

Random Forest를 활용한 산사태 피해 영향인자 평가: 충주시 산사태를 중심으로

이재호¹ · 정유진¹ · 최정해^{2*}

¹경북대학교 지질학과 석사과정, ²경북대학교 지구과학교육과 교수

Evaluation of the Importance of Variables When Using a Random Forest Technique to Assess Landslide Damage: Focusing on Chungju Landslides

Jaeho Lee¹ · Youjin Jeong¹ · Junghae Choi^{2*}

¹Master's Student, Department of Geology, Kyungpook National University

²Professor, Department of Earth Science Education, Kyungpook National University

Abstract

Landslides are natural disasters that causes significant property damage worldwide every year. In Korea, damage due to landslides is increasing owing to the effects of climate change, and it is important to identify the factors that increase the prevalence of landslides in order to reduce the damage they cause. Therefore, this study used a random forest model to analyze the importance of 14 factors in influencing landslide damage in a specific area of Chungju, Chungcheongbuk-do province, Korea. The random forest model performed accurately with an AUC of 0.87 and the most-important factors were ranked in the order of aspect, slope, distance to valley, and elevation, suggesting that topographic factors such as aspect and slope more greatly influence landslide damage than geological or soil factors such as rock type and soil thickness. The results of this study are expected to provide a basis for mapping and predicting landslide damage, and for research focused on reducing landslide damage.

Keywords: landslide damage area, variable importance, random forest, frequency ratio

OPEN ACCESS

*Corresponding author: Junghae Choi
E-mail: choi.jh@knu.ac.kr

Received: 3 January, 2024
Revised: 7 March, 2024
Accepted: 11 March, 2024

© 2024 The Korean Society of Engineering
Geology



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

초 록

산사태는 전 세계적으로 매년 큰 재산 피해를 야기하는 자연 재해로 알려져 있다. 국내에서도 기후 변화의 영향으로 산사태 피해가 증가하는 경향을 보이고 있으며, 이로 인한 피해를 줄이기 위해서는 산사태를 증가시키는 인자들을 파악하는 것이 중요하다. 따라서 본 연구는 충청북도 충주시에서 발생한 산사태 피해에 영향을 미치는 변수들의 중요도를 평가하기 위해 랜덤포레스트 모델을 활용하여 14개의 인자들 사이의 중요도를 분석하였다. 연구 결과, 모델의 성능은 AUC가 0.87로 높은 정확도를 보이며, 변수 중요도는 경사 방향, 경사, 계곡까지의 직선 거리, 고도 순으로 정해졌으며, 이는 경사방향과 경사 등의 지형인자가 암종과 유효토심과 같은 지질과 토양인자보다 산사태 피해에 더 큰 영향을 미친다는 것을 시사한다. 이 연구 결과는 산사태 피해 예측지도의 제작 및 산사태 피해 감소에 초점을 맞춘 연구에 기초 자료로서 활용될 수 있을 것으로 기대된다.

주요어: 산사태 피해면적, 변수 중요도, 랜덤포레스트, 빈도비율

서론

산사태는 전 세계적으로 재산피해와 인명피해를 야기하는 자연재해로(Gaidzik and Ramírez-Herrera, 2021), 최근 지구 온난화의 영향으로 산사태의 발생빈도와 규모가 점차 증가하고 있다. 우리나라에서는 7월과 8월에 태풍과 집중호우 등으로 인해 산사태 발생이 집중되는데, 최근 기상청 자료에 따르면 지구 온난화로 인한 여름철 극한 강우 현상의 지속기간이 길어지면서 산사태 피해 또한 증가하고 있는 것으로 나타난다. 최근 30년간 연평균 강수량은 135.3 mm 증가했으나 강수일수는 21.2일로 줄어드는 양상을 보이며, 특히 극한 강우로 인한 산사태 피해는 2020년에 크게 증가한 것으로 나타났다(Fig. 1). 과거 10년, 20년의 평균을 비교했을 때, 충청북도 충주시에서 산사태 피해 증가가 큰 폭으로 뚜렷하게 나타났으며, 2020년 7월 말부터 8월 초까지 총 210건의 산사태가 발생하였고, 이로 인한 피해 면적은 총 376.23 ha에 달하고(Shin et al., 2022), 경제적인 피해액은 총 4천 8백억 원으로 추정된다(KMA, 2020).

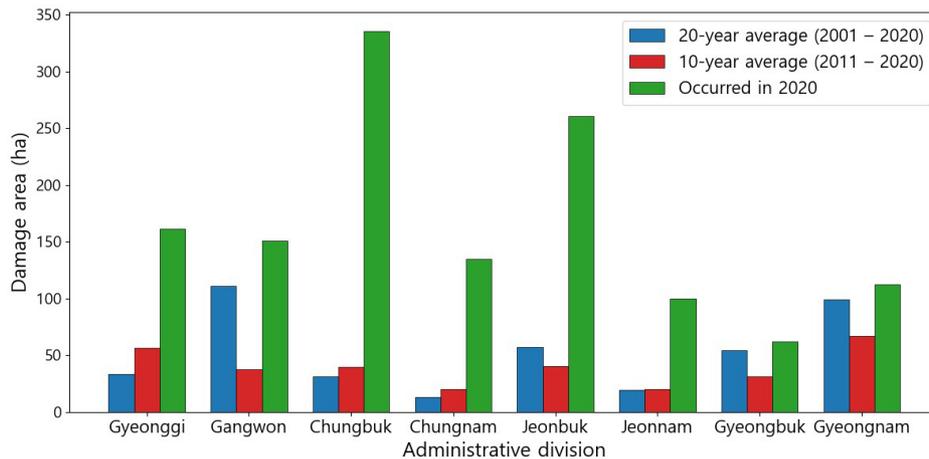


Fig. 1. Area of landslide damage in selected administrative areas (KFS, 2021).

최근 여러 분야에서 산사태 피해를 최소화하기 위한 산사태 리스크 분석(landslide risk analysis)과 산사태 위험평가(landslide hazard assessment)에 관한 연구가 진행되어 왔다. 산사태 리스크 분석은 산사태 발생 가능성과 사회적 경제적 피해를 예측하는 과정으로, 일반적으로 산사태 취약성 평가(landslide susceptibility assessment)가 우선적으로 수행되어야 한다(Lee, 2022). 산사태 취약성 평가는 크게 정성적 방법과 정량적 방법으로 구분된다(Van Westen et al., 1999). 지식 기반 분석기법은 정성적 방법 중 하나로써 산사태 취약성을 평가하는 전문가의 의견을 고려하여 산사태 발생여부를 추정하기 때문에 현장데이터 조사결과보다 전문가의 주관적인 판단이 더 중요하다. 따라서 이 기법은 현장데이터 조사결과보다 전문가의 주관적인 판단이 더 중요하고, 전문가의 현장 조사가 필수적이기 때문에 광역적인 지역에 적합하지 않다는 특징을 갖는다(Van Westen et al., 1999). 그에 반해 정량적 분석기법은 크게 데이터 기반 분석(data-driven analysis)과 물리사면모델 기반의 분석기법(physically based analysis)으로 구분된다. 데이터 기반 분석법은 과거의 산사태 발생 이력 정보를 이용하여 산사태에 영향을 주는 인자와 산사태 발생 사이의 상관성을 통계적으로 분석함으로써, 미래의 비슷한 조건에서 산사태가 일어날 가능성을 예측하는 방법이다(Lee, 2022). 최근 컴퓨터 기술의 발전으로 대량의 지질정보를 처리할 수 있는 머신 러닝(machine learning) 분석기법이 활용되고 있으며(Park and Lee, 2022), 그 예로는 의사결정나무(decision tree)(Pradhan, 2013), 서포트 벡터 머신(support vector machine)(Pradhan, 2013; Pham et al., 2019), 적응형 신경망 퍼지 추론모델(Adaptive Neuro-Fuzzy Inference System)(Pradhan, 2013), 회귀 모델(regression model)(Du et al.,

2017; Lombardo and Mai, 2018), 일반화 가법 모델(generalized additive models)(Goetz et al., 2015) 랜덤포레스트(random forest)(Goetz et al., 2015; Hong et al., 2015; Pham et al., 2016) 등이 있다. 현재까지 수행된 연구 대부분은 연구 지역에 다수의 알고리즘을 적용하여 정확도의 차이를 분석하는 비교연구나 하이브리드 머신러닝에 관한 연구이다. 본 연구에서는 선행연구 조사를 통해 다양한 머신러닝 모델의 성능을 비교검토한 결과, 결정나무기반의 앙상블 모델이 다른 모델에 비해 성능이 우수한 것으로 나타났다(Merghadi et al., 2020). 특히 랜덤포레스트는 훈련데이터에 대한 과적합문제가 없고 분류예측성능이 뛰어나다는 장점을 가지기 때문에(Goetz et al., 2015; Hong et al., 2015) 본연구의 분석모델로 채택하였다.

기존연구에서는 대체로 산사태 발생에 초점을 두고 산사태를 유발시킬 수 있는 변수에 대해 집중하였으며, 피해규모(Yang et al., 2014)나 피해 범위(Song and Lee, 2023)에 대한 연구는 있었지만, 산사태 피해증가에 영향을 미치는 변수에 대해서는 심도 깊게 다뤄지지 않았다. 따라서 본 연구에서는 2020년에 충청북도 충주시에서 발생한 산사태를 대상으로 산사태 피해 증가에 영향을 주는 변수들 중 어떠한 변수가 가장 영향을 크게 미치는지 확인하기 위해 랜덤포레스트를 활용하여 변수 중요도를 분석하였다. 이러한 분석은 향후 산사태의 피해 최소화화 및 피해 위험지도를 작성하는데 중요한 역할을 할 것으로 기대된다.

분석자료

산사태 현황 자료(landslide inventory)

산사태 현황 자료(landslide inventory)는 과거 발생한 산사태의 위치와 발생일시, 피해면적 등 과거 산사태의 공간분포에 대한 기본정보를 내포하고 있어 산사태 위험평가에 중요한 역할을 한다. 본 연구에서는 산림청에서 제공받은 191건의 산사태 현황 자료의 지면을 토대로 국토지리정보원에서 획득한 0.25 m 해상도의 항공사진과 대조하여 위치를 보다 정확하게 보정하였다. 충주시 산사태의 총 피해면적은 313.2 ha로 충주시 면적의 0.36%를 차지하며, 이 중 변수들의 값을 취득할 수 있는 104건의 산사태 현황자료를 이용하였다(Fig. 2). 보정된 산사태 현황자료를 살펴보면 전체 산사태의 29%는 1 ha이상의 산사태 피해를 입었고, 31%는 0.5 ha초과하고 1 ha미만의 산사태 피해를 입었으며, 나머지 51%는 0.5 ha 이하의 산사태 피해를 입었다.

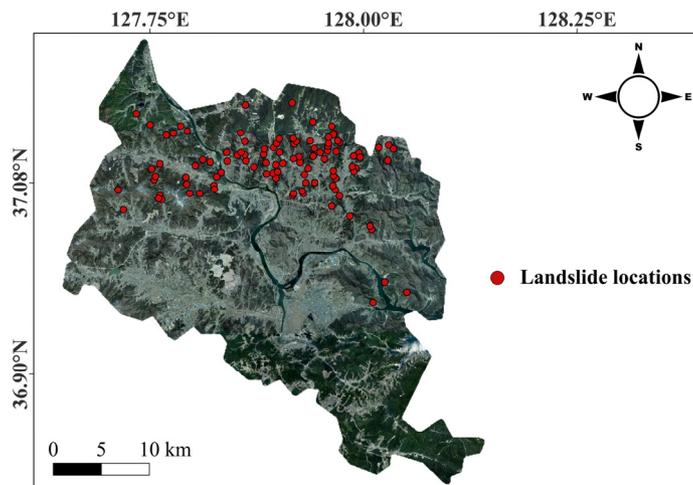


Fig. 2. Spatial distribution of landslides in the landslide inventory.

연구지역

본 연구지역은 충청북도 북부에 위치하며, 위도 36°47'40"에서 37°12'45"사이와 경도 127°39'40"와 128°08'10"에 위치해 있다(Fig. 2). 해당 지역은 전체 면적이 약 984 km²에 달하고, 고도는 39 m에서 982 m사이로 분지 형태를 이루고 있으며, 비교적 낮은 100 m 이하 지역에 사회기반 시설이 발달되어 있다. 충주시의 중심을 기준으로 동쪽지역은 험준한 산세를 보이는 반면에, 서쪽은 상대적으로 평탄한 지형을 보인다.

지질학적 특징을 살펴보면 전반적으로 화강암류와 편마암류로 구성되며, 충주시를 기준으로 북서와 북동방향에 있는 지역에서는 주로 선캄브리아기의 편마암류가, 중앙에는 중생대 쥐라기에 관입 되어 형성된 화강암류가 분포한다(Kim and Lee, 1965; Lee and Park, 1965; Kim et al., 1967a, 1967b; Park and Yeo, 1971; Lee and Kim, 1972; Yeo and Lee, 1975; Jeong et al., 1977; Ji et al., 1989)(Fig. 3).

연구지역에서는 2020년 7월 28일부터 8월 11일까지 발생한 집중 호우로 인하여 200여 건의 산사태가 발생하였으며, 이로 인해 인명과 재산피해가 야기되었다. 충주시의 산사태는 암종, 고도, 경사 등과 같은 비슷한 조건을 가지고 있음에도 불구하고 피해면적의 상이하게 나타났기 때문에 산사태 피해 연구지역으로 적합하여 연구를 수행하게 되었다.

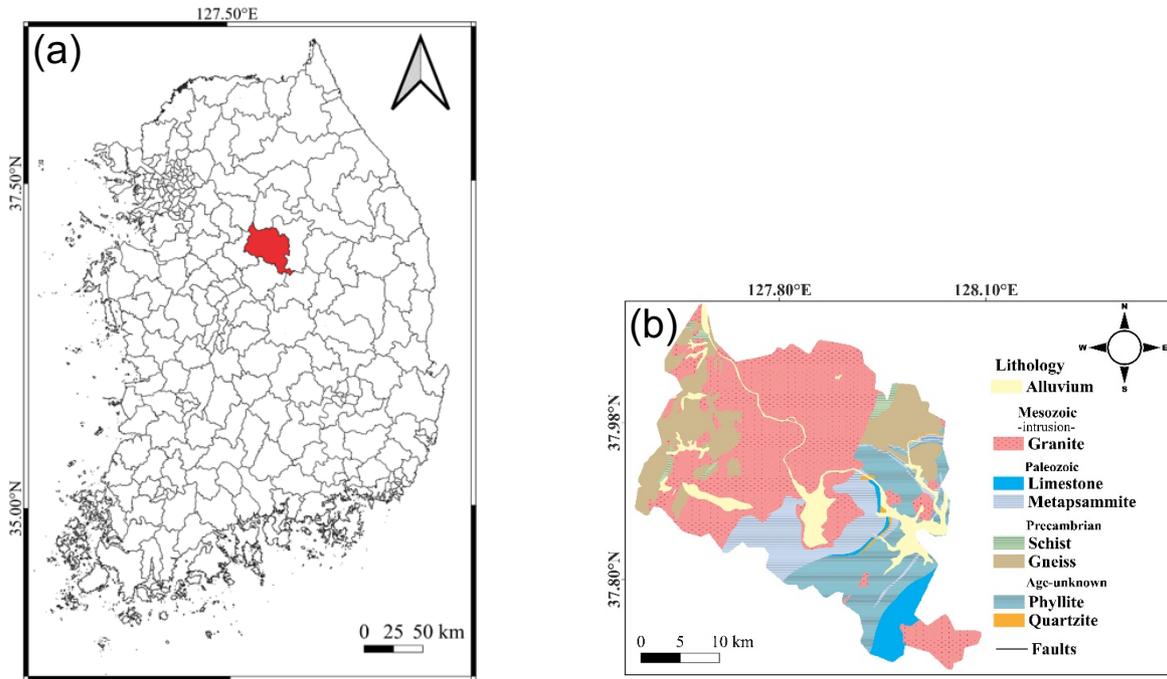


Fig. 3. Maps of the study area (WGS84 coordinate system): (a) location of Chungju-si in Korea, (b) geological map of Chungju-si.

영향인자 선택

산사태는 지형, 수문, 지질, 임상, 토양 등의 다양한 인자들에 의해 유발되고, 발생지역에 따라 피해면적도 상이하게 나타난다. 본 연구는 산사태의 피해 면적에 영향을 미치는 변수 중요도를 평가하는 것을 목적으로, 문헌 조사를 통해 산사태 유발인자 중에서 피해 면적에 크게 영향을 미치는 14개의 인자를 선별하였다. 선별된 피해 영향인자는 고도(elevation), 경사(slope), 경사방향(aspect), 수직곡률(profile curvature), 수평곡률(plan curvature), 계곡까지의 직선 거리(distance to valley), 경사길이와 기울기(LS-factor), 지형습윤지수(Topographic wetness index, TWI), 유효토심(soil thickness), 심도

토성(soil texture), 모암물질(soil material), 임상정보(forest cover), 암종(lithology)이다(Goetz et al., 2015; Lee and Kim, 2020; Fadhillah et al., 2022).

Table 1의 지형인자는 QGIS 프로그램을 이용하여 국토지리정보원(<https://www.ngii.go.kr>)에서 취득한 충주시의 90 m × 90 m 해상도를 가진 수치표고모델(DEM)에서 등고선을 추출한 후, TIN(Triangulated Irregular Network) 보간을 이용하여 10 m × 10 m 해상도를 가진 래스터 파일로 재구축되었다. 고도인자는 산사태와 관련하여 핵심적인 정보를 제공하는 동시에 해당지점의 기후도 영향을 미칠 수 있는 요소로, 기온, 강수량, 식생과 같은 여러 환경조건의 변화로 인해 산사태가 발생되기도 한다. 사면방향인자는 경사면의 방향에 따라 경사면이 받는 일조량, 풍량, 강수량이 달라지게 되며 (Cellek, 2021), 이러한 요소는 경사면의 수분 함량 등에 영향을 미쳐 산사태의 위험성을 증가시킬 수 있다(Roth, 1983). 수직곡률은 경사로를 통과하는 흐름의 가속과 감속에 영향을 미치기 때문에 수로의 침식과 퇴적정보를 제공하며, 수평곡률은 지표수 흐름의 수렴과 발산에 영향을 미친다(Wu et al., 2020). LS-factor은 경사면의 경사와 경사면의 길이를 같이 고려하여 나타낸 값으로 토양 침식에 대한 지형의 영향을 의미한다(Lee and Kim, 2020).

수문인자는 국토지리정보원(<https://www.ngii.go.kr>)에서 획득한 DEM을 10 m × 10 m로 재구축한 래스터 파일을 이용하여 각각의 주제도를 작성하였다. 계곡까지의 직선거리는 흐르는 물의 흐름을 분석하여 수로와 발생지점까지의 거리를 나타낸 인자이며, 경사면의 수위 변동에 영향을 미쳐 경사면의 불안정성을 초래할 수 있다. TWI는 사면의 습윤 정도를 판단하기 때문에 물의 공간적인 특성을 묘사할 수 있다. 추출된 수문인자와 지형인자는 연속형 데이터의 형태를 가지기 때문에, 이를 합리적인 범주화를 위해 초기에 변수의 항목(category)을 등간격으로 나눈 후 빈도비율(Frequency Ration, FR)을 사용하여 유사한 빈도비율 값을 가진 항목을 병합하여 재분류하였다(Guo et al., 2021).

토양인자는 국립농업과학원(<https://www.naas.go.kr>)에서 제공하는 정보를 활용하였다. 유효토심은 토양의 흡수성과 표면 유출수의 양에 영향을 미친다(Shahabi and Hashim, 2015). 모암물질과 심도토성은 토양의 침투성과 다공성을 결정하며, 이는 표면 유출수의 유속에 영향을 미친다(Regmi et al., 2013). 토양인자 항목의 분류는 국립농업과학원에서 제시하는 기준에 따라 분류하였다.

임상인자는 산사태 발생의 환경적인 영향을 파악하는데 도움을 주고, 지질인자의 암종은 투수성과 표면 유출수에 영향을 주며, 이러한 특성은 암석의 침식을 조절하고 산사태를 유발시키는 원인이 된다(Yesilnacar and Topal, 2005; Du et al., 2017). 전반적인 인자들의 출처 및 설명은 아래 Table 1에 제시되어 있다.

Table 1. Description of factors associated with landslides

Type	Factor	Source	Number of categories	Scale/Resolution
Topographical	Elevation	National Geographic Information Institute (NGII; https://map.forest.go.kr)	6	10 m × 10 m
	Slope		8	
	Aspect		7	
	Profile curvature		6	
	Plan curvature		9	
	LS-factor		5	
Hydrologic	Distance to valley		6	
	Topographic wetness index (TWI)		5	
Soil	Soil thickness	Korea National Institute of Agricultural Sciences (NAS; https://www.nas.go.kr)	6	1:5,000
	Soil texture		9	
	Soil material		6	

Table 1. Continued

Type	Factor	Source	Number of categories	Scale/Resolution
Forest type	Forest cover	Korea Forest Service (https://map.forest.go.kr)	4	1:5,000
Geological	Lithology	Korea Institute of Geoscience and Mineral Resources (KIGAM; https://data.kigam.re.kr)	3	1:50,000

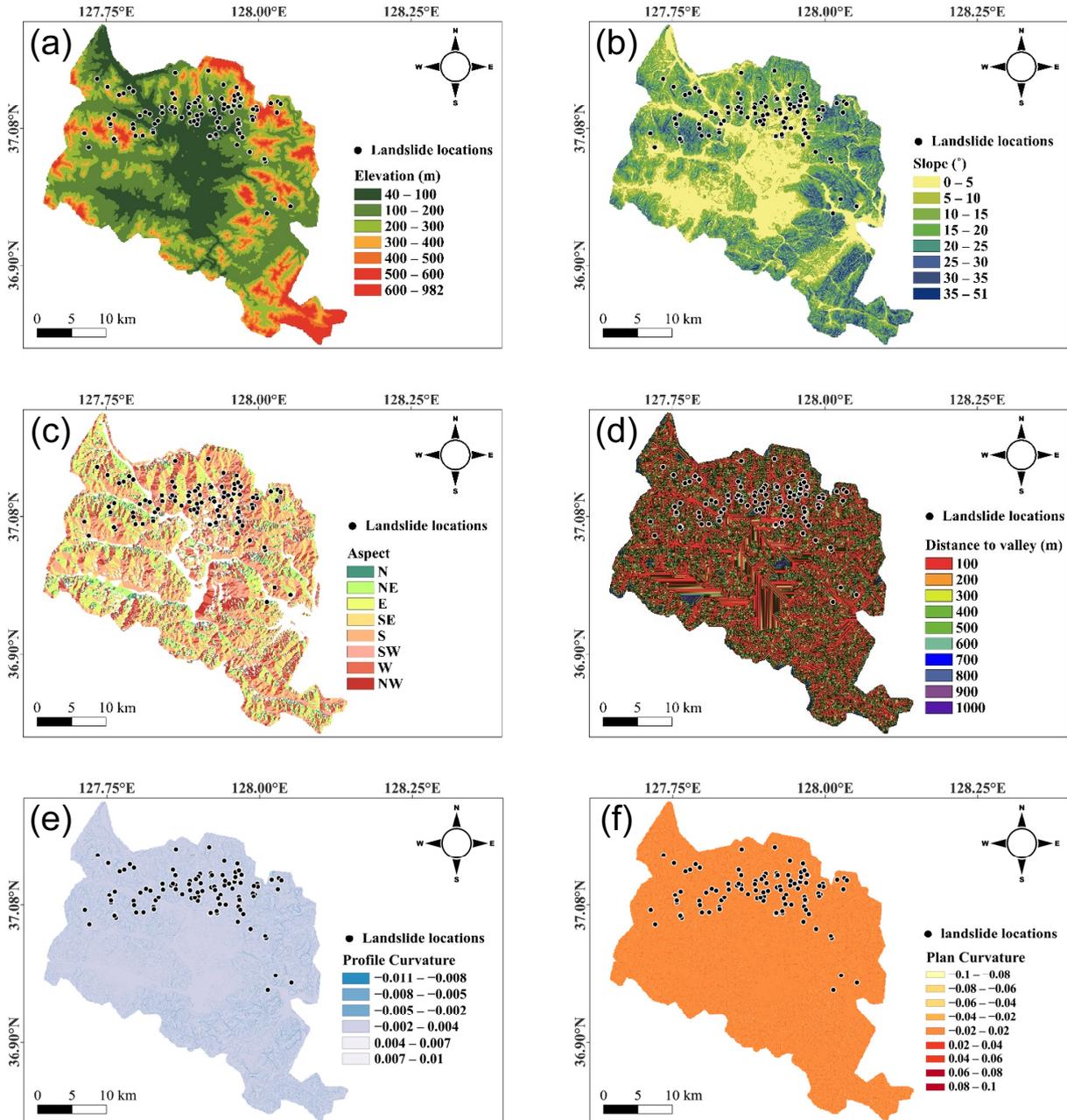


Fig. 4. Maps of influencing factors used for analysis of the importance of variables to landslide damage: (a) elevation, (b) slope, (c) aspect, (d) distance to valley, (e) plan curvature, (f) profile curvature, (g) TWI, (h) LS-factor, (i) lithology, and (j) forest cover.

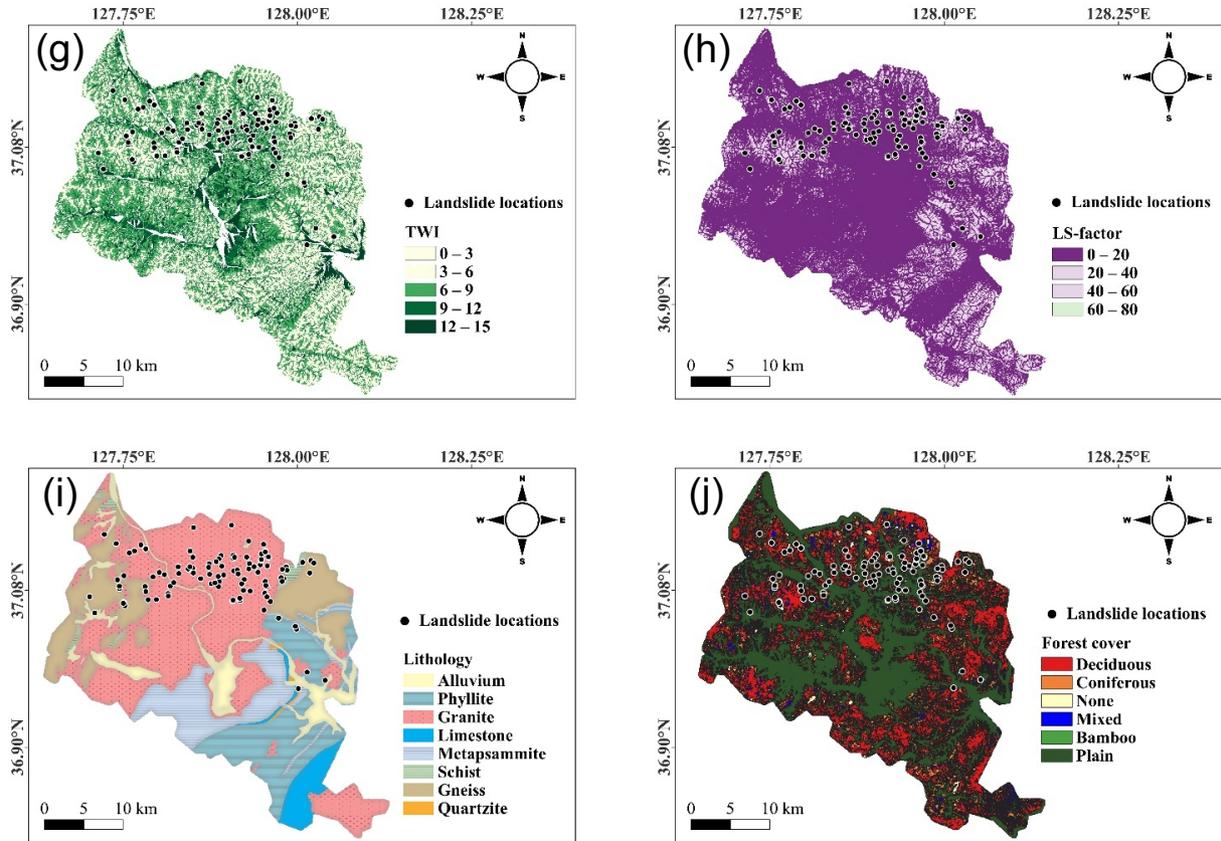


Fig. 4. Continued.

방법

빈도비율(Frequency Ratio, FR)

데이터 기반의 모델은 과거 산사태 발생 정보와 영향 인자 사이의 상관관계를 분석하여 산사태를 예측하며, 이는 과거에 발생한 산사태와 유사한 상황이나 조건에서 산사태가 발생할 것이라는 가정에 기초하고 있다. 이와 마찬가지로 산사태 피해면적과 영향 인자 사이의 상관관계를 분석하기 위해 과거 산사태 피해와 영향 인자들 간의 상관관계를 정확히 분석하는 것이 중요하며, 연속형 데이터의 범주화를 정량적으로 표현하기 위해 빈도비율(Frequency Ratio, FR)을 사용하였다. 빈도비율의 계산식은 다음 식 (1)과 같다.

$$FR = \frac{N_i / TN}{A_i / TA} \quad (1)$$

여기서, N_i 는 특정 범주에서의 산사태의 피해 면적을 나타내고 TN 은 연구 지역 전체에서의 산사태 총 피해 면적을 의미한다. 그리고 A_i 는 특정 범주의 총 면적을 의미하고 TA 는 연구 지역에서의 총면적을 의미한다. 빈도비율이 1보다 낮으면 산사태 피해와 주어진 인자의 속성 간의 관계가 약하다고 판단할 수 있으며(Lee and Talib, 2005), 1보다 크면 관계가 강하다고 해석된다.

랜덤포레스트(Random forest)

랜덤포레스트는 결정나무(decision tree) 기반으로 하는 앙상블 모델로서, 다른 성능을 가진 여러 개의 결정 나무를 결합한 모델이다(Breiman, 2001). 랜덤포레스트는 배깅(bagging) 알고리즘을 제한적으로 적용하여 축소된 특성을 가진 결정나무를 만들고 이를 병렬로 병합하여 하나의 강한 학습자를 형성할 수 있다(Appiahene et al., 2020). 동시에 알고리즘 내 제한적 배깅 알고리즘이 적용되지 않은 훈련 데이터인 Out-of-bag(OOB)을 사용하여 변수 중요도를 결정할 수 있다(Breiman, 2001). 이러한 특징을 가지는 랜덤포레스트는 노이즈(noise)에 대해 견고할 뿐만 아니라 병렬구조를 가지기 때문에 계산속도가 빠르고 의사결정나무의 단점인 과적합 문제를 해결할 수 있다는 장점을 가진다(Breiman, 2001). 또한 목표변수의 타입에 상관없이 적용할 수 있고 타 분류모델 대비 높은 예측 성능을 보이기 때문에(Goetz et al., 2015; Hong et al., 2015) 랜덤 포레스트를 본 연구의 분석모델로 채택하였다.

모델링 방법

연구 지역의 각 인자별 주제도를 구축하기 위해서 국토지리정보원에서 제공하는 90 m × 90 m 해상도의 수치표고모델

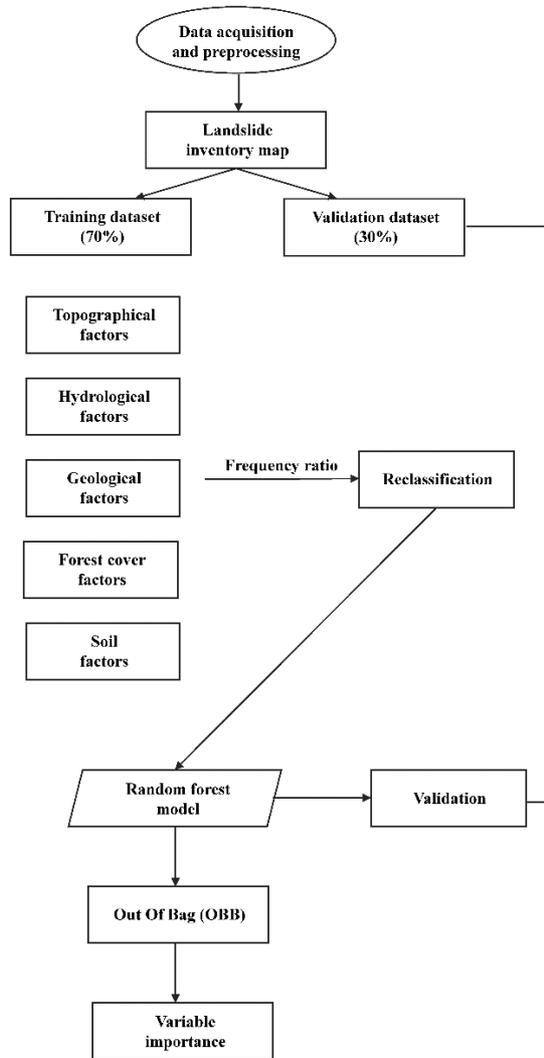


Fig. 5. Flow chart of the study.

(DEM)을 획득한 후, 고도 값에 해당하는 등고선 레이어를 추출하여 TIN(Triangulated Irregular Network) 보간을 실행하여 10 m × 10 m 해상도의 래스터 파일형태의 DEM으로 재구축하였다. 이를 기반으로 지형과 수문인자의 속성정보를 추출하였다. 지형인자와 수문인자를 범주화 시키기 빈도비율(FR)을 사용하여 정량화 하였으며, 피해면적과 인자 속성정보 사이의 관계를 분석하였다(Fig. 4). 본 연구에서는 종속변수의 이진분류시 데이터의 불균형을 고려하여 피해면적 자료의 분류비율이 1 대 1이 되는 0.5 ha를 기준으로 0.5 ha 이상일 때 1로 0.5 ha 이하일 때는 0으로 표현하였다. 전처리된 인자의 속성정보를 랜덤포레스트의 입력데이터로 사용하였고 변수 중요도는 훈련데이터에서 OOB 활용하여 이전 모델의 성능과 i번째의 변수의 데이터를 무작위로 섞은 후 예측성능의 차이를 이용하여 변수 중요도를 계산하였다. 본 연구의 전반적인 데이터 구축 및 분석과정은 흐름도(Fig. 5)에 제시되어 있다.

성능평가

랜덤포레스트의 예측 결과의 정확성을 파악하기 위해 ROC(Receiver Operation Characteristics) 그래프를 활용하였다. ROC 그래프는 혼동행렬(confusion matrix)을 기초로 계산된 민감도(sensitivity)와 특이도(specificity)를 이용하여 그래프로 나타내며, 그래프 곡선의 하단부 면적(Area Under Curve, AUC)을 계산하여 예측 성능의 정확성을 계산할 수 있다 (Frattini et al., 2010). 이를 바탕으로 왼쪽 상단에 가까울수록 예측의 정확성을 의미하는 AUC 값의 크기가 커지며, 이는 모델의 예측 정확성이 높은 것으로 해석한다(Frattini et al., 2010).

결과

최종적으로 도출된 각 인자의 분류체계와 각 카테고리의 빈도비율은 Table 2에 나와있으며, 산사태 피해에 대한 인자 내에서 어떤 구간(category)이 산사태 피해에 영향을 미치는 지 확인할 수 있다. 고도인자에서는 대부분 200~400 m의 항목에서 산사태 피해가 분포하였으며, 400 m 이상의 고도에서는 산사태 피해가 상대적으로 적었다. 경사인자에서는 경사가 높아질수록 빈도비율이 점차적으로 증가하는 양상을 보이며, 사면방향은 북서와 남동 방향이 가장 우세하게 나타난다. 수직곡률은 -0.011에서 -0.008일 때 빈도비율의 값이 가장 크게 나타났는데, 이는 수직곡률이 오목할수록 사면물질의 가속에 영향을 미치게 되고 결과적으로 사면물질의 침식과 퇴적의 양을 증가시켜 산사태 피해에 긍정적인 영향을 미친 것으로 해석된다.

Table 2. Frequency ratio (FR) of each category within the influencing factors (cells in the total study area: 10,692,363; cells in the total landslide damage area: 8,269)

Factor	Category	Number of cells in the study area (m ²)	Number of cells in the landslide damage area (m ²)	Number of landslides	FR
Elevation (m)	40~100	1,913,520	484	10	0.33
	100~200	3,811,858	2,398	41	0.81
	200~300	2,311,714	3,729	41	2.1
	300~400	1,315,278	1,320	9	1.3
	400~500	724,427	338	3	0.60
	500~982	615,566	0	0	0
Slope (°)	0~5	291,838,700	575	7	0.25
	5~10	172,372,300	1,089	21	0.82

Table 2. Continued

Factor	Category	Number of cells in the study area (m ²)	Number of cells in the landslide damage area (m ²)	Number of landslides	FR
Slope (°)	10~15	171,473,700	2,001	25	1.51
	15~20	167,944,400	951	16	0.73
	20~25	137,212,700	1,869	19	1.76
	25~30	83,076,300	13,200	11	2.05
	30~35	25,321,400	464	5	2.37
	35~51	199,968	0	0	0
Aspect	N	1,133,165	174	4	0.20
	NE	1,229,060	857	10	0.90
	E	1,158,654	1,195	18	1.33
	SE	1,117,052	1,520	12	1.76
	S	1,073,407	823	13	0.99
	SW	1,213,728	702	8	0.75
	W	3,010,994	1,345	17	0.58
	NW	1,240,257	1,653	22	1.72
Plan curvature	-0.1 to -0.08	58,540	0	0	0
	-0.08 to -0.06	35,679	45	1	1.63
	-0.06 to -0.04	88,159	38	1	0.56
	-0.04 to -0.02	247,536	261	3	1.36
	-0.02 to 0.02	9,826,400	7,283	92	0.96
	0.02 to 0.04	284,202	155	3	0.71
	0.04 to 0.06	95,867	418	4	5.64
	0.06 to 0.08	38,408	35	1	1.18
	0.08 to 0.1	17,572	34	1	2.50
Profile curvature	-0.011 to -0.008	87,750	222	4	3.27
	-0.008 to -0.005	231,161	385	6	3.11
	-0.005 to -0.002	744,602	485	7	0.84
	-0.002 to 0.004	9,116,269	7,030	90	1.00
	0.004 to 0.007	240,415	35	1	0.19
	0.007 to 0.01	95,815	112	1	1.51
Distance to valley (m)	0~100	3,677,639	1,632	17	0.57
	100~300	4,904,205	4,444	67	1.17
	300~400	1,292,615	1,641	15	1.6
	400~500	543,617	481	4	1.14
	500~600	163,411	71	1	0.56
	600~1,000	110,877	0	0	0
Forest cover	Deciduous	3,830,686	3,622	37	1.22
	Coniferous	1,648,382	2,316	30	1.82
	Mixed	682,467	536	5	1.02
	None	292,144	1,157	18	5.12

Table 2. Continued

Factor	Category	Number of cells in the study area (m ²)	Number of cells in the landslide damage area (m ²)	Number of landslides	FR
Forest cover	Bamboo	93	0	0	0
	Plain	4,238,591	0	0	0
Lithology	Granite	4,886,911	6,228	82	1.65
	Schist	116,549	864	5	9.59
	Gneiss	1,699,695	1,027	13	0.78
TWI	0~3	0	0	0	-
	3~6	4,130,694	4,962	63	1.55
	6~9	4,135,110	2,626	39	0.82
	9~12	977,274	681	5	0.90
	12~15	0	0	0	-
LS-factor	0~10	5,862,617	2,584	46	0.57
	10~20	3,022,383	3,183	37	1.36
	20~30	1,327,738	1,689	18	1.64
	30~40	331,451	490	6	1.91
	40~50	83,910	323	1	4.98

산사태 피해에 미치는 영향 인자의 중요도는 Fig. 6에서 확인할 수 있다. 14개의 인자 중 5개의 인자가 상대적으로 높은 기여를 하였으며, 이는 사면방향(IM = 0.17), 경사(IM = 0.15), 계곡까지의 직선 거리(IM = 0.12), 고도(IM = 0.11), 임상(IM = 0.10)으로 나타났다. 충주시의 산사태 피해면적은 전반적으로 지형인자가 지질인자 및 토양인자보다 더 중요하다는 것으로 나타났다. 본 연구에 사용된 모델의 성능은 AUC = 0.87의 정확도를 보여 신뢰성이 높은 것으로 해석된다 (Fig. 7).

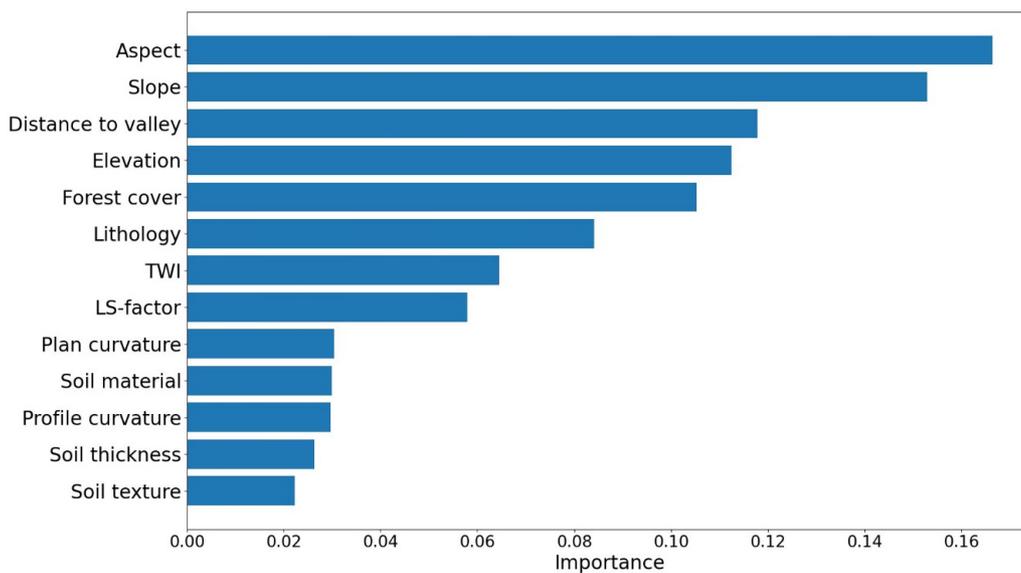


Fig. 6. Variable importance calculated by the random forest model.

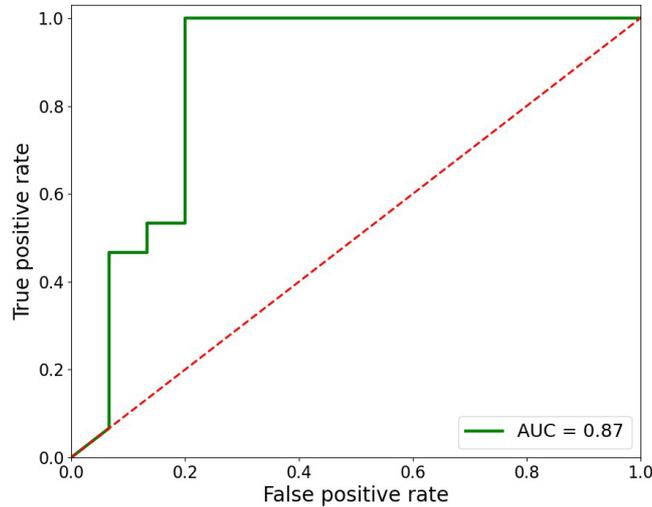


Fig. 7. ROC curve of the random forest model.

토의

우리나라에서 발생하는 산사태는 대부분 7월과 8월에 집중호우에 의해 발생하는 토석류에 해당하며(Chae et al., 2006), 이러한 산사태는 특정지역 내에서 토양, 암종, 지형조건 및 사면의 기하학적 형태의 유사성에도 불구하고 상이한 피해면적을 도출되기 때문에, 산사태 피해에 어떠한 변수가 피해 면적에 영향을 미치는지를 토석류의 확산(run out)과 연관지어 설명해 보려고 한다. 산사태 피해에 관련된 변수 중요도를 분석한 결과 경사방향(aspect)이 가장 중요한 인자로 나타났다. 이는 연구지역에서의 0.5 ha 이상의 피해를 입은 지역의 사면방향이 주로 남동쪽과 북서쪽 방향에 위치해 있으며 해당 사면의 계곡부 방향이 민가나 경작지에 집중되어 있어 피해면적이 크게 산출되었기 때문에 중요도에 영향을 미친 것으로 추정된다.

사면 경사(slope) 또한, 산사태 피해 증가에 영향을 주는 인자로 나타났으며, 이는 사면 경사가 토석류의 유동속도에 영향을 미치기 때문으로 판단된다. 선행연구에 따르면 사면 경사의 증가에 따라 토석류의 유동속도가 증가하는 현상이 발생하는데(Iverson et al., 2010), 이는 사면 경사가 가파를수록 토석류의 관성력이 증가하기 때문이다(Eu and Im, 2017). 따라서 사면 경사가 급해질수록 토석류의 운동에너지의 양이 증가하기 때문에 더욱 큰 산사태 피해가 나타날 수 있다.

토석류는 발생 형태에 따라 사면형 토석류와 계곡형 토석류로 구분되며 이에 따라 토석류의 피해규모가 달라지기 때문에 산사태 피해에 있어 계곡까지의 직선 거리는 중요한 인자로 작용한다. 사면형 토석류는 경사면을 따라 이동하며, 상대적으로 계곡형 토석류보다 유출거리(runout distance)가 짧아 피해면적이 작게 나타난다(Kim, 2008). 반면에 계곡형 토석류는 계곡을 따라 흘러내려가기 때문에 유출거리가 길고 이동할 때 토양과 암석을 침식시키고 주변을 휩쓸면서 내려가 새로운 산사태를 발생시키거나 기존의 산사태를 더욱 발달시키기도 한다. 이는 계곡의 존재는 산사태 피해를 증가시킬 수 있는 요인으로 작용된다는 것을 알 수 있다.

마지막으로 고도 또한 산사태 피해와 높은 연관성을 보였는데, 이는 토석류의 유출거리와 관련이 있는 것으로 보인다. 토석류의 유출거리를 예측하는 모델은 경험적 모델과 수치적 모델이 존재하며, 일반적으로 사용되는 경험적 모델은 도달각(reach angle)에 따라 토석류의 유출거리를 예측할 수 있다. 도달각은 발생지점과 산사태 퇴적물질의 끝단을 연결하는 선분과 수평선분의 각도를 의미하며, 발생지점과 퇴적지점사이의 높이차와 거리를 기반으로 계산할 수 있다. 우리나라에

서도 Choi and Baek(2012)은 2009년에서 2011년까지의 우리나라 전국의 443개 구역의 토석류를 현장조사를 하였고, 그 데이터를 Sled 모형에 적용하여 도출한 경험식을 식 (2)와 같이 제안하였다.

$$\frac{1}{R} = \frac{L}{H} \quad (2)$$

여기서, $1/R$ 은 순효율(net efficiency), L 은 토석류의 수평이동거리, H 는 토석류의 발생지점과 퇴적지점과의 표고차를 의미하며, Choi and Baek(2012)의 연구 결과에 따르면, 우리나라의 순효율은 4.3으로 나타났다. 식 (2)에서 순효율이 4.3으로 고정되어 있다고 가정하면 고도차가 증가할수록 그에 따른 수평이동거리가 증가하는 양상을 보이며, 또한 수평이동거리가 증가할수록 토석류의 부피가 증가하는 경향을 보인다(Kim, 2008). 이는 고도가 증가함에 따라 토석량이 증가하여 더 많은 산사태 피해를 야기시킬 수 있는 것으로 해석된다.

결론

우리나라의 산사태는 주로 여름철인 7월과 8월 사이에 집중호우로 인해 발생하여, 매년 심각한 인명 재산 피해를 야기하는 자연재해 중 하나이다. 최근 온난화로 인한 기후변화의 영향으로 산사태 피해가 점차 증가하는 양상을 보이기 때문에 산사태 피해를 최소화시키기 위해 산사태 피해를 증가시키는 인자를 파악하는 것이 중요하다. 대부분의 기존 연구는 산사태 발생 예측에 중점을 두고 연구를 수행하였으나, 본 연구는 산사태 피해에 초점을 맞추어 산사태 피해인자 간의 변수 중요도를 평가하는 것을 목적으로 산사태 피해면적, 고도, 사면경사 등 14개의 인자를 대상으로 랜덤포레스트를 활용하여 충청북도 충주에서 발생한 104개의 산사태 현황자료를 분석하였다.

각 변수들의 빈도비율을 계산한 결과 사면방향이 주로 북서, 남동일 때 산사태 피해에 더 큰 영향을 미치며, 고도는 200 m에서 400 m의 범위 일 때 그리고 경사가 증가할수록 산사태 피해가 증가하는 경향을 보인다. 랜덤포레스트 분석 결과, AUC는 0.87로 높은 예측 성능을 보였으며, 변수 중요도의 순위는 사면방향이 북동쪽과 남동쪽 일 때, 경사가 증가할수록, 계곡까지의 거리가 짧을수록 산사태 피해에 더 큰 영향을 미치는 것으로 나타난다. 이와 같이 피해에 높은 영향을 주는 인자들이 토석류의 유출거리를 증가시키는 인자로 작용하기 때문에 다른 인자들에 비해 산사태 피해와 범위에 영향을 주는 것으로 판단된다.

이러한 연구 결과는 산사태 피해 예측지도 제작의 기초자료 활용될 수 있을 것이며, 도시계획 및 정책 결정에 있어 유용한 참고자료가 될 수 있다.

References

- Appiahene, P., Missah, Y.M., Najim, U., 2020, Predicting bank operational efficiency using machine learning algorithm: Comparative study of decision tree, random forest, and neural networks, *Advances in Fuzzy Systems*, 2020, 8581202.
- Breiman, L., 2001, Random forests, *Machine Learning*, 45(1), 5-32.
- Cellek, S., 2021, The effect of aspect on landslide and its relationship with other parameters, *Landslides*, IntechOpen, 13-29.
- Chae, B.G., Kim, W.Y., Jo, Y.C., Kim, K.S., Lee, C.O., Song, Y.S., 2006, Field investigation and prediction techniques of landslides, *Proceedings of the Korean Society of Geological Engineering Symposium: Landslides and Disaster Prevention Measurement*, Seoul, 149-184.

- Choi, D.Y., Baek, J.C., 2012, Characteristics of runout distance of debris flows in Korea, *Journal of the Korean Society of Civil Engineers*, 32(3B), 193-201 (in Korean with English abstract).
- Du, G.L., Zhang, Y.S., Iqbal, J., Yang, Z.H., Yao, X., 2017, Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu Province, China, *Journal of Mountain Science*, 14(2), 249-268.
- Eu, S., Im, S.J., 2017, Examining velocity estimation equations of debris flow using small-scaled flume experiments, *Journal of Korean Forest Society*, 106(4), 424-430 (in Korean with English abstract).
- Fadhillah, M.F., Hakim, W.L., Panahi, M., Rezaie, F., Lee, C.W., Lee, S., 2022, Mapping of landslide potential in Pyeongchang-gun, South Korea, using machine learning meta-based optimization algorithms, *Egyptian Journal of Remote Sensing and Space Science*, 25(2), 463-472.
- Frattoni, P., Crosta, G., Carrara, A., 2010, Techniques for evaluating the performance of landslide susceptibility models, *Engineering Geology*, 111(1), 62-72.
- Gaidzik, K., Ramírez-Herrera, M.T., 2021, The importance of input data on landslide susceptibility mapping, *Scientific Reports*, 11(1), 19334.
- Goetz, J.N., Brenning, A., Petschko, H., Leopold, P., 2015, Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling, *Computers & Geosciences*, 81, 1-11.
- Guo, Z., Shi, Y., Huang, F., Fan, X., Huang, J., 2021, Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management, *Geoscience Frontiers*, 12(6), 101249.
- Hong, H., Pradhan, B., Xu, C., Tien Bui, D., 2015, Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines, *Catena*, 133, 266-281.
- Iverson, R.M., Logan, M., LaHusen, R.G., Berti, M., 2010, The perfect debris flow? Aggregated results from 28 large-scale experiments, *Journal of Geophysical Research*, 115, F03005.
- Jeong, C.H., Park, Y.A., Kim, H.M., 1977, Geological report of Eumseong sheet (1:50,000), Geological Survey of Korea, <https://doi.org/10.22747/data.20211022.4507>.
- Ji, J.M., Yoon, S., Lee, C.J., 1989, Geological report of Munmak sheet (1:50,000), Geological Survey of Korea, <https://doi.org/10.22747/data.20211221.4679>.
- Kim, G.W., Lee, H.G., 1965, Geological report of Chungju sheet (1:50,000), Geological Survey of Korea, <https://doi.org/10.22747/data.20211022.4505>.
- Kim, G.W., Park, B.S., Lee, H.G., 1967a, Geological report of Jecheon sheet (1:50,000), Geological Survey of Korea, <http://doi.org/10.22747/data.20211203.4611>.
- Kim, K.S., 2008, Characteristics of basin topography and rainfall triggering debris flow, *Journal of the Korean Society of Civil Engineers*, 28(5C), 263-271 (in Korean with English abstract).
- Kim, N.J., Choi, S.O., Kang, P.J., 1967b, Geological report of Mungyeong sheet (1:50,000), Geological Survey of Korea, <https://doi.org/10.22747/data.20211214.4646>.
- KMA (Korea Meteorological Administration), 2020, Landslide damage status of Chungju city, Chungcheongbuk-do, 2020, Retrieved from <https://www.data.go.kr/data/15102424/fileData.do>.
- KFS (Korea Forest Service), 2021, 2020 Forest disaster white paper, 94p.
- Lee, J.H., 2022, Landslide susceptibility assessment using coupled initiation and runout prediction model, Doctoral Dissertation, Sejong University Graduate School, 2-4 (in Korean with English abstract).
- Lee, J.H., Kim, J.H., 1972, Geological report of Goesan sheet (1:50,000), Geological Survey of Korea, <https://doi.org/10.22747/data.20211214.4653>.
- Lee, M.S., Park, B.S., 1965, Geological report of Hwanggang-ri sheet (1:50,000), Geological Survey of Korea, <https://doi.org/10.22747/data.20211110.4596>.

- Lee, S., Kim, M., 2020, Construction of topographic/hydrologic data using DEM and its service, *GEO DATA*, 2(2), 36-44 (in Korean with English abstract).
- Lee, S., Talib, J.A., 2005, Probabilistic landslide susceptibility and factor effect analysis, *Environmental Geology*, 47(7), 982-990.
- Lombardo, L., Mai, P.M., 2018, Presenting logistic regression-based landslide susceptibility results, *Engineering Geology*, 244, 14-24.
- Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R., Abderrahmane, B., 2020, Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance, *Earth-Science Reviews*, 207, 103225.
- Park, B.S., Yeo, S.C., 1971, Geological report of Mokgye sheet (1:50,000), Geological Survey of Korea, <https://doi.org/10.22747/data.20211203.4618>.
- Park, H.J., Lee, J.H., 2022, A review of quantitative landslide susceptibility analysis methods using physically based modelling, *The Journal of Engineering Geology*, 32(1), 27-40 (in Korean with English abstract).
- Pham, B.T., Pradhan, B., Tien Bui, D., Prakash, I., Dholakia, M.B., 2016, A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India), *Environmental Modelling & Software*, 84, 240-250.
- Pham, B.T., Prakash, I., Khosravi, K., Chapi, K., Trinh, P.T., Ngo, T.Q., Hosseini, S.V., Bui, D.T., 2019, A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling, *Geocarto International*, 34(13), 1385-1407.
- Pradhan, B., 2013, A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS, *Computers & Geosciences*, 51, 350-365.
- Regmi, A.D., Yoshida, K., Nagata, H., Pradhan, A.M.S., Pradhan, B., Pourghasemi, H.R., 2013, The relationship between geology and rock weathering on the rock instability along Mugling-Narayanghat road corridor, Central Nepal Himalaya, *Natural Hazards*, 66(2), 501-532.
- Roth, R.A., 1983, Factors affecting landslide-susceptibility in San Mateo County, California, *Environmental & Engineering Geoscience*, xx(4), 353-372.
- Shahabi, H., Hashim, M., 2015, Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment, *Scientific Reports*, 5, 9899.
- Shin, W.J., Hong, S.C., Kim, J.Y., 2022, A study on the factors affecting debris disasters in Chungbuk province, *Journal of the Association of Korean Geographers*, 11(1), 105-120 (in Korean with English abstract).
- Song, Y.S., Lee, M.S., 2023, A random walk model for estimating debris flow damage range, *The Journal of Engineering Geology*, 33(1), 201-211 (in Korean with English abstract).
- Van Westen, C.J., Seijmonsbergen, A.C., Mantovani, F., 1999, Comparing landslide hazard maps, *Natural Hazards*, 20(2), 137-158.
- Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H., Hong, H., 2020, Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping, *Catena*, 187, 104396.
- Yang, I.T., Park, J.K., Park, K., 2014, An evaluation of damage scale on the local governments in Gangwon-do using landslide risk maps, *Journal of the Korean Society for Geospatial Information Science*, 22(4), 71-80 (in Korean with English abstract).
- Yeo, S.C., Lee, I.G., 1975, Geological report of Yeosu sheet (1:50,000), Geological Survey of Korea, <https://doi.org/10.22747/data.20211104.4577>.
- Yesilnacar, E., Topal, T., 2005, Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey), *Engineering Geology*, 79(3-4), 251-266.