

# 통계분석 기법과 머신러닝 기법의 비교분석을 통한 건물의 지진취약도 공간분석

김성훈<sup>1</sup>, 김상빈<sup>2\*</sup>, 김대현<sup>3</sup>

<sup>1</sup>남서울대학교 공간정보공학과 교수, <sup>2</sup>남서울대학교 공간정보공학과 대학원 박사과정학생, <sup>3</sup>한국전력 데이터사이언스연구소 일반연구원

## A Spatial Analysis of Seismic Vulnerability of Buildings Using Statistical and Machine Learning Techniques Comparative Analysis

Seong H. Kim<sup>1</sup>, Sang-Bin Kim<sup>2\*</sup>, Dae-Hyeon Kim<sup>3</sup>

<sup>1</sup>Professor, Department of GIS Engineering, Namseoul University

<sup>2</sup>Doctorate course, Department of GIS Engineering, Namseoul University

<sup>3</sup>Research Engineer, AI Research Team, Data Science Lab, Korea Electric Power Corporation

**요약** 최근 지진 발생 빈도가 증가하고 있는 반면 국내 지진 대응 체계는 취약한 현실에서, 본 연구의 목적은 통계분석 기법과 머신러닝 기법을 활용한 공간분석을 통해 건물의 지진취약도를 비교분석 하는 것이다. 통계분석 기법을 활용한 결과, 최적화척도법을 활용해 개발된 모델의 예측정확도는 약 87%로 도출되었다. 머신러닝 기법을 활용한 결과, 분석된 4가지 방법 중, Random Forest의 정확도가 Train Set의 경우 94%, Test Set의 경우 76.7%로 가장 높아, 최종적으로 Random Forest가 선정되었다. 따라서, 예측정확도는 통계분석 기법이 약 87%, 머신러닝 기법이 76.7%로, 통계분석 기법의 예측정확도가 더 높은 것으로 분석되었다. 최종 결과로, 건물의 지진취약도는 분석된 건물데이터 총 22,296개 중, 1,627(0.1%)개의 건물데이터는 통계분석 기법 사용 시 더 위험하다고 도출되었고, 10,146(49%)개의 건물데이터는 동일하게 도출되었으며, 나머지 10,523(50%)개의 건물데이터는 머신러닝 기법 사용 시 더 위험하게 도출되었다. 기존 통계분석 기법에 첨단 머신러닝 기법 활용결과가 추가로 비교검토 됨으로써 공간분석 의사결정에 있어서, 좀더 신뢰도가 높은 지진대응책 마련에 도움이 되길 기대한다.

**키워드** : 지진, 통계, 머신러닝, 공간분석, 취약도

**Abstract** While the frequency of seismic occurrence has been increasing recently, the domestic seismic response system is weak, the objective of this research is to compare and analyze the seismic vulnerability of buildings using statistical analysis and machine learning techniques. As the result of using statistical technique, the prediction accuracy of the developed model through the optimal scaling method showed about 87%. As the result of using machine learning technique, because the accuracy of Random Forest method is 94% in case of Train Set, 76.7% in case of Test Set, which is the highest accuracy among the 4 analyzed methods, Random Forest method was finally chosen. Therefore, Random Forest method was derived as the final machine learning technique. Accordingly, the statistical analysis technique showed higher accuracy of about 87%, whereas the machine learning technique showed the accuracy of about 76.7%. As the final result, among the 22,296 analyzed building data, the seismic vulnerabilities of 1,627(0.1%) buildings are expected as more dangerous when the statistical analysis technique is used, 10,146(49%) buildings showed the same rate, and the remaining 10,523(50%) buildings are expected as more dangerous when the machine learning technique is used. As the comparison of the results of using advanced machine learning techniques in addition to the existing statistical analysis techniques, in spatial analysis decisions, it is hoped that this research results help to prepare more reliable seismic countermeasures.

**Key Words** : Earthquake, Statistics, Machine learning, Spatial analysis, Seismic vulnerability

This paper was carried out with the support of Namseoul University.

\*Corresponding Author : Kim Sang Bin(kcb9241@naver.com)

Received November 15, 2022

Accepted January 20, 2023

Revised December 14, 2022

Published January 28, 2023

## 1. 서론

### 1.1 연구 배경 및 목적

우리나라는 지진이 과거로부터 현재까지 자주 일어났고, 최근 발생한 경주, 포항, 괴산에서의 지진은 우리나라가 더 이상 지진 안전지대가 아님을 보여주고 있다. 건물의 지진취약도를 결정하는 주요 요소 중 하나인 내진설계가 고려된 건물은 많지 않다. 기상청에서 조사한 통계표에 의하면, 지진 발생은 1999년 기준 이전 연간 19.1회였으나, 이후 70.6회 일어나고 있는 것으로 파악된다. 또한, 국내 지진 규모 순위를 보면 2000년 이후 발생한 지진이 상위 14개 지진 중 10개로 파악된다. 그러나 본 연구의 대상 지역인 강남구가 포함된 서울시 건물은 전체 중 18.9%만 내진설계가 되어있어 지진에 상당히 취약한 것으로 판단된다. 따라서, 본 연구는 통상적으로 예측에 많이 사용되고 있는 통계분석 기법과 머신러닝 기법을 적용 및 비교하고 추가적인 도구로써 지역 시스템 분석에 유리한 공간분석 기법을 활용하여 건물의 지진취약도를 비교 분석 할 예정이다.

### 1.2 연구 범위 및 절차

지진 연구는 예방, 피해 저감 활동, 사후 복구 활동의 영역으로 구분될 수 있다. 본 연구는 위 영역 중 예방영역

에 해당하며 연구 수행 절차는 Fig. 1과 같다.

연구 수행 절차는 모델링 섹션, 통계분석 및 머신러닝 섹션, GIS 섹션으로 크게 3가지로 나뉜다. 모델링 섹션은 영향요인 및 모델의 구성요소 연구 및 관련 데이터를 수집, 모델 개발과 모델 유효성 검증으로 구성된다. 통계분석 및 머신러닝 섹션은 통계분석 및 머신러닝을 활용하여 모델을 개발하고 통계지표와 머신러닝지표 검증 부분으로 구성된다. 여기서 모델 유효성 검증이 통과하면 다음 단계로 진행되고, 못하면 통계분석 및 머신러닝 부분으로 돌아간다. 마지막으로 GIS 섹션은 GIS의 툴을 선택한 뒤 분석 범위, 공간 및 건물 데이터를 입력하고 개발된 모델 및 선택된 머신러닝 모델을 적용한 뒤 공간분석실행으로 구성된다.

## 2. 이론적 동향

### 2.1 공간분석의 동향

공간분석이란, 의사결정을 지원하기 위해 공간 자료와 속성자료를 함께 분석할 수 있게 하는 분야이다.

본 연구에서 주로 참고한 공간분석을 활용한 연구는 다음과 같다. 첫 번째 관련 연구는 과거 지진 발생 빈도가 높았던 포항, 경주 근처의 데이터를 주로 참고하였다[1]. 그리고 예측 모델로는, 2000년대 이전은 우리나라에 지진피해를 본 건물데이터가 축적이 안 돼 있어, 토목, 건축

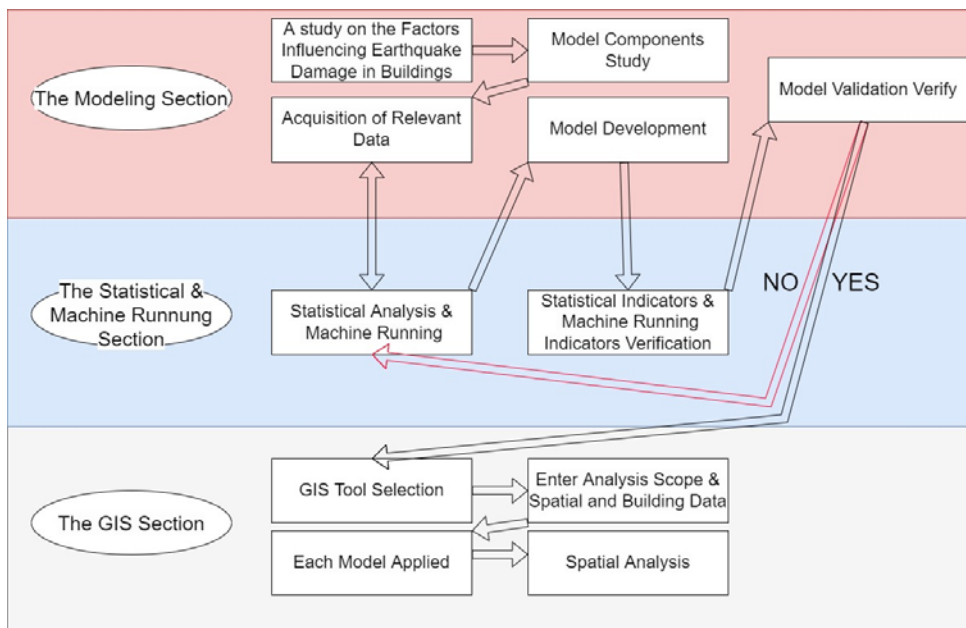


Fig. 1. Scope and procedure of current study

학적인 공식이 많이 사용되었다[2]. 또한, 본 연구의 주제로 선택된 건물을 대상으로 기존에 있던 모델을 활용하여 지진취약도 분석을 수행한 연구는 우리나라에서 2000년대 이후부터 시도되기 시작했다[3-4]. 다음으로 AHP(Analytic Hierarchy Process) 기법을 이용하여 18개의 하위지표를 선정해 GIS로 지진피해 예측을 경주지역을 대상으로 한 연구가 있다[5]. 그리고 대전지역을 중심으로 지진재해 구역화를 GIS 기반의 지반정보 시스템을 활용하여 예측한 연구도 있다[6].

2.2 통계분석의 동향

통계분석이란, 자료를 통계학적으로 수집, 정리하여 분석한 뒤 의미 있는 결과를 도출하는 분야이다.

본 연구 관련 통계분석 기법을 활용한 공간분석에 관한 특징적인 연구는 다음과 같다. 해당 연구에서는 건물 지진 피해에 영향을 미치는 인자를 FEMA(Federal Emergency Management Agency) 보고서를 활용해 조사한 뒤, NIST(National Institute of Standards and Technology)에서 모델을 구성하는 데이터를 습득한다 [7-9]. 습득된 국내의 총 348개의 지진피해 경험데이터를 사용하여 통계분석을 시행해 모델을 개발하였다. 일련의 통계분석 결과, 최적화최도법에 따라 개발된 모델이 최선의 결과를 보였으며, 모델의 예측정확도는 87%로 산정되었다. 개발된 모델 식을 공간분석에 적용하기 위해, 서울시 구도심과 신도심의 특성을 대표하는 종로구와 강남구를 대상 지역으로 선정하였고, 분석 결과 지진피해 관점에서 강남구의 건물이 종로구의 건물보다 위험한 것으로 예상되었다[10-11].

2.3 머신러닝의 동향

머신러닝이란, 인공지능의 한분야로 사람이 학습하듯이 컴퓨터도 데이터를 학습시켜 새로운 지식을 얻어내는 분야이다.

본 연구에서 주로 참고한 머신러닝 기법을 활용한 공간분석에 관한 연구는 다음과 같다. 첫 연구에서 Longhai (China) 지역의 LSM (Landslide Susceptibility Mapping)을 위한 3개의 기계 학습 알고리즘(MLA)이 평가·비교되었다. 우선 다양한 데이터 소스에서 총 14개의 산사태 관련 변수를 얻은 뒤, 주파수 비율(FR) 및 SVM(Support Vector Machine) 방법을 이용하여 변수들의 인과관계를 분석하고 학습에 유용하다고 생각되는 변수가 선정되었

다. 그리고 RBF(Radial Basis Fuction) Classifier, RBF Network, NB 3가지 모델을 이용하여 데이터를 학습한 후 성능을 비교하였다[12].

다음으로 Lao Cai(Vietnam) 지역의 열대 산불에 관한 사례연구를 통해 산불위험의 공간 패턴을 분석하고 예측하기 위한 새로운 기계 학습 방법론이 제안되었다. 이를 위해 연구영역에 대한 GIS 데이터베이스가 구축되었으며, 여기에는 화재 발생 257개의 케이스 관련 10가지 영향 요인(경사도, 지형, 고도, 풍속 및 습도 등)이 포함된다. 산불피해도와 이러한 요인의 관련성 수준이 확률적 최적화 알고리즘을 사용하여 분석·평가되었다[13].

2.4 기존연구와 차별점

우리나라 지진연구는 2016, 2017년 경주, 포항지진 이후 연구가 활발히 진행되었다. 그 이전의 공간분석 연구는 건축, 토목학적인 모델을 사용하여 시각화에만 공간 분석을 사용하였고, 데이터는 가상의 데이터를 많이 사용하였다. 또한, 기존 연구는 통계분석 기법과 머신러닝 기법 중 한 가지 방법만 활용해 분석함으로써 예측정확도의 검증에 대한 다소간의 아쉬움이 있다.

하지만 본 연구는 실제 지진피해를 입은 건물 데이터를 활용하여 모델링을 통계분석 기법과 머신러닝 기법 두 가지 방법을 모두 활용해 공간분석에 적용해, 좀더 신뢰도가 높은 지진대응책 마련에 도움이 될 수 있다는 차별점을 가지고 있다.

3. 통계분석과 머신러닝 기법을 활용한 모델 도출

3.1 통계분석 기법을 활용한 모델링 결과

건물의 지진취약도 공간분석을 위해 다음과 같이 자체 개발된 통계분석 모델이 사용되었다.

Table 1. Model summary using optimization scale method

Model Summary						
Multiple R(M)	R <sup>2</sup>	Revised R <sup>2</sup>	Normalization "R <sup>2</sup> " (1-Error)	Prediction error	Expected prediction error	
					Estimated value	Standard error
.931	.867	.862	.859	.141	.149	.011

최적화최도법을 사용해 개발된 “모형 요약”(Table 1 참고)에서 R 제곱을 보면, 0.867로 약 87%의 매우 높은 정확도를 가진 모델이 도출되었다.

**Table 2. ANOVA of the developed model using optimization scale method**

ANOVA					
	Sum of squares	Degree of freedom	Average <sup>2</sup>	F	Significance probability
Regression analysis(R)	299.095	13	23.007	157.132	.000
Residual	48.905	334	.146		
The entire	348.000	347			

다음으로 분산분석표(Table 2. 참고)를 보면, 개발된 모델의 유의확률은 0.000으로 유의한 모델이 나왔다는 것을 확인할 수 있다.

**Table 3. Optimization scale factor**

	Coefficient				
	Standardization factor		Degree of freedom	F	Significance probability
	Beta	Standard brush strap (1000) Estimate error			
Shape of the building	.228	.022	3	111.932	.000
Material of the building	.169	.021	2	63.902	.000
Dating of the building	.161	.020	1	62.326	.000
Height of the building	.174	.021	3	69.154	.000
The foundation (Ground Condition)	.112	.019	2	35.547	.000
The strength of an earthquake (MMI)	.438	.025	2	316.216	.000

또한, 최적화척도법의 계수(Table 3 참고)를 보면, 모델을 구성하는 각 인자의 유의확률은 모두 0.000으로 유의한 것을 확인할 수가 있다.

통계분석에서 선정된 각 인자들은 미국 FEMA의 보고서를 참고하여, 우리나라에서 구할 수 있는 GIS 데이터에 맞춰 편집·구성하였다.

통계분석 결과, 최적화척도법을 활용해 개발된 모델의 모델 식은 Fig. 2와 같다. 해당 식은 각 구성인자의 Beta 값(Table 3 참고)인 가중치를 활용하여 만들어졌다.

$$Y = 0.228 \text{ shape of the building} + 0.169 \text{ material of the building} + 0.161 \text{ dating of the building} + 0.174 \text{ height of the building} + 0.112 \text{ the foundation} + 0.438 \text{ the strength of an earthquake MMI}$$

**Fig. 2. Model expression**

건물의 지진 경험데이터 348개를 위의 식에 대입한 뒤, Y값의 범위를 확인하였다.

확인 결과, 다음과 같이 나왔다. 1(안전) = 1.282 ~ 1.736, 2(반파) = 1.737 ~ 2.835, 3(완파) = 2.836 ~ 4.739로, 안전의 1.282는 최솟값으로 모든 인자가 1일 경우의 값이고 완파의 4.739는 모든 인자가 최댓값인 경우의 값이다. 안전은 지진피해 후 바로 사용이 가능한 건물, 반파는 지진피해 후 수리한 뒤 사용이 가능한 건물, 완파는 지진피해 후 무너지거나 수리하여도 사용할 수 없는 건물이다.

**3.2 머신러닝 기법을 활용한 적정 모델 도출 결과**

건물의 지진취약도 공간분석을 위해 다음과 같은 머신러닝 모델들이 실험되었다.

- ① k-NN - 비모수 적용 방식으로 분류 또는 회귀 분석에 사용되는 방법이다. 전체 입력은 특정 공간 안에 k개의 특성을 가진 유사한 데이터셋으로 구성된다.
- ② Random Forest - 여러 개의 학습 모델을 결합하는 앙상블(Ensemble) 학습 모델의 하나로, 여러 개의 결정 트리(Decision Tree)로부터 나온 결과를 투표하여 선정 혹은 평균을 내는 방식으로 동작한다.
- ③ Decision Tree - 입력 변수를 이용하여 목표한 변수의 값을 예측하는 학습 모델이다. 데이터를 분석한 뒤 데이터 사이에 존재하는 패턴을 규격화하여 입력값이 분류되거나 회귀된다.
- ④ SVM - SVM 알고리즘은 주어진 데이터를 기준으로 새로운 데이터가 어떤 카테고리에 속할지 예측하는 비확률적 이진 선형 분류 모델이다.

각 모델의 실험을 위해 Python의 Package인 Scikit-Learn<sup>1)</sup>을 사용하였으며, 각 모델의 매개변수는 Train Data에 실험했을 때 가장 좋은 성능을 보였던 매개변수로 최종적으로 선택하였다. 각 모델의 실험 결과는 Table 4와 같다.

Train Set에 대한 정확도는 Random Forest와 Decision Tree가 94%로 가장 높은 결과를 보였다. 또한, Test Set에 대한 정확도는 Random Forest가 76.7%로 Decision Tree의 73.3%보다 3.4% 더 높은 정확도를 보여주므로 최종적으로 강남에 대한 건물의 지진취약도 산

1) Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830.

정을 위한 머신러닝 모델로 Random Forest가 선정되었다.

**Table 4. Machine learning results**

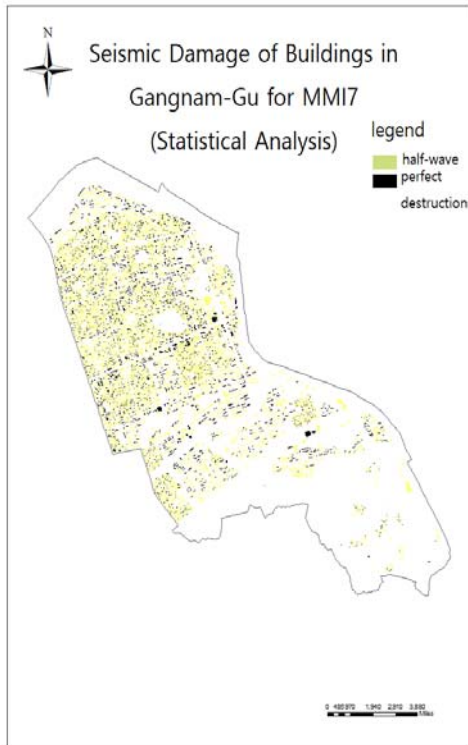
Machine Learning Techniques	Train Accuracy	Test Accuracy
k-NN	82%	73.3%
Random Forest	94%	76.7%
Decision Tree	94%	73.3%
SVM	84%	66.7%

### 4. 공간분석의 분석 결과

#### 4.1 통계분석 결과를 활용한 공간분석

통계분석 기법 활용 결과로 개발된 모델식 Fig. 2를 공간분석에 적용한 결과는 다음과 같이 나왔다. MMI7의 경우 안전이 0%, 반파가 83%이고 완파가 17%로 예측되었다. 대부분 건물이 반파로 나왔으며, 안전한 건물은 없는 것으로 확인되었다.

통계분석 기법을 활용해 개발된 식을 공간분석에 적용한 결과는 Fig. 3과 같다.

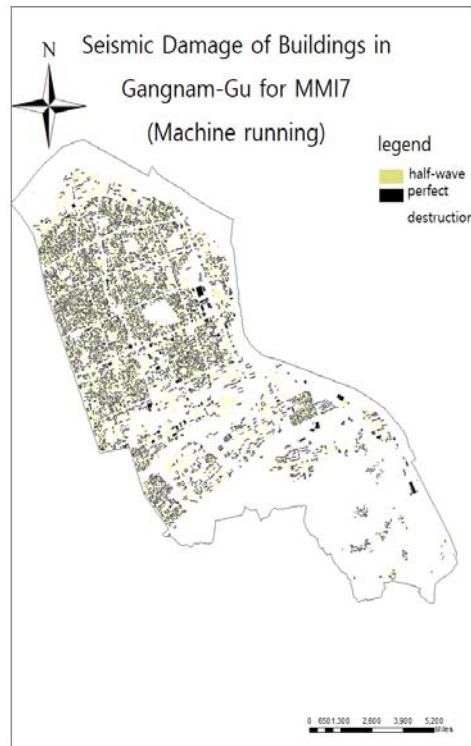


**Fig. 3. Spatial analysis using statistical analysis**

#### 4.2 머신러닝 결과를 활용한 공간분석

머신러닝 기법의 비교검토 결과(Table 4 참고)로 도출된 Random Forest 모델을 공간분석에 적용한 결과는 다음과 같이 나왔다. MMI7의 경우 안전이 0%, 반파가 43%, 완파가 57%로 예측되었다. 반파와 완파의 등급을 받은 건물의 숫자가 비슷하게 나왔으며, 안전한 건물은 없는 것으로 확인되었다.

머신러닝 기법을 활용해 도출된 모델을 공간분석에 적용한 결과는 Fig. 4와 같다.



**Fig. 4. Spatial analysis using machine running**

### 5. 결론

건물데이터 총 22,296개 중 1,627(0.1%)개의 건물데이터는 통계분석 기법을 활용한 결과 개발된 모델 사용시 더 위험하다고 도출되었고, 10,146(49%)개의 건물데이터는 동일하게 도출되었으며, 나머지 10,523(50%)개 건물데이터는 머신러닝 기법을 활용해 도출된 모델 사용시 더 위험하게 도출되었다.

통계분석 기법과 머신러닝 기법을 활용한 건물의 지진취약도 평가 결과를 보면, 모델의 예측정확도는 통계분석

을 사용한 결과 약 87%, 머신러닝을 사용한 결과 약 76.7%로 도출되었고, 두 기법의 정확도 차이는 10.3%이다. 그러나 두 기법의 공간분석 결과를 비교하면, 약 46%의 건물만이 같은 예측 결과를 보이고, 나머지 약 54%의 건물은 다른 예측 결과를 보인다.

결론적으로, 통계분석 기법을 활용한 경우의 예측 결과에 대비하여 머신러닝 기법을 활용한 경우의 예측 결과가 전반적으로 위험한 건물이 더 많은 것으로 도출되고 있다.

본 연구에서는 우선 건물의 내적 요인과 환경적 요인만을 고려하여 지진피해를 예측하였다. 추후 사회적 요인(주거인구, 도로와의 거리 등)을 고려하여 분석한다면 더 유용한 결과획득이 가능할 것으로 판단된다.

## REFERENCES

- [1] Park, S. J. (2007). Tectonic movement in the Korean peninsula (II): A geomorphological interpretation of the spatial distribution of earthquakes. *Journal of the Korean Geographical Society*, 42(4), 488-505.
- [2] Chun, Y. (2017). A Study on Earthquake Hazard Mapping using Risk Factors. *Proc. of Korean Society for Geospatial Information Science, Seoul, Korea, May*, 25-26.
- [3] Koh, J. H., Kwon, J. H., & Choi, Y. S. (2005). Error Assessment of Attitude Determination Using Wireless Internet-Based DGPS. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, 23(3), 239-249.
- [4] Hwang, C. H. (2012). A Study on Urban Damage Assessment by Earthquake Damage Assessment System. *Seoul Venture University*.
- [5] Han, J., & Kim, J. (2019). A GIS-based seismic vulnerability mapping and assessment using AHP: A case study of Gyeongju, Korea. *Korean Journal of Remote Sensing*, 35(2), 217-228.
- [6] Sun, C. G. (2009). Seismic zonation on site responses in Daejeon by building geotechnical information system based on spatial GIS framework. *Journal of the Korean Geotechnical Society*, 25(1), 5-19.
- [7] Federal Emergency Management Agency (US) (Ed.). (2015). *Rapid visual screening of buildings for potential seismic hazards: a handbook*. Government Printing Office.
- [8] Federal Emergency Management Agency (US) (Ed.). (2015). *Rapid visual screening of buildings for potential seismic hazards: a handbook. a Handbook (Third Edition)*. Government Printing Office.
- [9] Catlin, A. C., & Pujol, S. (2015). NIST Disaster and Failure Studies Data Repository: *The Chile Earthquake Database—Ground Motion and Building Performance Data from the 2010 Chile Earthquake—User Manual*.
- [10] S. B. Kim. (2019). A Development of a Seismic Vulnerability Model and Relative Spatial Analysis for Buildings. Paper of Master's Degree. *Namseoul University*, Cheonan.
- [11] Kim, S. B., & Kim, S. H. (2020). A Development of a Seismic Vulnerability Model and Spatial Analysis for Buildings. *Journal of the Korea Convergence Society*, 11(10), 9-18.
- [12] He, Q., Shahabi, H., Shirzadi, A., Li, S., Chen, W., Wang, N., ... & Ahmad, B. B. (2019). Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. *Science of the total environment*, 663, 1-15.
- [13] Bui, D. T., Hoang, N. D., & Samui, P. (2019). Spatial pattern analysis and prediction of forest fire using new machine learning approach of Multivariate Adaptive Regression Splines and Differential Flower Pollination optimization: A case study at Lao Cai province (Viet Nam). *Journal of environmental management*, 237, 476-487.

김성훈(Seong H. Kim)

[정회원]

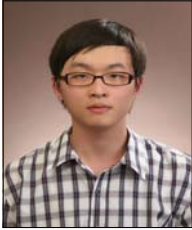


- 1983년 2월 : 연세대학교 토목공학 과(학사)
- 1987년 5월 : 뉴욕주립대 토목공학 과(공학석사)
- 1993년 8월 : 뉴욕주립대 토목공학 과(공학박사)

- 1993년 9월~1999년 8월 : 삼성 SDI 책임연구원
- 1999년 9월~현재 : 남서울대학교 드론공간정보공학과 교수
- 관심분야 : 공간분석, 모델링, 컨설팅
- E-Mail : gotit@nsu.ac.kr

김 상 빈(Sang-Bin Kim)

[정회원]



- 2017년 2월 : 남서울대학교 공간정보공학(학사)
- 2019년 6월 : 남서울대학교 공간정보공학(공학석사)
- 2020년 5월~2021년 5월 : 울산연구원 전문연구원

- 2021년 5월~2022년 3월 : 한국환경연구원 위촉연구원
- 2022년 9월~현재 : 남서울대학교 공간정보공학(박사)재학
- 관심분야 : 공간분석, 통계, 모델링
- E-Mail : kcb9241.naver.com

김 대 현(Dae-Hyeon Kim)

[정회원]



- 2018년 2월 : 홍익대학교 컴퓨터공학(학사)
- 2020년 2월 : 성균관대학교 인공지능학(석사)
- 2020년 11월~현재 : 한국전력공사 데이터사이언스연구소

- 관심분야 : 머신러닝, 패턴인식, 음성처리
- E-Mail : kdh92@skku.edu