

AutoML을 이용한 산사태 예측 및 변수 중요도 산정

남경훈¹ · 김만일² · 권오일³ · 왕파우⁴ · 정교철^{5*}

¹일본 시마네대학 지구과학 및 지구환경과학과 박사후 연구원, ²산림조합중앙회 산림공학연구소 책임연구원, ³한국건설기술연구원 연구위원, ⁴중국 통지대학 토목공학과 교수, ⁵안동대학교 지구환경과학과 교수

Prediction of Landslides and Determination of Its Variable Importance Using AutoML

KoungHoon Nam¹ · Man-Il Kim² · Oil Kwon³ · Fawu Wang⁴ · Gyo-Cheol Jeong^{5*}

¹Post Dr. Researcher, Geoscience and Geoenvironmental Science, Shimane University, Japan

²Principal Researcher, Forest Engineering Research Institute, National Forestry Cooperative Federation

³Research Fellow, Road Infrastructure Project Team, Korea Institute of Civil Engineering and Building Technology

⁴Professor, Department of Civil Engineering, Tongji University, China

⁵Professor, Department of Earth and Environmental Sciences, Andong National University

Abstract

This study was performed to develop a model to predict landslides and determine the variable importance of landslides susceptibility factors based on the probabilistic prediction of landslides occurring on slopes along the road. Field survey data of 30,615 slopes from 2007 to 2020 in Korea were analyzed to develop a landslide prediction model. Of the total 131 variable factors, 17 topographic factors and 114 geological factors (including 89 bedrocks) were used to predict landslides. Automated machine learning (AutoML) was used to classify landslides and non-landslides. The verification results revealed that the best model, an extremely randomized tree (XRT) with excellent predictive performance, yielded 83.977% of prediction rates on test data. As a result of the analysis to determine the variable importance of the landslide susceptibility factors, it was composed of 10 topographic factors and 9 geological factors, which was presented as a percentage for each factor. This model was evaluated probabilistically and quantitatively for the likelihood of landslide occurrence by deriving the ranking of variable importance using only on-site survey data. It is considered that this model can provide a reliable basis for slope safety assessment through field surveys to decision-makers in the future.

Keywords: AutoML, XRT, landslide prediction model, probabilistic prediction, variable importance

초 록

이 연구는 도로 비탈면에서 발생하는 산사태의 확률론적 예측에 기반된 산사태 발생에 영향을 미치는 인자의 중요도 산정 및 예측 모델을 개발하는 것이다. 산사태 예측 모델을 개발하기 위해 한반도 전 지역을 대상으로 2007년부터 2020년까지 조사된 30,615사면의 현장조사 자료를 활용하였다. 전체 131개의 변수 인자 중 지형인자 17개, 지질인자 114개(기반암 89개를 포함), 도로와의 이격거리를 사용하였다. 산사태 발생에 영향을 미치는 인자를 자동화된 머신러닝인 AutoML을 실시하여 예측 성능이 뛰어난 XRT (extremely randomized trees)를 선정하였다. 변수 중요도 분석결과 지형적 요인 10개, 지질인자 9개, 사

OPEN ACCESS

*Corresponding author: Gyo-Cheol Jeong
E-mail: jeong@anu.ac.kr

Received: 26 August, 2020

Revised: 17 September, 2020

Accepted: 21 September, 2020

© 2020 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.

회적 영향성인 도로와의 이격 거리와 관련된 항목순으로 급경사지 불안정에 가장 많은 영향을 주는 것으로 분석되었다. 개발된 모델의 신뢰성 검증을 수행한 결과 AUC 83.977%의 예측율을 확보한 것으로 나타났다. 이 모델은 산사태 이력을 기반으로 한 현장조사 자료만을 이용하여 변수 중요도의 순위를 도출함으로써 그에 따른 산사태 발생 가능성을 확률적 및 정량적으로 평가하였다. 향후 의사 결정자들에게 현장조사를 통한 사면진단 안전평가 시 신뢰성 있는 근거를 제공하리라 판단된다.

주요어: AutoML, XRT, 산사태 예측 모델, 확률론적 예측, 변수 중요도

서론

확률론적 분석에 의한 통계적 방법은 산사태에 관련된 각종 자료를 획득하고 이를 확률론적인 이론을 도입하여 산사태 발생 유무를 예측하는 방법이다. 자연사면의 산사태 발생 여부는 일반적으로 지형학적, 지질학적 및 지반공학적 토질 물성에 따라 달라지게 된다. 이를 고려하여 산사태 발생에 대한 예측을 로지스틱 회귀분석(logistic regression, LR)(Chae et al., 2004)과 인공신경망(artificial neural network)(Hong et al., 2004; Byun and Moon, 2010)을 사용한 연구가 수행되었다. 또한 위성영상에서 추출한 다양한 Band math 기법을 적용한 산사태 취약성 맵핑이 연구되어 왔으며 예측모델 기법으로 최근 RF(random forest)와 GBM(gradient boosting machine)과 같은 앙상블(ensemble) 기반된 배깅(bagging) 및 부스팅(boosting) 계열의 모델이 강력한 예측 성능을 나타내는 것으로 보고되고 있다(Kim et al., 2017; Di et al., 2019; Park and Kim, 2019; Merghadi et al., 2020).

그러나 강력한 예측 성능의 모델을 사용하는 것을 차치하고 기존의 산사태 예측 모델 및 취약성 맵핑은 GIS에서 인공적 또는 무작위로 비산사태 지역을 생성하는 것에 한계점을 두고 있다(Nam and Wang, 2019). 즉, 산사태 지역과 비산사태 지역을 이진분류(binary classification) 하기에 앞서 비산사태 지역의 샘플링을 명확히 해야 한다. 현재까지 국내외의 경우 산사태를 예측하기 위하여 산사태 지역과 비산사태 지역을 구분하여 현장 조사한 빅데이터가 희귀함에 따라 그에 따른 예측 모델에 적용한 연구가 없는 상태이다.

자동화된 머신러닝인 AutoML은 머신러닝과 딥러닝 모델을 구축하는데 있어 기술력을 갖춘 데이터 과학자란 필요조건을 극복하기 위하여 개발되었으며 현존하는 대부분의 예측 모델을 적용해보고 최적의 모형을 선택한다(Nagarajah and Poravi, 2019). 또한 초매개변수(hyperparameter)를 최적화하기 위한 탐색방식으로 Grid Search와 Randomized Search를 지원한다. 모든 머신러닝 모델에는 매개변수인 모델의 각 변수와 특성에 대한 가중치가 있다. 매개변수는 일반적으로 확률적 기울기 하강(stochastic gradient descent)과 같은 옵티마이저의 통제 하에 반복과 오류의 역전파(back-propagation)에 의해 결정된다. 학습률, 배제율, 노드, 은닉층 및 랜덤 포레스트(random forest)의 트리 수와 같은 모델별 매개변수 등이 초매개변수에 포함되거나 영향을 받는다. 이를 최적화시키고 모델을 다시 학습시키는 것에 많은 연산 시간이 소비될 수 있다. 하지만 AutoML은 이러한 과정들을 자동화하여 가장 효율적인 모델을 연구자에게 제공한다. 이 연구는 국내외 산사태 예측 모델에 처음 사용하는 AutoML을 적용하여 현장에서 빠르게 평가하기 위한 방법을 소개하고자하며 정량적 산사태 예측모델 제안 및 발생 요인의 중요도를 산정하고자 한다.

연구지역

우리나라 전 범위를 대상으로 도로 비탈면 지역을 연구지역으로 선정하였다. 강원도(6,696개소), 경상북도(5,886개소), 전라남도(4,645개소), 경상남도(3,871개소), 경기도(2,574개소), 전라북도(2,569개소), 충청북도(2,329개소), 충청남도(1,933개소), 울산광역시(115개소), 대구광역시(56개소), 부산광역시(48개소), 인천광역시(20개소), 광주광역시(7

개소)를 대상으로 총 30,615개소의 도로 비탈면의 암반사면(14,114개소), 혼합사면(10,729개소), 토사사면(5,399개소), 자연사면(507개소)에서 산사태 발생지역(17,761개소)과 산사태 미발생지역(12,854개소)을 구분하여 조사한 현장 데이터를 사용하였다(Table 1). 한국건설기술연구원에서 2007년 10월부터 2020년 3월까지 조사한 현장 데이터를 이용하여 산사태 예측 모델링을 수행하였다.

Table 1. Number of landslides and non-landslides on slopes along the road investigated from 2007 to 2020 in Korea

Province	No. of landslides	Province	No. of non-landslides
Kangwon-do	4,753	Gyeongsangbuk-do	3,570
Jeollanam-do	2,870	Gyeongsangnam-do	1,960
Gyeongsangbuk-do	2,316	Kangwon-do	1,941
Gyeongsangnam-do	1,911	Jeollanam-do	1,774
Chungcheongbuk-do	1,699	Gyeonggi-do	1,103
Jeollabuk-do	1,651	Jeollabuk-do	918
Gyeonggi-do	1,348	Chungcheongnam-do	867
Chungcheongnam-do	1,060	Chungcheongbuk-do	628
Ulsan Metropolitan City	79	Ulsan Metropolitan City	36
Busan Metropolitan City	34	Daegu Metropolitan City	22
Daegu Metropolitan City	34	Incheon Metropolitan City	14
Incheon Metropolitan City	6	Busan Metropolitan City	14
Gwangju Metropolitan City	0	Gwangju Metropolitan City	7
Sum	17,761	Sum	12,854

산사태를 발생 시키는 요인

한국건설기술연구원에서 도로 비탈면의 사면안정성 평가를 위한 현장조사 시 사용한 데이터는 (1) 지형, (2) 지질, (3) 수리, (4) 상부자연사면, (5) 상태, (6) 보호시설, (7) 사회적 영향성, (8) 기타 항목으로 크게 8가지로 구분해서 조사하였다. 결측치가 많은 항목들을 제외하고 사용한 데이터는 각각의 항목에서 세부적으로 기입된 사면길이, 사면최대높이, 경사각도, 경사구배, 상부경사, 이격거리, 사면종류, 주변지형, 지하수, 풍화도, 불연속면방향성, 사면형상, 측면형상, 붕괴이력, 토층심도, 암반형태, 불연속면, 기반암, 도로와의 이격거리, 붕괴유형을 조사하였으며 이를 산사태 발생 인자로써 예측 모델 적용 시 변수로 활용하였다(Table 2).

Table 2. Description of landslide susceptibility factors for input data used in AutoML modelling

Landslide susceptibility factors	Continuous data	Categorical data		
		Type	No. of type	
Geomorphological factor	Slope angle	30,615	-	-
	Upper slope angle	30,615	-	-
	Slope height	30,615	-	-
	Slope length	30,615	-	-
	Slope gradient	30,615	-	-

Table 2. Continued

Landslide susceptibility factors		Continuous data	Categorical data	
			Type	No. of type
Geomorphological factor	Slope type	-	Rockmass slope	14,109
		-	Merged slope	10,614
		-	Soil slope	5,385
		-	Natural slope without retaining wall	507
	Slope shape	-	Straight	17,858
		-	Convex	7,460
		-	Concave	4,444
		-	Wave	853
	Slope side shape	-	Straight	18,698
		-	Lumpy	11,143
		-	Drop out	432
		-	Protuding	342
	Soil depth	30,615	-	-
	Bedrock	-	Granite	9,979
		-	Gneiss	7,631
-		Sandstone	2,651	
-		Tuff	2,396	
-		Shale	1,334	
-		
-		etc	6,624	
Groundwater	-	Damp	14,847	
	-	Complete dry	14,362	
	-	Wet	1,052	
	-	Dripping	218	
Geological factor	Rockmass type	-	Flowing	136
		-	Irregular	12,977
		-	Crushed	9,515
		-	Tabular	3,583
		-	Blocky	3,308
		-	Massive	891
	Weathering (Rating)	-	Columnar	341
		-	Severe (IV)	12,503
		-	Normal (III)	8,903
		-	Complete (V)	6,254
Discontinuity	-	Residual deposit (VI)	2,502	
	-	Slightly weathered (II)	436	
	-	Fresh (I)	17	
	-	Joint	15,178	
	-	Non discontinuity	10,822	
	-	Foliation	2,194	
	-	Stratification	2,053	
Social effect	Distance to road	-	Fault	213
		-	Dike	117
Landslide	Landslide inventory	-	Shear zone	28
		-	Crack	10
		30,615	-	-
		-	Landslides	17,761
		-	Non landslides	12,854

연구방법 및 토의

AutoML 모델링 및 학습

도로 비탈면의 사면 불안정에 영향을 미치는 요인들의 상관관계 분석을 위해 파이썬 오픈소스 라이브러리인 'h2o'(www.h2o.ai)의 'AutoML'을 Google platform의 Jupyter 환경인 Google colab(<https://colab.research.google.com>)에서 모델링을 수행하였다. 자동화된 머신러닝인 AutoML은 인공지능망을 기반으로 하는 DNN(deep neural network), 배깅(bagging) 계열의 DRF(distributed random forest)와 XRT(extremely randomized trees), 부스트(boost) 계열의 GBM(gradient boosting machines)과 XGBoost(extreme gradient boosting) 모델 등을 지원한다.

Table 2에서 보는 바와 같이 사면경사, 상부사면경사, 사면높이, 사면길이, 사면구배, 토층심도, 도로와의 이격거리인 연속형데이터(continuous data)를 제외하고 대부분의 데이터가 범주형 데이터(categorical data)이다. 범주형 데이터를 통계 분석에 용이하게 하기 위해 수치형 데이터로 인코딩 한 후 이진분류를 위해 산사태 발생지역을 1, 산사태 미발생지역을 0으로 변환하였다. 전체 30,615사면에서 21,431사면(70%)을 훈련데이터(training data)로 9,184사면(30%)을 검증데이터(test data)로 사용하였다. AutoML을 이용한 상위 30개 모델을 분석결과 AUC가 높을수록 예측 성능이 뛰어난 트리 기반된 앙상블 예측 모델들이 상위 순위에 선정되었다. AutoML을 이용한 자동화된 지도학습과정은 예측 기법의 각각의 파라미터도 자동으로 조정되어 연구자의 주관적인 판단을 배제할 수 있으며 그에 따른 연산 시간도 대폭 줄일 수 있다는 장점이 있다. 이 연구에서 구축된 트리 기반된 모델(XRT, GBM, RF, XGBoost)은 Table 3에 기술하였으며 인공지능망 기반된 모델(DNN)은 Table 4와 같다.

Table 3. Parameters of tree-based models automatically selected from the AutoML modelling

Model	No. of trees	Min. depth	Max. depth	Min. leaves	Max. leaves
XRT	49	20	20	2,387	3,588
GBM	75	13	13	104	855
RF	50	20	20	2,698	3,675
XGBoost	64	-	-	-	-

Table 4. Parameters of deep neural network automatically selected from the AutoML modelling

Model	Layer	Units	Type	Dropout	Mean weight	Mean bias
DNN	1	121	Input	20	-	-
	2	200	Rectifier dropout	40	-0.009	-3.466
	3	200	Rectifier dropout	40	-0.088	-0.082
	4	2	Softmax	-	-0.024	-0.012

AutoML에서 선정된 XRT 모델

Geurts et al.(2006)에 의해 제안된 Extremely Randomized Trees(XRT)는 회귀적으로 훈련된 Decision Tree(DT)의 앙상블이며 최종 모델은 대형 DT를 사용하여 구축된다. 각 모델은 전체 데이터 세트를 사용하여 트리를 구성하고 각 분할에 대해 정보 이득을 기반으로 최상의 컷 포인트를 찾는다. XRT 모델의 주요 개선점은 다음과 같다. (i) 컷 포인트를 사용하여 노드를 무작위로 분할하고 (ii) 복제본이 아닌 DT를 구축하는 데 총 학습 데이터 세트를 사용한다(Xia et al., 2015). 랜덤 포레스트 모델 및 다른 의사 결정 트리 모델에 대해 부트스트랩을 사용하여 생성한다.

Galelli and Castelletti(2013)에 따르면 XRT 모델은 독립형 DT 모델에 비해 몇 가지 장점은 다음과 같다. (i) 독립형 DT를 사용하여 발생하는 과적합을 방지하는 데 도움이 되고, (ii) 계산적으로 더 효율적이고, (iii) 완전히 무작위 앙상블 방법, (iv) 단일 모델의 앙상블 평균화를 통해 표준 의사 결정 트리 모델보다 더 높은 정확도를 보유하고, (v) 입력 및 분할 값이 무작위로 선택된다. XRT 모델의 우수성과 높은 정확성은 복잡한 산사태 취약성 인자들의 평가를 위한 모델링으로부터 입증되었다(Merghadi et al., 2020).

Table 5는 수치 속성(numerical attributes)에 대한 Extra-Trees의 분할 절차이며(Geurts et al., 2006). Fig. 1은 AutoML에서 선정된 XRT 모델의 단일 DT(max depth: 3)를 시각화 한 것이다. 지니 지수(gini index)는 특정 변수가 무작위로 선택될 때 잘못 분류될 가능성에 대하여 평가한다.

$$Gini = 1 - \sum_{i=1}^n (p_i)^2 \quad (1)$$

여기서, p_i 는 객체가 특정 클래스로 분류될 확률이다. 일반적으로 XRT 모델에서는 지니 지수가 가장 낮은 기능이 루트 노드(첫번째 노드)로 선택된다. 수치 속성에 대한 Extra-Trees 분할 절차는 두 개의 매개변수가 존재하며 각 노드에서 무작위로 선택된 속성 수 K 와 노드 분할을 위한 최소 샘플 크기인 n_{\min} 이다. 앙상블 모델을 생성하기 위해 전체 학습 샘플과 함께 모델링되며 트리의 예측은 분류 문제에서 산술 평균으로 최종 예측을 산출하기 위해 집계된다. K 는 속성 선택 프로세스의 강도, n_{\min} 은 평균 출력 노이즈 강도, 트리의 개수 M 은 앙상블 모델 집계의 분산 감소 강도를 결정한다. 이러한 매개변수는 교차검증으로 문제 특성에 맞게 조정된다.

Table 5. Extra-Trees splitting algorithm modified from Geurts et al. (2006)

Split a node (S)
Input : the local learning subset S corresponding to the node we want to split
Output : split $[a < a_c]$ or nothing
- If `Stop_split` (S) is TRUE then return nothing.
- Otherwise select K attributes $\{a_1, \dots, a_k\}$ among all non constant (in S) candidate attributes;
- Draw K splits $\{s_1, \dots, s_k\}$, where $s_i = \text{Pick_a_random_split}(S, a_i), \forall i = 1, \dots, K$;
- Return a split s_* such that $\text{Score}(s_*, S) = \max_{i=1, \dots, K} \text{Score}(s_i, S)$.

Pick a random split (S, a_i)
Input : a subset S and an attribute a
Output : a split
- Let a_{\max}^S and a_{\min}^S denote the maximal and minimal value of a in S ;
- Draw a random cut-point a_c uniformly in $[a_{\max}^S, a_{\min}^S]$;
- Return the split $[a < a_c]$

Stop_split (S)
Input : a subset S
Output : a boolean
- If $|S| < n_{\min}$, then return TRUE;
- If all attributes are constant in S , then return TRUE;
- If the output is constant in S , then return TRUE;
- Otherwise, return FALSE.

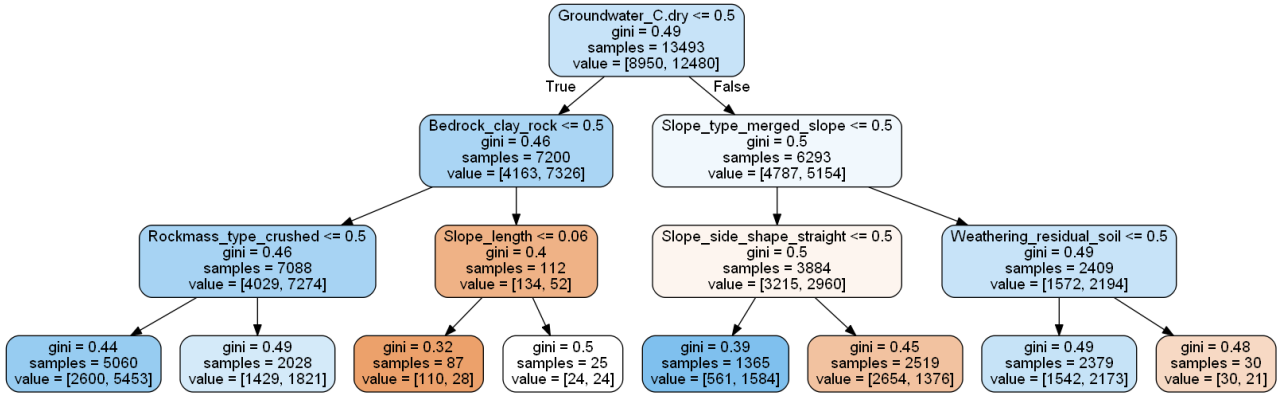


Fig. 1. Visualization of single decision tree obtained by the extremely randomized tree (e.g. max. depth: 3) used in this study.

AutoML 모델링에서 예측 성능이 가장 뛰어난 XRT 모델은 49개의 트리 수, 20개의 최소 심도, 2,387개의 최소 가지 수, 3,588개의 최대 가지 수가 사용되었으며 5번의 교차검증이 수행되었다.

XRT 모델 적합성 검증

평가항목에 대한 변수 중요도 산정에 대한 검증은 9,184(30%)개 사면에 대해 AutoML 모델에서 학습 과정에 사용되지 않은 검증데이터를 사용하여 분석하였다. 개발된 모델의 신뢰성 검증은 일반적으로 혼돈 행렬에 기반되어 분석하며(Fig. 2) 다변량 분석기법 중 하나로 널리 쓰이는 LR(Chae et al., 2004)도 함께 비교 및 검증하였다. 분석결과 AutoML 모델 중 예측 성능이 가장 좋은 XRT 모델은 AUC 83.977% 를 획득하였으며 다음으로 GBM: 83.969%, RF: 83.652%, XGBoost: 83.514%, DNN: 79.830%, LR: 76.6%의 예측율을 보이는 것으로 나타났다(Fig. 3). 검증데이터의 경우 전체자료에서 무작위로 선택되었고 AutoML 모델 구축 시 학습과정에 전혀 사용되지 않았음에도 불구하고 학습 자료의 경향을 반영하여 정확한 예측이 수행되었음을 나타낸다. 따라서 AutoML 모델링이 도로 비탈면의 산사태 취약성 요인의 상관관계를 규명 하는데 효과적이라 할 수 있다.

		True landslides		
		Landslides (Ls)	Non-Ls	
Predicted landslides	Landslides	TP (True Positive)	FP (False Positive)	Accuracy $= \frac{TP+TN}{TP+FP+TN+FN}$
	Non-Ls	FN (False Negative)	TN (True Negative)	Precision $= \frac{TP}{TP+FP}$
		TPR (True Positive Rate) $= \frac{TP}{TP+FN}$	FPR (False Positive Rate) $= \frac{FP}{FP+TN}$	TNR (True Negative Rate) $= \frac{TN}{TN+FP}$

Fig. 2. Confusion matrix for validation of landslide susceptibility assessment.

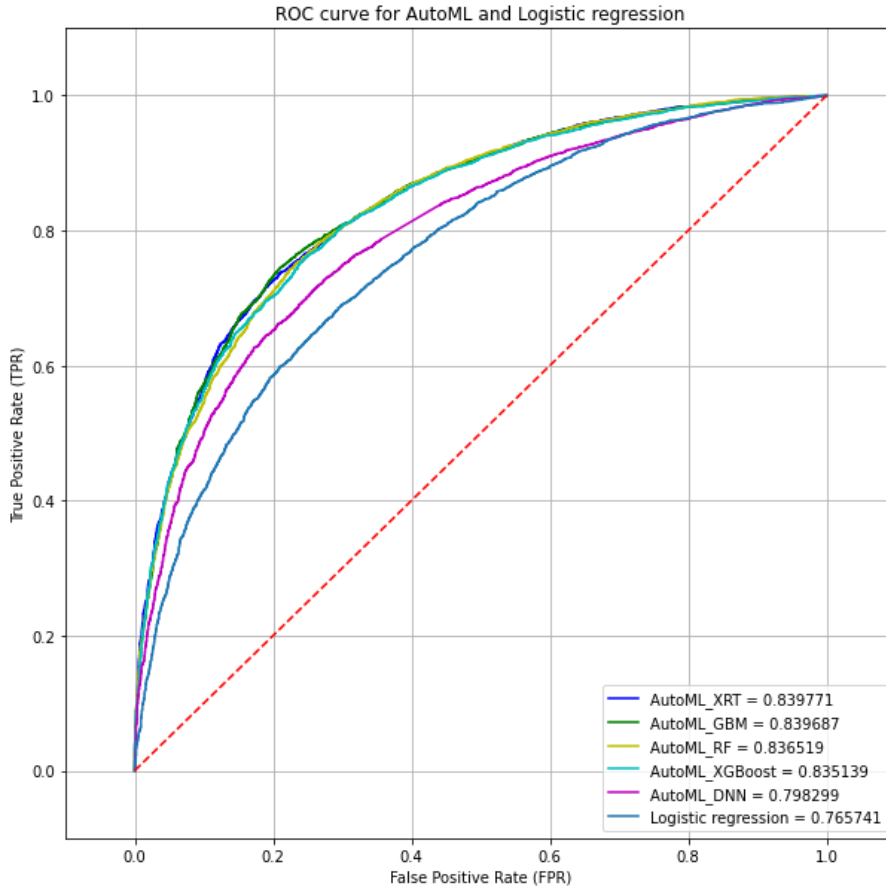


Fig. 3. ROC (receiver operating characteristic) curve for logistic regression and best model (XRT: 83.977%) selected from AutoML on test data.

변수 중요도 산정

산사태 발생 가능성 예측에 있어서 현장 조사 자료만으로 산사태 발생 가능성을 확률적 및 정량적으로 평가하기 위해서는 산사태 취약 변수의 중요도 산정이 신뢰성 있는 근거를 제공할 수 있을 것이다.

도로 비탈면 131개의 산사태 취약성 변수 중요도 분석 결과 토양심도(8.52%), 상부사면경사(8.21%), 사면경사(7.04%), 사면길이(6.84%), 최대사면높이(6.76%), 도로이격거리(6.17%), 사면측면형상-직선형(6.03%), 사면측면형상-요철형(5.71%), 사면구배(4.73%), 지하수-완전건조(2.21%), 암반유형-불규칙(1.80%), 기반암-편마암(1.75%), 기반암-화강암(1.53%), 불연속면-확인불가(1.53%), 사면유형-암반사면(1.52%), 사면유형-혼합사면(1.52%), 사면형상-직선형(1.48%), 지하수-축축한(1.44%), 풍화도-심각한(1.33%), 풍화도-보통(1.32%) 순으로 변수 중요도가 산정되었다(Fig. 4 and Table 6). 이러한 결과는 취약성 요인의 대분류 항목 중 지형적 요인 10개, 지질인자 9개, 사회적 영향성인 도로와의 이격 거리와 관련된 항목순으로 급경사지 불안정에 가장 많은 영향을 주는 것으로 분석되었다. 특히 토양심도는 산사태 발생지역의 근원지(source area)로서 토양심도에 따른 퇴적부(deposited area) 및 재해지역에 영향력이 크게 나타난다고 보고되었으며(Sassa et al., 2010; Dai et al., 2014), 이 연구에서도 산사태 취약성 변수 중 가장 중요도가 크게 산정되었다. 또한 상부사면경사는 사면경사, 사면길이, 최대사면높이보다 중요도가 높게 산정되었으며 이러한 결과는 도로 비탈면의 인위적 절취 사면(cut slope)에 따른 상부사면경사가 큰 영향력을 끼치는 것으로 판단된다.

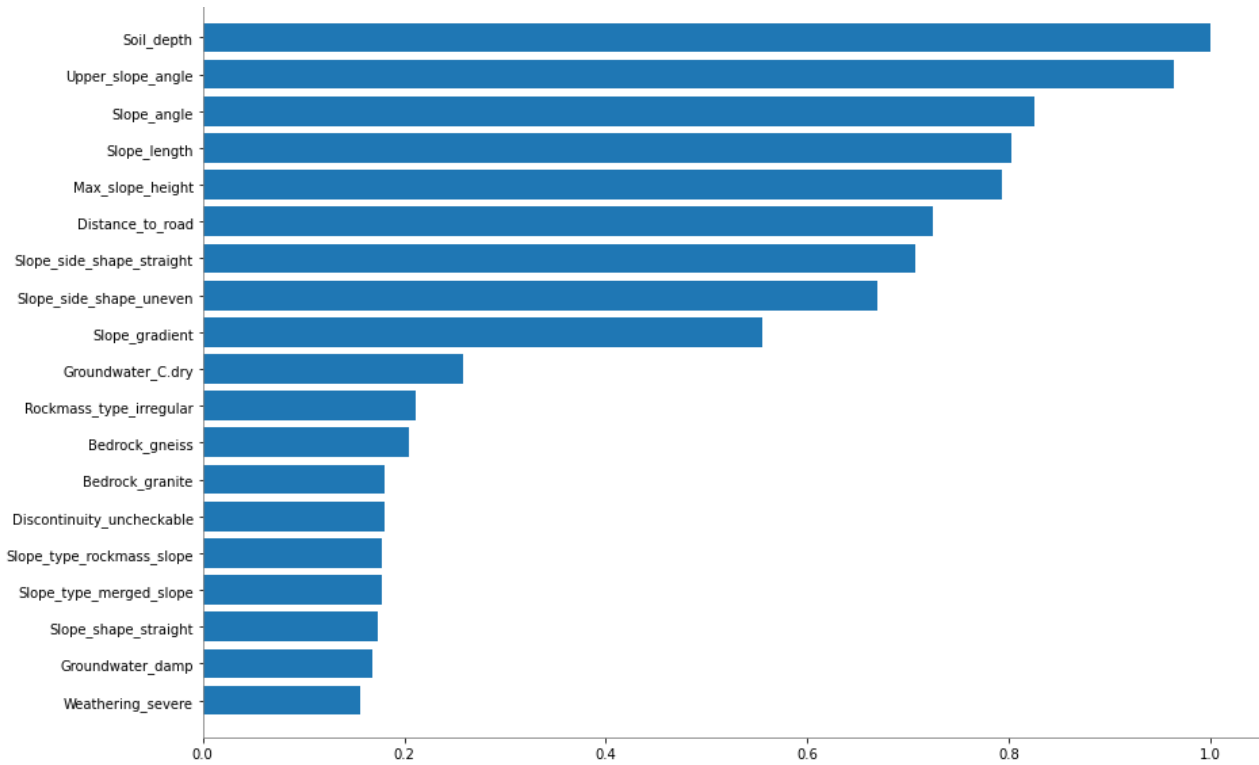


Fig. 4. Variable importance plot selecting landslide susceptibility factors for best model (XRT) selected from AutoML modelling.

Table 6. Ranking of variable importance and percentage using best model (XRT) selected from AutoML modelling

Ranking	Variable importance	Relative importance	Scaled importance	Percentage
1	Soil depth	10,652.473	1	0.085
2	Upper slope angle	10,272.392	0.964	0.082
3	Slope angle	8,800.288	0.826	0.070
4	Slope length	8,554.624	0.803	0.068
5	Max slope height	8,453.980	0.794	0.068
6	Distance to road	7,723.669	0.725	0.062
7	Slope side shape (straight)	7,541.327	0.708	0.060
8	Slope side shape (uneven)	7,139.971	0.670	0.057
9	Slope gradient	5,917.200	0.555	0.047
10	Groundwater (complete dry)	2,759.304	0.259	0.022
11	Rockmass type (irregular)	2,249.989	0.211	0.018
12	Bedrock (gneiss)	2,184.671	0.205	0.017
13	Bedrock (granite)	1,917.600	0.180	0.015
14	Discontinuity (uncheckable)	1,915.901	0.180	0.015
15	Slope type (rockmass slope)	1,899.607	0.178	0.015
16	Slope type (merged slope)	1,896.642	0.178	0.015
17	Slope shape (straight)	1,854.336	0.174	0.015
18	Groundwater (damp)	1,799.982	0.169	0.014
19	Weathering (severe)	1,664.631	0.156	0.013
20	Weathering (normal)	1,655.238	0.155	0.013

현장 적용성 검토

현장조사자료만을 바탕으로 국내에서 산사태 연구에 적용한 사례는 현재까지 없는 상태이다. 이 연구는 산사태 취약성 인자의 체계적이고 객관적인 선정과 이에 대한 변수 중요도 선정을 위해 AutoML을 이용하였다. 이를 통해 산사태에 영향을 미치는 인자들을 추출하였고 이 인자들을 포함한 발생 가능성 판별력을 분석하였다. 기술한 모델을 이용하여 현장조사 자료만으로 다른 지역의 산사태 발생 여부를 예측하는 연구에 신뢰성을 가지고 접근해 볼 수 있는 계기를 마련하게 되었다. 향후 도로 비탈면의 산사태 발생 가능성 여부와 산사태 취약성 인자들을 고려한 사면 안전 진단 평가 시 고려할 수 있을 것이라 판단된다. 또한 이 기법은 현장조사 시 지형 및 지질인자의 131가지 취약성 인자를 기반으로 산사태 발생 가능성을 예측함으로써 그에 따른 변수 중요도를 알 수 있었다.

이번 연구에서 제안한 산사태 예측 기법은 현장 조사자의 주관적 견해가 크게 영향을 미칠 것으로 생각되며 지역별 강우 사상 및 토질특성을 반영하지 못한 것은 한계점으로 작용한다. 그러나 제안되는 방법을 이용해 상부자연사면을 고려한 비탈면 위험 및 취약성 평가를 위한 지형 및 지질인자가 과거 비탈면의 붕괴 이력을 통한 검증 결과가 도출되어 기초조사 자료로써 데이터베이스 구축에 의의가 있다고 판단된다.

결론

도로 비탈면 현장 조사자료를 바탕으로 30,615개소의 암반사면, 토사사면, 혼합사면, 자연사면을 대상으로 AutoML 모델링을 구축하고 검증하였다. 도로 비탈면 불안정에 영향을 미치는 요인들의 복잡한 상관관계 분석은 131가지 주요 산사태 취약성 인자를 도출하여 입력변수로 설정하였다. 변수 중요도 분석결과 토양심도(8.52%), 상부사면경사(8.21%), 사면경사(7.04%), 사면길이(6.84%), 최대사면높이(6.76%), 도로이격거리(6.17%), 사면측면형상-직선형(6.03%), 사면측면형상-요철형(5.71%), 사면구배(4.73%), 지하수-완전건조(2.21%), 암반유형-불규칙(1.80%), 기반암-편마암(1.75%), 기반암-화강암(1.53%), 불연속면-확인불가(1.53%), 사면유형-암반사면(1.52%), 사면유형-혼합사면(1.52%), 사면형상-직선형(1.48%), 지하수-축축한(1.44%), 풍화도-심각한(1.33%), 풍화도-보통(1.32%) 순으로 변수 중요도가 산정되었다.

분석된 산사태 영향을 끼치는 요인 가중치에 대한 검증 결과 오차가 비교적 크지 않아 급경사지 안정성 평가에 대한 적절한 기준을 제시할 수 있을 것이라 기대된다. 개발된 모델의 신뢰성 검증을 수행한 결과 AUC 83.977%의 예측율을 확보한 것으로 나타났다. 이 모델은 현장 조사 자료만으로 산사태 발생가능성을 확률적 및 정량적으로 평가하였고 향후 의사결정자들에게 현장조사를 통한 사면진단 안전평가 시 신뢰성 있는 근거를 제공하리라 판단된다.

사사

본 연구는 국토교통부 국토교통과학기술진흥원의 건설기술연구사업(비탈면 대상 조사·점검 개선 및 보강시설 노후화 대응 기술 개발, 20SCIP-C151408-02)의 지원으로 수행되었으며 이에 깊은 감사를 드립니다.

References

Byun, L.S., Moon, H.K., 2010, The prediction of rock slope stability in Korea using artificial neural network, The Korean Society of Mineral and Energy Resources Engineers, 47(1), 31-44 (in Korean with English abstract).

- Chae, B.G., Kim, W.Y., Cho, Y.C., Kim, K.S., Lee, C.O., Choi, Y.S., 2004, Development of a logistic regression model for probabilistic prediction of debris flow, *The Journal of Engineering Geology*, 14(2), 211-222 (in Korean with English abstract).
- Dai, Z., Huang, Y., Cheng, H., Xu, Q., 2014, 3D numerical modeling using smoothed particle hydrodynamics of flow-like landslide propagation triggered by the 2008 Wenchuan earthquake, *Engineering Geology*, 180, 21-33.
- Di, B., Zhang, H., Liu, Y., Chen, N., Stamatopoulos, C., Luo, Y., Zhan, Y., 2019, Assessing susceptibility of debris flow in southwest China using gradient boosting machine, *Scientific Reports*, Published Online.
- Galelli, S., Castelletti, A., 2013, Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling, *Hydrology and Earth System Sciences*, 17(7), 2669-2684.
- Geurts, P., Ernst, D., Wehenkel, L., 2006, Extremely randomized trees, *Machine Learning*, 63(1), 3-42.
- Hong, W.P., Kim, W.Y., Song, Y.S., Lim, S.G., 2004, Prediction of landslide using artificial neural network model, *Journal of the Korean Geotechnical Society*, 29(8), 67-75 (in Korean with English abstract).
- Kim, J.C., Lee, S.M., Jung, H.S., Lee, S.R., 2017, Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea, *Geocarto International*, Published Online, 1000-1015.
- Merghadi, A., Yunus, A., 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.
- Nagarajah, T., Poravi, G., 2019, A review on automated machine learning (AutoML) systems, *Proceedings of the 2019 5th International Conference for Convergence in Technology (I2CT)*, 1-6.
- Nam, K.H., Wang, F.W., 2019, The performance of using an autoencoder for prediction and susceptibility assessment of landslides: a case study on landslides triggered by the 2018 Hokkaido Eastern Iburi earthquake in Japan, *Geoenvironmental Disasters*, 6(19), 1-14.
- Park, S.Y., Kim, J.S., 2019, Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance, *Applied Sciences*, Published Online, 1-19.
- Sassa, K., Nagai, O., Solidum, R., Yamazaki, Y., Ohta, H., 2010, An integrated model simulating the initiation and motion of earthquake and rain induced rapid landslides and its application to the 2006 Leyte landslide, *Landslides*, 7, 219-236.
- Xia, B., Zhang, H., Li, Q., Li, T., 2015, PETs: A stable and accurate predictor of protein-protein interacting sites based on extremely-randomized trees, *IEEE Transactions on Nanobioscience*, 14(8), 882-893.