

# 랜덤 포레스트 기법을 이용한 건설현장 안전재해 예측 모형 기초 연구

## Basic Study on Safety Accident Prediction Model Using Random Forest in Construction Field

강 경 수\*      류 한 국\*\*

Kang, Kyung-Su      Ryu, Han-Guk

### Abstract

The purpose of this study is to predict and classify the accident types based on the KOSHA (Korea Occupational Safety & Health Agency) and weather data. We also have an effort to suggest an important management method according to accident types by deriving feature importance. We designed two models based on accident data and weather data (model(a)) and only weather data (model(b)). As a result of random forest method, the model(b) showed a lack of accuracy in prediction. However, the model(a) presented more accurate prediction results than the model(b). Thus we presented safety management plan based on the results. In the future, this study will continue to carry out real time prediction to occurrence types to prevent safety accidents by supplementing the real time accident data and weather data.

키 워 드 : 건설안전, 기계 학습, 랜덤 포레스트, 특징 중요도

Keywords : construction safety, machine learning, random forest, feature importance

### 1. 서 론

건설안전재해를 다룬 대부분의 연구는 재해유형에 영향을 미치는 설명요인에 대한 중요도나 가설검정이 주를 이룬다. 하지만 안전재해는 추락, 충돌과 같은 발생형태를 실시간 예측하여 예방하는 것이 중요하다. 이를 예측하기 위해 기상관측 데이터를 사용하였다. 대부분 건설현장은 옥외작업으로 기후환경에 영향을 받으며 이러한 기후요소가 작업 생산성 또는 작업자에게 예기치 않은 위험을 초래할 수도 있다.<sup>1)2)</sup> 따라서 본 연구에서는 안전보건공단(KOSHA)에서 수집한 산업재해원인조사 데이터와 기상관측 데이터를 활용하여 발생형태 예측 가능성을 살펴보고, 특징변수들의 중요도를 도출하여 안전재해를 예방하기 위한 중점 관리요인을 제시하고자 한다. 본 연구는 KOSHA에서 수집한 산업재해원인조사 데이터(2008년에서 2014년까지)를 이용하였다. 해당 데이터는 총 55개 변수 중 23개의 변수로 축소하였고, 불균형 문제를 해결하기 위해 random under-sampling 기법을 통해 총 9,795개에서 6,374개로 조절하였다.<sup>3)</sup> 그리고 기상관측 데이터를 재해발생일자 기준으로 기온, 습도, 풍속 등 기상요인을 수집하고 이를 활용하여 블래지수와 체감온도를 파생변수로 생성·병합하였다.

### 2. 모형 구축 및 성능 평가

안전재해 발생형태를 예측하고 분류하기 위해 기계학습 기법 중 랜덤 포레스트(random forest)를 구축하였다. 산업재해원인조사와 기상관측 데이터를 모두 사용한 모형<sub>(a)</sub>과 기상관측 데이터만 사용한 모형<sub>(b)</sub>를 구축하였다.<sup>3)</sup> 모형<sub>(a)</sub>의 정확도는 0.739이며 모형<sub>(b)</sub>의 정확도는 0.326이다. 그림 1과 같이 macro-average 평가를 기준으로 충돌·접촉(0.96), 낙하·비래(0.91), 추락(0.91)으로 우수하게 평가됐지만 전도·전복(0.87)은 기준에 충족되지 못한 것으로 나타났다. 모형<sub>(b)</sub> 경우 정확도가 낮고 클래스별 분류 정확도가 충돌·접촉에 편향되어있다. 따라서 현재 기상관측 데이터로 구축한 모형<sub>(b)</sub>의 예측성은 낮은 것으로 나타났다.

### 3. 특징 중요도에 따른 중점 관리요인

모형<sub>(a)</sub>에서 도출된 발생형태 분류 정확도와 추출된 특징 중요도를 통해 발생형태별 중점 요인을 제시하였다. 분류 정확도가 기준치에 만족하는 것은 충돌·접촉, 낙하·비래, 추락이며 특징 중요도는 기인물(0.264), 가해물(0.184), 불안전 상태(0.131)이다. 표 1과 같이 빈도가 높은 요인을 중점 관리하고 예방할 필요가 있다. 충돌·접촉의 경우 휴대용 공구 중 절단용 공구와 같이 기계, 설비 취급상 비중이 높으며, 낙하·비래는 재료나 조립식 거푸집에서 조립, 해체 작업절차 중 부적절한 상태로 인한 발생빈도가 높다. 추락의 경우 기계 및

\* 창원대학교 경영학과 박사수료

\*\* 창원대학교 건축공학과 교수, 교신저자(hgryu@changwon.ac.kr)

작업 발판에서 옥외바닥 및 지표면으로 추락하고 단부 방호시설 미설치 또는 불량으로 인한 재해발생빈도가 높다. 따라서 본 연구에서 제시한 중점 요인들을 관리할 필요가 있으며 공통으로 작업절차 및 취급 부주의로 인한 빈도가 높으므로 이를 집중적으로 관리하여 작업자의 안전을 지키는 것이 중요하다.

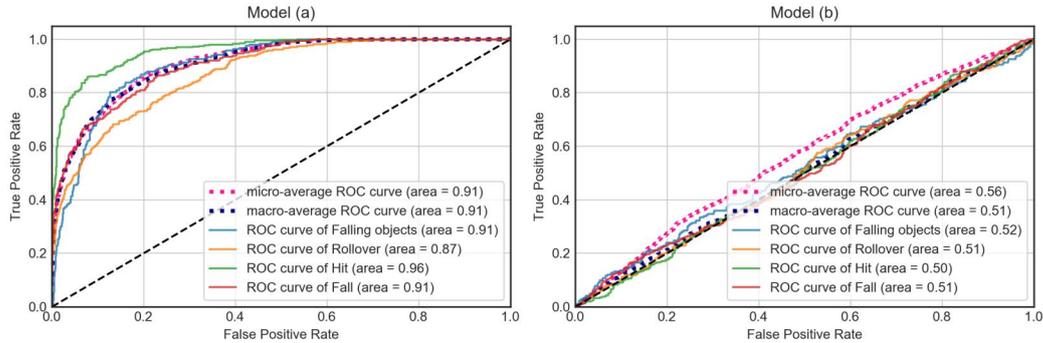


그림 1. 각 클래스별 ROC(Receiver Operating Characteristic) 평가

표 1. 발생형태와 특징 중요도에 따른 요인 빈도 분석

	총돌·접촉 #1(0.96)			낙하·비래 #2(0.91)			추락 #3(0.91)		
	기인물 #1(0.264)	가해물 #2(0.184)	불안전 상태 #3(0.131)	기인물 #1(0.264)	가해물 #2(0.184)	불안전 상태 #3(0.131)	기인물 #1(0.264)	가해물 #2(0.184)	불안전 상태 #3(0.131)
1	휴대용 공구 (443/0.260)	절단용 휴대공구 (312/0.183)	기계, 설비 취급상 위험 (412/0.242)	재료 (321/0.266)	조립식 거푸집 (171/0.142)	조립, 해체 작업절차 부적절 (208/0.172)	비계 및 작업발판 (618/0.329)	옥외바닥 및 지표면 (1036/0.552)	단부 방호시설 미설치 또는 불량 (386/0.206)
2	수공구 (188/0.110)	타격용 수공구 (85/0.050)	공구, 기구 취급상 위험 (236/0.139)	거푸집 및 동바리 (214/0.177)	철골빔 등 (104/0.086)	물 취급상 위험 (198/0.164)	계단 및 사다리 (404/0.215)	옥내바닥 (612/0.326)	작업발판 미설치 및 불량 (346/0.184)
3	재료 (187/0.110)	동근톱 (78/0.046)	작업유도자 배치 및 신호체계 미흡 (168/0.099)	설비·기계기구 부품 및 부속물 (97/0.080)	벽돌, 블록, 기와 (69/0.057)	지지 및 체결상태 불량 (91/0.075)	기타 건물·구조물 (261/0.139)	이동식 사다리 (27/0.014)	이동식 사다리 미설치 및 불량 (209/0.111)
4	육상교통수단(179/0.105)	표면처리용 휴대공구 (76/0.045)	물 취급상 위험 (154/0.090)	비금속광물 제품 (88/0.073)	기타 금속재료 (52/0.043)	기타 그 밖의 작업수행절차, 방법 부적절 (68/0.056)	단부 및 개구부 (227/0.121)	영구용계단 (17/0.009)	보호구 착용상태 및 성능 불량 (203/0.108)
5	일반제조 가공 설비·기계 (160/0.094)	철골빔 등 (76/0.045)	차량의 운전상 결함 (126/0.074)	휴대용 공구 (84/0.070)	못, 리벳, 나사 (49/0.041)	보호구 착용상태 및 성능 불량 (63/0.052)	육상교통수단 (76/0.040)	작업발판(작업대) (14/0.007)	신체반응, 동작상의 결함 (163/0.087)

#### 4. 결 론

본 연구는 안전보건공단에서 수집한 산업재해원인조사와 기상관측 데이터를 기반으로 안전재해 발생형태를 예측하고 분류하고자 하였다. 또한, 특징 중요도를 도출하여 발생형태에 따른 중점 관리요인을 제시하고자 하였다. 데이터별 모형을 구축하고 평가한 결과, 기상관측 데이터 기반으로 학습한 모형은 낮은 정확도를 보였다. 향후 본 연구의 한계점을 보완하여 기상관측 데이터를 활용한 실시간 안전재해 예측 연구를 지속해서 수행할 예정이다.

#### Acknowledgement

이 성과는 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2018021921).

#### 참 고 문 헌

1. 안전보건공단, 동절기 건설현장 안전보건 가이드라인, pp.1~45, 2016.10
2. 안전보건공단, 장마철 건설현장 안전보건 가이드라인, pp.4~68, 2018.5
3. Aurelien Geron, Hands-On Machine Learning with Scikit-Learn & Tensorflow, O'reilly, pp.225~267, 2018