12.20	1.08
	5350-8145	6,883	7.51	540	5.48	0.73
	8145-12291	519	0.57	6	0.06	0.11
Distanceto fire stations (km)	0-1431	34,930	38.12	4,363	44.31	1.16
	1431-2766	22,245	24.27	2,233	22.68	0.93
	2766-4102	14,318	15.62	1,357	13.78	0.88
	4102-5533	12,231	13.35	1,212	12.31	0.92
	5533-8204	7,425	8.10	670	6.80	0.84
	8204-12164	493	0.54	12	0.12	0.23
Distanceto hospitals (km)	0-0828	35,977	39.26	4,115	41.79	1.06
	0828-1919	20,364	22.22	1,976	20.07	0.90
	1919-3011	15,086	16.46	1,637	16.62	1.01
	3011-4216	10,804	11.79	1,044	10.60	0.90
	4216-5646	6,744	7.36	840	8.53	1.16
	5646-9599	2,667	2.91	235	2.39	0.82
Distanceto roads (km)	0-0116	54,351	59.31	6,110	62.05	1.05
	0116-0311	22,932	25.02	2,371	24.08	0.96
	0311-0610	9,430	10.29	993	10.08	0.98
	0610-1025	2,706	2.95	258	2.62	0.89
	1025-1609	1,674	1.83	103	1.05	0.57

	1609-3310	549	0.60	12	0.12	0.20
Distance to gas stations (km)	0-0680	47,099	51.39	5,860	59.51	1.16
	0680-1391	19,988	21.81	2,006	20.37	0.93
	1391-2195	13,483	14.71	1,158	11.76	0.80
	2195-3091	6,706	7.32	571	5.80	0.79
	3091-4390	3,363	3.67	216	2.19	0.60
	4390-7884	1,003	1.09	36	0.37	0.33



Table 4. Logistic regression between seismic vulnerability and related indicators

Sub-indicators	Logistic regression coefficient	Significance coefficient
Elevation	0.0003	0.458
Slope	-0.0005	0.362
Groundwater level	-0.0088	0.001
Distance to faults	0.0000	0.004
Distance to epicenters	0.0001	0.000
PGA	0.0087	0.001
Age of buildings	-0.0005	0.568
Number of floors	-0.0292	0.004
Construction materials		
Materials1 (masonry)	0.7526	0.608
Materials2 (concrete)	0.5296	0.718
Materials3 (wood)	0.7093	0.629
Materials4 (steel)	0.8158	0.578
Materials5 (concrete + steel)	0.1529	0.918
Density of buildings	-0.0004	0.000
Child population	0.0000	0.508
Elderly population	0.0000	0.924
Population density	0.0000	0.359
Distance to police stations	0.0001	0.000
Distance to fire stations	-0.0001	0.001
Distance to hospitals	-0.0002	0.000
Distance to roads	0.0001	0.080
Distance to gas stations	0.0002	0.000
Constant	-0.7536	

The RF models identify the importance of predictors through %IncMSE and IncNodePurity during regression tree analysis.

% increase in mean squared error (%IncMSE) means an increase in the percentage of mean square error (MSE), which results in the greatest increase in error value when the most important variable in the model is removed. Increase in node purity (IncNodePurity) is a measure of the reduction in the Gini coefficient and includes the reduction of the model's agreement on the square of residuals. The Gini coefficient is a homogeneity measure of the three nodes of an RF model, meaning that the higher the value, the more important the variable is (Paul et al., 2019).

Table 5 shows that Distance to epicenters is the highest value in terms of %IncMSE, and is an important variable in the order of Distance to fire stations, PGA. In the order of Distance to epicenters, PGA, and Elevation, it was found that the most important criteria were the IncNodePurity. In common, Distance to epicenters and PGA have been identified as the factors that have the greatest impact on seismic vulnerabilities.

Table 5. Importance variables of random forest

Sub-indicators	%IncMSE	IncNodePurity
Elevation	284.749	253.147
Slope	267.526	158.825
Groundwater level	232.894	244.861
Distance to faults	286.345	229.250
Distance to epicenters	332.423	288.229
PGA	315.220	271.325
Age of buildings	298.413	223.064
Number of floors	165.124	95.954
Construction materials		
Materials1 (masonry)	122.533	21.600
Materials2 (concrete)	69.515	21.038
Materials3 (wood)	87.179	13.428
Materials4 (steel)	119.914	46.754
Materials5 (concrete + steel)	39.538	2.169
Materials6 (Etc)	1.000	0.050
Density of buildings	289.661	202.302
Child population	125.958	62.316
Elderly population	113.949	81.016
Population density	170.006	114.353
Distance to police stations	310.752	201.540
Distance to fire stations	324.077	207.242
Distance to hospitals	290.231	204.810
Distance to roads	286.129	156.594
Distance to gas stations	306.848	199.011

4.3. Seismic vulnerability map

The seismic vulnerability maps were constructed by applying the best model of each methodology to a total of four maps. The SVM applied the RBF kernel-based model with the highest accuracy.

The FR map was applied to each sub-indicator divided into six classes, and 18 sub-indicators with FR values were overlaid to produce the final seismic vulnerability map.

$$\begin{aligned} \text{LR map} = & -0.7536 + (0.0003 \times \text{Elevation}) + (-0.0005 \times \text{Slope}) \\ & + (-0.0088 \times \text{Groundwater level}) \\ & + (0.0000 \times \text{Distance to faults}) \\ & + (0.0001 \times \text{Distance to epicenters}) \\ & + (0.0087 \times \text{PGA}) + (-0.0005 \times \text{Age of buildings}) \\ & + (-0.0292 \times \text{Number of floors}) \\ & + (0.7526 \times \text{Materials1}) + (0.5296 \times \text{Materials2}) \\ & + (0.7093 \times \text{Materials3}) + (0.8158 \times \text{Materials4}) \\ & + (0.1529 \times \text{Materials5}) \\ & + (-0.0004 \times \text{Density of buildings}) \\ & + (0.0000 \times \text{Child population}) \\ & + (0.0000 \times \text{Elderly population}) \\ & + (0.0000 \times \text{Population density}) \\ & + (0.0001 \times \text{Distance to police stations}) \\ & + (-0.0001 \times \text{Distance to fire stations}) \\ & + (-0.0002 \times \text{Distance to hospitals}) \\ & + (0.0001 \times \text{Distance to roads}) \\ & + (0.0002 \times \text{Distance to gas stations}) \end{aligned} \quad (5)$$

LR map was produced by applying the logistic coefficient. The seismic vulnerability maps based RBF-SVM and RF models are constructed by the model's predicted values, and maps are classified into five classes (safe, low, moderate, high, and 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.

We compared the percentages of buildings assigned to each grade in the vulnerability maps. In the FR map, the buildings in “safe” class among the total 71,888 buildings were 589 (0.82%), those in “low” class were 9,999 (13.91%), those in “moderate” class were 36,172 (50.32%), those in “high” class were 21,299 (29.63%), and those in “very high” class were 3,829 (5.33%). In the LR map, the buildings in “safe” class were 3,232 (4.50%), those in “low” class were 12,254 (17.05%), those in “moderate” class were 21,868 (30.42%), those in “high” class were 28,786 (40.04%), and those in “very high” class were 5,748 (8.00%) In the RBF-SVM map, the buildings in “safe” class were 434 (0.60%), those in “low” class were 31,171 (43.36%), those in “moderate” class were 31,466 (43.77%), those in “high” class were 8,662 (12.05%), and those in “very high” class were 155 (0.22%), whereas in the RF map, the buildings in “safe” class were 23,803 (33.11%), those in “low” class were 26,429 (36.76%), those in “moderate” class were 13,669 (19.01%), those in “high” class were 6,548

(9.11%), and those in “very high” class were 1,439 (2.00%) (Table 6). The result of mapping produced by five grades is shown in Fig. 7.



Table 6. Distribution of building by five classes

Model		Safe	Low	Moderate	High	Very high	Sum
FR	Number of buildings	589	9,999	36,172	21,299	3,829	71,888
	Ratio	0.82	13.91	50.32	29.63	5.33	100.00
LR	Number of buildings	3,232	12,254	21,868	28,786	5,748	71,888
	Ratio	4.50	17.05	30.42	40.04	8.00	100.00
SVM (RBF)	Number of buildings	432	31,171	31,466	8,662	155	71,888
	Ratio	0.60	43.36	43.77	12.05	0.22	100.00
RF	Number of buildings	23,803	26,429	13,669	6,548	1,439	71,888
	Ratio	33.11	36.76	19.01	9.11	2.00	100.00

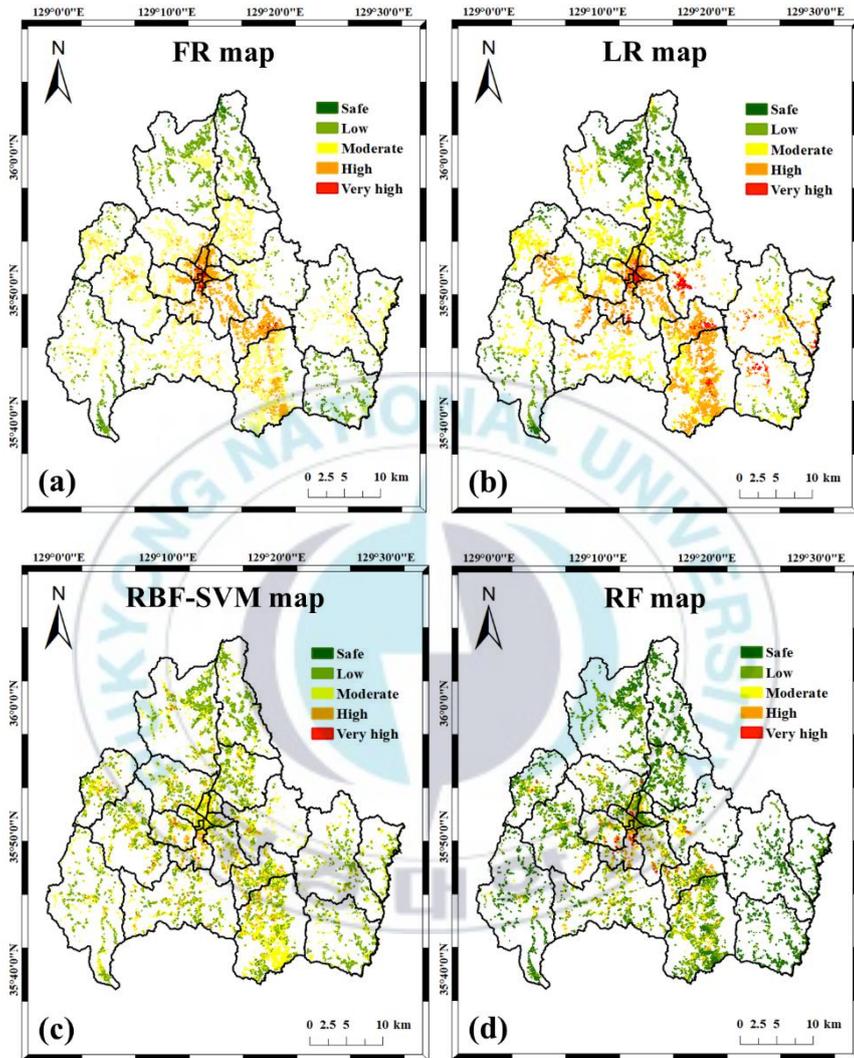


Fig. 7. Seismic vulnerability maps (a) FR map; (b) LR map; (c) RBF-SVM map; (d) RF map.

4.4. Building class distribution by Gyeongju districts

We examined the seismic vulnerability distribution of building located with the 23 administrative districts of Gyeongju. The distribution of buildings by grade within the administrative region is as shown in Fig. 8.

Based on the FR map, the regions most vulnerable to earthquake (i.e., high and very high classes) were Jungbu-dong, Hwangoh-dong, Hwangseong-dong, Seonggun-dong, and Wolseong-dong; based on the LR map, they were Jungbu-dong, Hwangoh-dong, Dongcheon-dong, Hwangnam-dong, and Seonggun-dong; based on the RBD-SVM map, they were Hwangnam-dong, Seondo-dong, Wolseong-dong, Geoncheon-eup, Bulguk-dong; based on the RF map, Hwangnam-dong, Seondo-dong, Wolseong-dong, Bulguk-dong, and Hwangoh-dong.

The safest regions (i.e., safe and low classes) based on the FR map were Gangdong-myeon, Yangnam-myeon, Angang-eup, Sannae-myeon, and Yangbuk-myeon; based on the LR map, they were Angang-eup, Gangdong-myeon, Sannae-myeon, Gampo-eup, and Hyungok-myeon.; based on the RBD-SVM map, they were Gangdong-myeon, Yangnam-myeon, Angang-eup, Yangbuk-myeon, and Cheonbuk-myeon; based on the RF map, Gangdong-myeon, Yangbuk-myeon, Gampo-eup, Gampo-eup, and Cheonbuk-myeon.

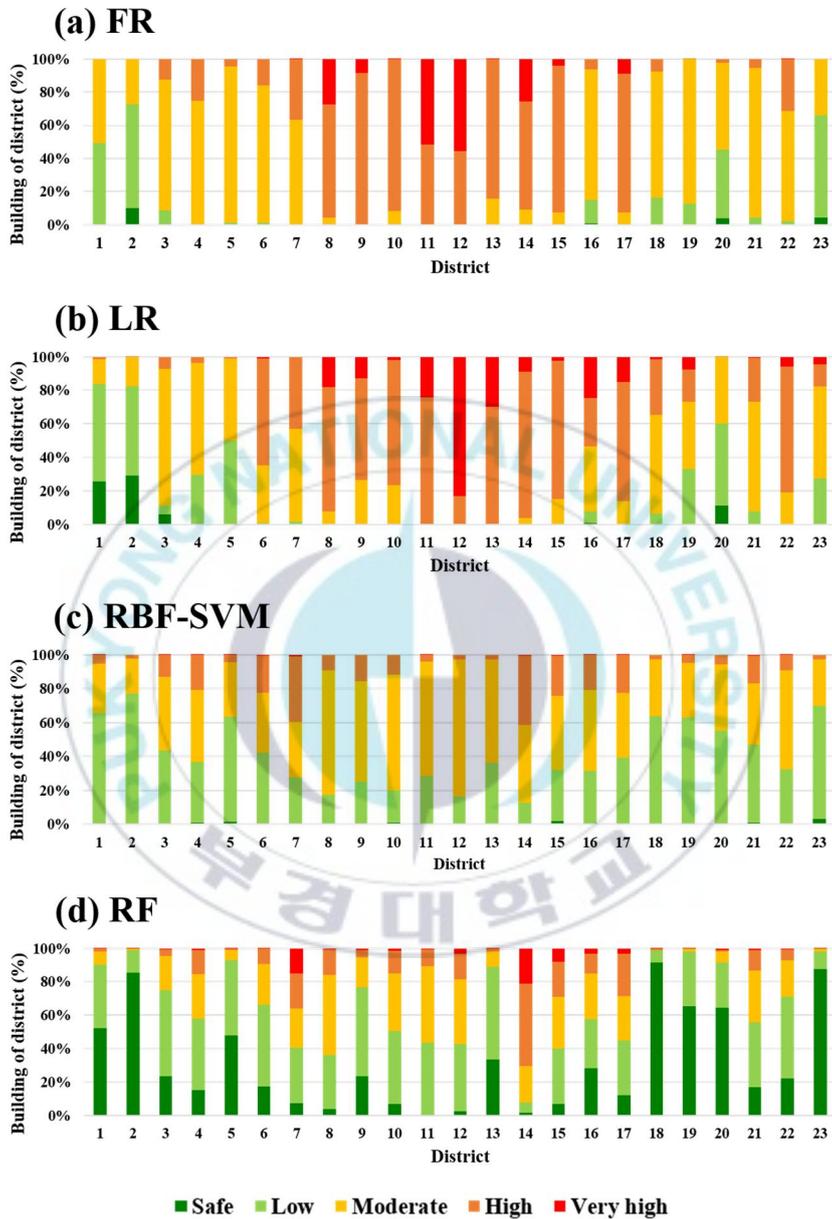


Fig. 8. Ratio of vulnerable buildings of Gyeongju districts by each model.

5. Discussion

We compared the model and prediction accuracy to see the differences in function. The LN-SVM, SIG-SVM, FR, and LR models were very similar in terms of success and prediction rates, as having slightly higher predictive with the exception of FR. The negligible difference in accuracy between the training and verification datasets may indicate underfitting problems for these two models, such that they are too simplistic to extract data diversity. Therefore, the LN-SVM, SIG-SVM, FR, and LR models may be inappropriate for predicting seismic vulnerability. The PL-SVM, RBF-SVM, and RF models showed high accuracy with 86.2% and 98.0%, and 100.0%, with a predication rate of 82.5%, 89.0%, and 94.9%, respectively. The PL-SVM and RBF-SVM models as nonlinear SVM kernels, and would therefore be useful for creating complex decision-making borders, even with small numbers of features, and are advantageous in that they operate smoothly for various datasets. These two models showed high accuracy using the training and validation datasets, and are therefore considered reliable. For the RF model, it complements the disadvantages of a single tree and works well with large datasets. Therefore, based on the relatively large data set used in this study, a model with high performance appears to have been created, which is

considered most suitable for predicting buildings vulnerable to earthquakes.

Finally, the RF model appeared as the best performance model, and the RBF-SVM model showed the next best performance. The SIG-SVM model has been identified as the lowest predictive efficiency model, indicating that it is most unsuitable for assessing seismic vulnerability.

The results of the present study are consistent with those of previous studies. We looked at studies comparing the accuracy of the SVM's kernel-specific models. Xu and Xu (2012) compared SVM's Kernel-specific performance as a study to generate a spatial prediction model for earthquake-causing landslides. As a result, the RBF (0.8434) kernel function showed the highest model accuracy, and the higher accuracy was calculated in the order of the PL (0.8373), LN (0.8009), and SIG (0.652) kernels. Xu et al. (2016) produced the ANN model and SVM model by kernel, the RBF (0.821) and PL (0.881) kernels showed similar levels of accuracy, followed by models of higher accuracy in the order ANN (0.8641), LN (0.7952), and SIG (0.5019). Feizizadeh et al. (2017) assessed the accuracy of each kernel of the SVM for landslide usability mapping, the RBF (0.893) kernel function was evaluated as the most suitable model for assessing the landscape usability, and the SIG (0.828) function showed the lowest accuracy. We found that the SVM model based on the RBF

kernel showed the most predictive accuracy, based on the SVM's Kernel-specific model performance

We looked at studies comparing the performance of different methodological models. Merghadi et al. (2018) compared their predictive abilities using several machine learning methods for landslide susceptibility mapping. Based on each methodology of RF, SVM, LR, gradient boosting machine (GBM), and artificial neural network (NNET), it was found that the optimal model was created with high predictability in the order of GBM, RF, NNET, SVM, and LR models. RF (0.8957) and SVM (0.8818) and LR (0.8575) were both shown to be over mid-80% accuracy. Kavzoglu et al. (2019) used the machine learning method Bagging, Rotation Forest (RotFor), RF, SVM and statistical-based LR models to select models suitable for landslide usability mapping. The results yielded the predicted accuracy of RF (0.963), Rotfor (0.959), SVM (0.955), Bagging (0.931), and LR (0.868), and showed high performance in tree-lined methodologies during machine learning. The results yielded the predicted accuracy of RF (0.963), Rotfor (0.959), SVM (0.955), Bagging (0.931), and LR (0.868), and showed high performance in methodologies of tree type during machine learning. Garosi et al. (2019) generated and analyzed the accuracy of machine learning methodology RF, SVM, generalized additive model (GAM), and naïve bayes (NB)

models to predict the usability of gully erosion. As a result, high accuracy models were created in the order RF (0.9236), SVM (0.9087), NB (0.8987), and GAM (0.8715). Xiao et al. (2019) predicted the landside usability region using the machine learning method RF and statistical-based FR certainty factor (CF), index of entropy (IOE) methodology. As a result, RF (0.801) showed the highest performance, followed by IOE (0.738), CF (0.732), and FR (0.728).

A literature review showed that machine learning-based models performed better than statistics and probability-based methodologies, among which the tree type methods were suitable for analysing vulnerabilities. And among the four kernel of the SVM, the performance of the RBF kernel was the highest, so we could see that we adopted a model that was created based on that kernel.

Next, we compared seismic vulnerability maps based on the four models. The RBF-SVM and RF models based on machine learning showed vulnerable and safe regions almost identically.

And the safety zones of the RBF-SVM and the FR have also been identified with similar trends.

Finally, we examined the distribution of vulnerability grades by administrative district. The FR and LR maps indicated that the center of the city, RBF-SVM map indicated southwestern, and RF map showed

south region was vulnerable. Whereas FR and LR maps showed that the north and outer areas of Gyeongju, RBF-SVM map indicated north and west coastal regions, and RF map showed west region were safe.



6. Conclusions

In this paper, the seismic vulnerability of Gyeongju was assessed by applying 18 earthquake sub-indicators related to geotechnical, physical, structural, social, and capacity indicators as independent variables, and buildings damaged during the 9.12 Gyeongju Earthquake as dependent variables. We generated 7 models using the FR, LR, SVM (including the four kernel of the SVM; LN, PL, RBF, and SIG), and RF methodologies. The four seismic vulnerability maps, based on the most accurate models of each methodology, constructed and a comparative analysis was performed. The model and prediction accuracy of RF were shown high performance, with 100.0% and 94.9%, respectively, followed by RBF-SVM, FR, and LR. The most vulnerable areas derived from three of the four maps as Hwangnam-dong (LR, RBF-SVM, and RF map) and Wolseong-dong (FR, RBF-SVM, and RF map) should be managed first.

In further studies, independent variables are reconstructed and analyzed according to the importance of independent variables evaluated in this study, which is judged to help predict seismic vulnerabilities more accurately. In additional, it is also considered that additional identification and analysis of factors affecting seismic vulnerabilities will help predict seismic vulnerabilities.

The seismic vulnerability maps produced in this study have the advantage of easy to identify intuitively vulnerable areas, and can provide considerable assistance in environmental, land, buildings and facility management in advance in preparation for possible future earthquakes. It can also be used as an important basic data in the policy establishment related to earthquake disasters, and it is expected that this will reduce damage caused.



References

- Ahmed, M., & Morita, H. (2018). An Analysis of Housing Structures' Earthquake Vulnerability in Two Parts of Dhaka City. *Sustainability*, 10, 1106.
- Aliabadi, S. F., Sarsangi, A., & Modiri, E. (2015). The Social and Physical Vulnerability Assessment of Old Texture Against Earthquake (Case Study: Fahadan District in Yazd City). *Arabian Journal of Geosciences*, 8, 10775-10787.
- Alizadeh, M., Alizadeh, E., Asadollahpour K. S., Shahabi, H., Beiranvand P. A., Panahi, M., Bin A. B., & Saro, L. (2018). Social Vulnerability Assessment using Artificial Neural Network (ANN) Model for Earthquake Hazard in Tabriz City, Iran. *Sustainability*, 10, 3376.
- Amiri, A., Delavar, M., Zahrai, S., & Malek, M. (2007). Tehran Seismic Vulnerability Assessment using Dempster–Shafer Theory of Evidence. In *Proc. Map Asia Conference*, Kuala Lumpur, Malaysia, August, pp. 14-16.
- Armaş, I. (2012). Multi-Criteria Vulnerability Analysis to Earthquake Hazard of Bucharest, Romania. *Nat. Hazards*, 63, 1129-1156.
- Bahadori, H., Hasheminezhad, A., & Karimi, A. (2017). Development of an Integrated Model for Seismic Vulnerability Assessment of

- Residential Buildings: Application to Mahabad City, Iran. *Journal of Building Engineering*, 12, 118-131.
- Borfecchia, F., De Cecco, L., Pollino, M., La Porta, L., Lugari, A., Martini, S., Ristoratore, E., & Pascale, C. (2010). Active and Passive Remote Sensing for Supporting the Evaluation of the Urban Seismic Vulnerability. *Italian Journal of Remote Sensing*, 42, 129-141.
- Breiman, L. (2001). Random forests. *Mach. Learn*, 45(1), 5–32.
- Chen, W., Cutter, S. L., Emrich, C. T., & Shi, P. (2013). Measuring Social Vulnerability to Natural Hazards in the Yangtze River Delta Region, China. *International Journal of Disaster Risk Science*, 4, 169-181.
- Chen, W., Sun, Z., & Han, J. (2019). Landslide Susceptibility Modeling using Integrated Ensemble Weights of Evidence with Logistic Regression and Random Forest Models. *Applied Sciences*, 9, 171.
- Colkesen, I., Sahin, E. K., & Kavzoglu, T. (2016). Susceptibility Mapping of Shallow Landslides using Kernel-Based Gaussian Process, Support Vector Machines and Logistic Regression. *J. Afr. Earth Sci*, 118, 53-64.
- Conforti, M., Pascale, S., Robustelli, G., & Sdao, F. (2014). 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.

- Feizizadeh, B., Roodposhti, M. S., Blaschke, T., & Aryal, J. (2017). Comparing GIS-Based Support Vector Machine Kernel Functions for Landslide Susceptibility Mapping. *Arabian Journal of Geosciences*, 10, 122.
- Garosi, Y., Sheklabadi, M., Conoscenti, C., Pourghasemi, H. R., & Van Oost, K. (2019). Assessing the Performance of GIS-Based Machine Learning Models with Different Accuracy Measures for Determining Susceptibility to Gully Erosion. *Sci. Total Environ*, 664, 1117-1132.
- Guettiche, A., Guéguen, P., & Mimoune, M. (2017). Seismic Vulnerability Assessment using Association Rule Learning: Application to the City of Constantine, Algeria. *Nat. Hazards*, 86, 1223-1245.
- Gyeongju city hall, "<http://www.gyeongju.go.kr/>" (accessed on 31 October 2019).
- Kadavi, P. R., Lee, C., Lee, S. (2019). Landslide-Susceptibility Mapping in Gangwon-do, South Korea, using Logistic Regression and Decision Tree Models. *Environmental earth sciences*, 78, 116.
- Kang T. S., & Kim. D. K. (2017). Convergence research review. *Convergence Research Policy Center (CRPC)*, Seoul, Korea, 3(1), 38-63.

- Kavzoglu, T., Sahin, E. K., & Colkesen, I. (2014). Landslide Susceptibility Mapping using GIS-Based Multi-Criteria Decision Analysis, Support Vector Machines, and Logistic Regression. *Landslides*, 11, 425-439.
- Kavzoglu, T., Colkesen, I., & Sahin, E. K. (2019). Machine learning techniques in landslide susceptibility mapping: a survey and a case study. In *Landslides: Theory, Practice and Modelling.*; Anonymous.; Springer, pp. 283-301.
- Khosravi, K., Nohani, E., Maroufinia, E., & Pourghasemi, H. R. (2016). 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*, 83, 947-987.
- Kim, J., Lee, S., Jung, H., & Lee, S. (2018). Landslide Susceptibility Mapping using Random Forest and Boosted Tree Models in Pyeong-Chang, Korea. *Geocarto Int*, 33, 1000-1015.
- Kim, K. H., Kang, T. S., Rhie, J., Kim, Y., Park, Y., Kang, S. Y., Han, M., Kim, J., Park, J., & Kim, M. (2016). The 12 September 2016 Gyeongju Earthquakes: 2. Temporary Seismic Network for Monitoring Aftershocks. *GEOSCIENCES JOURNAL -SEOUL-*, 20, 753-757.

- Kim, K., & Yoon, S. (2018). Assessment of Building Damage Risk by Natural Disasters in South Korea using Decision Tree Analysis. *Sustainability*, 10, 1072.
- Kim, Y., Rhie, J., Kang, T. S., Kim, K. H., Kim, M., & Lee, S. J. (2016). The 12 September 2016 Gyeongju Earthquakes: 1. Observation and Remaining Questions. *GEOSCIENCES JOURNAL -SEOUL-*, 20, 747-752.
- Kim, Y., Kim, T., Kyung, J. B., Cho, C. S., Choi, J., & Choi, C. U. (2017). Preliminary Study on Rupture Mechanism of the 9.12 Gyeongju Earthquake. *Journal of the Geological Society of Korea*, 53, 407-422.
- Kleinbaum, D. G. (1998). *Survival Analysis, a self-learning Text*. *Journal of Mathematical Methods in Biosciences*, 40, 107-108.
- Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine Learning in Geosciences and Remote Sensing. *Geoscience Frontiers*, 7, 3-10.
- Lee, K. (2010). Comments on Seismicity and Crustal Structure of the Korean Peninsula. *Geophysics and Geophysical Exploration*, 13, 256-267.
- Lee, M., & Kang, J. (2012). Predictive Flooded Area Susceptibility and Verification Using GIS and Frequency Ratio. *Journal of the Korean*

- Association of Geographic Information Studies, 15(2), 86-102 (2012).
- Lee, S. (2005). Application of Logistic Regression Model and its Validation for Landslide Susceptibility Mapping using GIS and Remote Sensing Data. *Int. J. Remote Sens*, 26, 1477-1491.
- Lee, W. J. (2018). Application of Satellite Imagery to Research on Earthquake and Volcano. *Korean Journal of Remote Sensing*, 34, 1469-1478.
- Liaw, A., Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Liu, Y., Li, Z., Wei, B., Li, X., & Fu, B. (2019), Seismic Vulnerability Assessment at Urban Scale using Data Mining and GIScience Technology: Application to Urumqi (China). *Geomatics, Natural Hazards and Risk*, 10, 958-985.
- McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. Academic Press: New York, NY, USA, pp 105–142.
- Merghadi, A., Abderrahmane, B., & Tien Bui, D. (2018). Landslide Susceptibility Assessment at Mila Basin (Algeria): A Comparative Assessment of Prediction Capability of Advanced Machine Learning Methods. *ISPRS International Journal of Geo-Information*, 7, 268.

- Ministry of Public Safety and Security (MPSS). (2017). Report on the 9.12 earthquake and countermeasures, Seoul, Korea.
- Moustafa, S. S., Al-Arifi, N.S., Jafri, M.K., Naeem, M., Alawadi, E.A., & Metwaly, M.A. (2016). First Level Seismic Microzonation Map of Al-Madinah Province, Western Saudi Arabia using the Geographic Information System Approach. *Environmental earth sciences*, 75, 251.
- National Emergency Management Agency (NEMA). (2012). Construction of active fault map and seismic risk map. National Emergency Management Agency, Seoul, Korea, pp. 899.
- Panahi, M., Rezaie, F., & Meshkani, S. (2014). Seismic Vulnerability Assessment of School Buildings in Tehran City Based on AHP and GIS. *Natural Hazards and Earth System Sciences*, 14, 969-979.
- Paul, S. S., Li, J., Li, Y., & Shen, L. (2019). Assessing Land use-land Cover Change and Soil Erosion Potential using a Combined Approach through Remote Sensing, RUSLE and Random Forest Algorithm. *Geocarto Int*, 1-15.
- Pourghasemi, H. R., Gayen, A., Panahi, M., Rezaie, F., & Blaschke, T. (2019). Multi-Hazard Probability Assessment and Mapping in Iran. *Sci. Total Environ*, 692, 556-571.

- Pradhan, B. A. (2013). 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*, 51, 350-365.
- Rahmati, O., Pourghasemi, H. R., & Zeinivand, H. (2016). Flood Susceptibility Mapping using Frequency Ratio and Weights-of-Evidence Models in the Golastan Province, Iran. *Geocarto Int*, 31, 42-70.
- Rezaie, F., & Panahi, M. (2015). GIS Modeling of Seismic Vulnerability of Residential Fabrics Considering Geotechnical, Structural, Social and Physical Distance Indicators in Tehran using Multi-Criteria Decision-Making Techniques. *Natural Hazards and Earth System Sciences*, 15, 461-474.
- Riedel, I., Guéguen, P., Dalla Mura, M., Pathier, E., Leduc, T., & Chausson, J. (2015). Seismic Vulnerability Assessment of Urban Environments in Moderate-to-Low Seismic Hazard Regions using Association Rule Learning and Support Vector Machine Methods. *Nat. Hazards*, 76, 1111-1141.
- Şengezer, B., Ansal, A., & Bilen, Ö. (2008). Evaluation of Parameters Affecting Earthquake Damage by Decision Tree Techniques. *Nat. Hazards*, 47, 547-568.

- Shin, D. H., Hong, S., & Kim, H. (2016). Investigation on Effective Peak Ground Accelerations Based on the Gyeongju Earthquake Records. *Journal of the Earthquake Engineering Society of Korea*, 20, 425-434.
- Son, J. (2017). Susceptibility assessment of landslide and land subsidence applying the radius of influence to frequency ratio model. Graduate School of Seoul National University, Seoul, Korea.
- Taalab, K., Cheng, T., & Zhang, Y. (2018). Mapping Landslide Susceptibility and Types using Random Forest. *Big Earth Data*, 2, 159-178.
- Tesfamariam, S., & Liu, Z. (2010). Earthquake Induced Damage Classification for Reinforced Concrete Buildings. *Struct. Saf*, 32, 154-164.
- Thaker, T. P., Savaliya, P. K., Patel, M. K., & Patel, K. A. (2018). GIS Based Seismic Risk Analysis of Ahmedabad City, India. In *GeoShanghai International Conference, Shanghai, China, May*; pp. 108-116.
- Vapnik, V., (1995). *The Nature of Statistical Learning Theory*, Springer: New York, NY, USA.
- Vapnik, V. (1998). *Statistical Learning Theory*, Wiley: New York, NY, USA, pp. 156-160.

- Walker, B. B., Taylor-Noonan, C., Tabbernor, A., Bal, H., Bradley, D., Schuurman, N., & Clague, J. J. (2014). A Multi-Criteria Evaluation Model of Earthquake Vulnerability in Victoria, British Columbia. *Nat. Hazards*, 74, 1209-1222.
- Wallemacq, P. (2018). *Economic Losses, Poverty & Disasters: 1998-2017*, Centre for Research on the Epidemiology of Disasters, Brussels, Belgium.
- Wang, L., Guo, M., Sawada, K., Lin, J., & Zhang, J. (2016). A Comparative Study of Landslide Susceptibility Maps using Logistic Regression, Frequency Ratio, Decision Tree, Weights of Evidence and Artificial Neural Network. *Geosciences Journal*, 20, 117-136.
- Wikipedia, Google, “<https://ko.wikipedia.org/wiki/%EA%B2%BD%EC%A3%BC%EC%8B%9C>” (accessed on 12 July 2019).
- Xiao, T., Yin, K., Yao, T., & Liu, S. (2019). Spatial Prediction of Landslide Susceptibility using GIS-Based Statistical and Machine Learning Models in Wanzhou County, Three Gorges Reservoir, China. *Acta Geochimica*, 38, 654-669.
- Xu, C., Xu, X., Dai, F., & Saraf, A. K. (2012). Comparison of Different Models for Susceptibility Mapping of Earthquake Triggered Landslides Related with the 2008 Wenchuan Earthquake in China. *Comput. Geosci*, 46, 317-329.

- Xu, C., Shen, L., & Wang, G. (2016). Soft Computing in Assessment of Earthquake-Triggered Landslide Susceptibility. *Environmental Earth Sciences*, 75, 767.
- Yalcin, A., Reis, S., Aydinoglu, A. C., & Yomralioglu, T. (2011). 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*, 85, 274-287.
- Yilmaz, I. (2010). Comparison of Landslide Susceptibility Mapping Methodologies for Koyulhisar, Turkey: Conditional Probability, Logistic Regression, Artificial Neural Networks, and Support Vector Machine. *Environmental Earth Sciences*, 61, 821-836.
- Youssef, A. M., Pradhan, B., & Sefry, S. A. (2016). Flash Flood Susceptibility Assessment in Jeddah City (Kingdom of Saudi Arabia) using Bivariate and Multivariate Statistical Models. *Environmental Earth Sciences*, 75, 12.