

# Top-K 기반 건설현장 사고유형 예측 및 사고위험 특성 해석

김기남\*, 한재영\*, 고태규\*, 이민재\*  
\*충남대학교 토목공학과  
e-mail:go7140@cnu.ac.kr  
e-mail:jyh10240@naver.com  
e-mail:ktg7699@gmail.com  
e-mail:lmjcm@cnu.ac.kr

## Top-K Prediction of Construction Site Accident Types and Interpretation of Accident Risk Characteristics

Ki-Nam Kim\*, Jae-Young Han\*, Tae-Gyu Ko\*, Min-Jae Lee\*  
\*Dept. of Civil Engineering, Chungnam National University

### 요약

본 연구에서는 건설현장 작업조건을 기반으로 사고유형 발생 가능성을 예측하고, 사고위험 특성을 해석하기 위하여 Random Forest 기반의 다중분류 예측모형과 Top-K 기반 분석 접근을 적용하였다. 분석 결과, 제안된 모델은 정확도 0.436, Hit@3 0.755, Hit@5 0.926의 성능을 나타내어 기준모델 대비 우수한 예측 성능을 보였으며, 특히 단일 예측보다 복수의 사고유형을 고려하는 Top-K 접근이 효과적임을 확인하였다. 또한 Top-K 기반 패턴 분석을 통해 가시성 작업에서 떨어짐-넘어짐-물체에 맞음, 건설공구 작업에서 절단-베임-끼임-물체에 맞음과 같은 대표 사고유형 조합이 도출되었으며, 이는 건설사고가 복수의 사고유형이 동시에 발생 가능한 구조를 가짐을 보여준다. 아울러, WISE-SHAP 분석을 통해 사고유형별 영향 구조를 해석한 결과, '떨어짐'은 특정 작업조건에 의해 명확하게 설명되는 반면, '물체에 맞음'은 다양한 변수들이 복합적으로 작용하는 특성을 보였다. 따라서 본 연구는 건설사고를 단일 사고유형이 아닌 복수의 잠재 위험 관점에서 해석하고, 작업조건 기반 복합 위험 구조를 정량적으로 제시함으로써 현장 안전관리 의사결정 지원에 활용 가능한 분석 프레임워크를 제시하였다.

## 1. 서론

건설산업은 안전 관련 법·제도와 관리 체계가 지속적으로 강화되어 왔음에도 불구하고, 사고 발생이 반복되는 대표적인 고위험 산업으로 인식되고 있다. 선행연구에서는 사고 예방의 실효성을 확보하기 위해 제도적 규제뿐 아니라, 실제 작업이 수행되는 현장에서 적용 가능한 정량적 안전관리 기준이 필요하다는 점을 지적하고 있다[1]. 기존 건설안전 연구는 주로 사고 이후의 원인 분석이나 제도적·관리적 접근에 초점을 두어 왔으며, 실제 작업조건 변화를 반영하여 사고 위험을 사전에 정량적으로 판단하는 데에는 한계가 있다[2]. 이에 따라 현장 안전관리는 점검 및 교육 중심의 절차적 관리에 머무르는 경향이 있으며, 작업조건에 따른 위험 수준을 체계적으로 평가하는 접근은 상대적으로 미흡하다. 건설사고는 공중, 작업 프로세스, 사고객체, 작업 위치 등 다양한 작업조건 조합에 의해 발생하며, 동일 작업에서도 발생 가능한 사고유형과 위험 구조가 달라질 수 있다. 그러나 기존 연구에서 사고유형은 주로 사후 분류 항목으로 활용되어 왔다. 또한,

최근 머신러닝 기반 사고 예측 연구가 증가하고 있으며, 사고 발생 특성 및 심각도를 예측하고 SHAP을 활용하여 영향요인을 해석하는 연구가 수행된 바 있다[3]. 그러나 다수 연구는 전체 정확도 중심의 성능 평가 또는 단일 사고유형 예측에 초점을 두고 있어, 실제 현장에서 요구되는 복수 위험 시나리오를 충분히 반영하지 못하는 한계가 있다. 따라서 건설사고 예측은 단일 분류 문제가 아니라, 복수의 위험 후보를 식별하는 Ranking 기반 의사결정 문제로 해석될 필요가 있다.

따라서, 본 연구는 건설사고를 사고유형 중심의 조건부 구조로 재해석하고, 작업조건과 사고유형 간 관계를 정량적으로 분석하였다. 또한 다중분류기반 예측모형을 활용하여 사고유형별 발생 확률을 산정하고, 상위 K개 예측 결과를 활용하는 Top-K 기반 평가 방식을 적용하였다.

따라서 본 연구의 목적은 작업조건 기반 사고유형 발생 가능성을 도출 및 해석을 통해 현장 안전관리자의 의사결정을 지원할 수 있는 정량적 분석 프레임워크를 제시하는 데 있다.

## 2. Top-K 기반 건설현장 사고유형 예측모델 구축

## 2.1 데이터 구성

본 연구에서는 건설현장에서 발생한 사고 데이터를 활용하여 사고유형 예측을 위한 분석을 수행하였다. 데이터는 공중, 작업 프로세스, 사고객체, 작업 위치 등 작업조건을 설명하는 범주형 변수들로 구성되며, 사고유형을 종속변수로 설정하였다. 입력변수는 작업조건을 반영할 수 있도록 공중 분류, 세부 작업단계, 사고객체 특성, 작업 위치 정보 등으로 구성하였으며, 각 변수는 범주형 형태로 정의하였다. 이를 통해 특정 작업조건이 사고유형 발생에 미치는 영향을 분석할 수 있도록 하였다.

또한, 사고유형 데이터는 유형 간 발생 빈도의 차이가 크게 나타나는 불균형 특성을 가지므로, 단일 정확도 지표만으로는 모델 성능을 충분히 평가하기 어려운 구조를 가진다. 이에 따라 본 연구에서는 사고유형을 다중분류 문제로 설정하고, 확률 기반 예측 결과를 활용하여, 이를 사고유형 발생의 상대적 위험 분포로 해석하였다.

## 2.2 예측모델 구축

본 연구에서는 건설현장 작업조건과 사고유형 간의 관계를 학습하기 위하여 Random Forest 기반의 다중분류 예측모델을 구축하였다. Random Forest는 다수의 의사결정나무를 결합하는 앙상블 기법으로, 다양한 변수 간의 비선형 관계를 반영할 수 있어 건설사고와 같은 복잡한 데이터 분석에 적합하다[4].

모델 학습을 위해 전체 데이터는 학습 데이터와 테스트 데이터로 구분하였으며, 학습 데이터에서의 사고유형 분포를 기준으로 주요 사고유형을 선정하고 동일한 기준을 테스트 데이터에 적용하여 학습과 평가 간의 일관성을 유지하였다. 입력 변수는 공중, 작업 프로세스, 사고객체, 작업 위치 등 작업조건을 설명하는 범주형 변수로 구성되며, 이를 기반으로 사고유형 간의 관계를 학습하도록 하였다.

구축된 모델은 각 입력 데이터  $x$ 에 대해 사고유형별 예측 확률 벡터  $P(x)$ 를 출력한다. 본 연구에서는 이 확률값을 기반으로 상위  $K$ 개의 사고유형을 도출하는 Top-K 기반 예측 방식을 적용하였다.

$$Top-K(x) = \text{argsort}(P(x))_{1:K}$$

이때 각 변수는 다음과 같이 정의된다.

$x$  : 입력데이터 (작업조건 변수)

$P(x)$  : 입력  $x$ 에 대한 사고유형별 예측 확률

$K$  : 상위 예측 결과로 선택되는 사고유형의 개수

본 연구에서는  $K=3$ 을 기준으로 상위 3개의 사고유형을 도출하였다. 이는 단일 사고유형을 예측하는 방식보다 복수의 잠재 위험을 반영할 수 있어 실제 건설현장의 의사결정 구

조에 보다 적합하다. 또한, 제안 모델의 성능을 비교하기 위하여 사고유형의 발생 빈도를 기반으로 예측을 수행하는 기준모델(baseline)을 함께 구성하였다.

## 2.3 평가 지표

본 연구에서는 다중 사고유형 예측 결과를 평가하기 위하여 정확도(Accuracy)와 함께 Top-K 기반 평가지표를 적용하였다. 정확도는 전체 예측 결과 중 정답을 정확히 예측한 비율로 정의되며, 다중분류 문제에서 일반적으로 사용되는 성능지표이다.

그러나 건설사고 데이터는 클래스 간 불균형이 존재하고, 실제 현장에서는 단일 사고유형을 정확히 예측하는 것보다 발생 가능성이 높은 복수의 사고유형을 함께 고려하는 것이 중요하다. 따라서 본 연구에서는 상위  $K$ 개의 예측 결과 내에 실제 사고유형이 포함되는지를 평가하는 Top-K 기반 지표를 적용하였다. 이러한 평가지표는 정보검색 및 추천시스템 분야에서 활용되는 ranking 기반 평가 방식으로, 상위 결과의 적중 여부를 평가하는 데 적합하다[5].

Top-K 지표는 다음과 같이 정의된다.

$$Hit@K = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(y_i \in Top-K(x_i))$$

이때 각 변수는 다음과 같이 정의된다.

$N$  : 전체 샘플 수

$y_i$  : 실제 사고유형

$Top-K(x_i)$  : 입력  $x_i$ 에 대해 예측된 상위  $K$ 개의 사고유형 집합

$\mathbb{I}(\cdot)$  : 조건이 참일 경우 1, 거짓일 경우 0을 반환하는 지시함수

본 연구에서는  $K=3$ 을 기준으로 Top-3 성능을 평가하였다. 이는 단일 사고유형 예측보다 복수의 잠재 위험을 고려하는 실제 현장 의사결정 구조를 반영하기 위함이다.

또한 정확도 지표와 Top-K 지표를 병행하여 적용함으로써, 단일 예측 성능과 함께 복수의 위험 후보를 식별하는 성능을 동시에 평가하고자 하였다.

## 3. 예측 결과 및 사고위험 특성 분석

### 3.1 예측 성능 분석

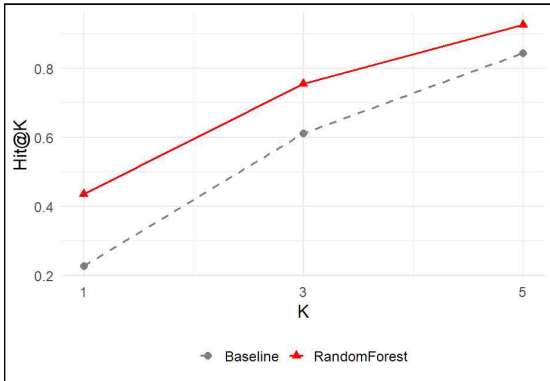
본 연구에서는 Random Forest 기반 예측모델의 성능을 평가하기 위하여 정확도(Accuracy)와 Top-K 지표를 적용하였으며, 기준모델(baseline)과의 비교를 통해 성능을 분석하였다. 분석 결과는 표 1에 제시하였다.

[표 1] 모델 성능 비교

Model	Accuracy	Hit@1	Hit@3	Hit@5
Baseline	0.228	0.228	0.611	0.844
<b>Random Forest</b>	<b>0.436</b>	<b>0.436</b>	<b>0.755</b>	<b>0.926</b>

표 1에 나타난 바와 같이, Random Forest 모델은 정확도 0.436, Hit@3 0.755로 나타났으며, 기준모델의 정확도 0.228, Hit@3 0.611 대비 전반적으로 우수한 성능을 보였다. 특히 Hit@3 기준에서 약 14% 이상의 성능 향상이 나타나, 상위 K개의 예측 결과를 고려하는 경우 모델의 예측 성능이 효과적으로 평가될 수 있음을 확인하였다.

또한 K 값이 증가함에 따라 Hit@K 성능은 0.436에서 0.926까지 크게 향상되는 경향을 보였으며, 이는 단일 사고유형 예측보다 복수의 사고유형을 함께 고려하는 접근이 데이터 특성을 보다 잘 반영함을 의미한다. 그림 1은 K 값에 따른 Hit@K 변화를 나타낸 것으로, K가 증가함에 따라 예측 성능이 향상되는 경향을 확인할 수 있다. 특히 Random Forest 모델은 모든 K 구간에서 기준모델 대비 높은 성능을 보였다.



[그림 1] Hit@K 기반 예측 성능 비교

한편 기준모델은 사고유형의 빈도 분포에 기반한 예측으로 인해 전반적으로 낮은 성능을 보였으며, 이는 작업조건을 반영한 예측 모델의 필요성을 뒷받침한다.

### 3.2 사고위험 특성 분석

본 연구에서는 Top-K 기반 예측 결과를 활용하여 특정 작업조건에서 반복적으로 나타나는 사고유형 조합을 도출하고, 이를 기반으로 사고 발생 구조를 분석하였다. 또한 대표 사고유형을 대상으로 WISE-SHAP 분석을 수행하여, 각 사고유형의 발생 가능성에 영향을 미치는 주요 변수와 그 영향을 해석하였다. 표 2와 같이, 동일 작업조건에서도 사고유형은 단일 유형이 아닌 복수의 유형으로 분산되어 나타나는 경향이 확인되었다. 예를 들어 가시

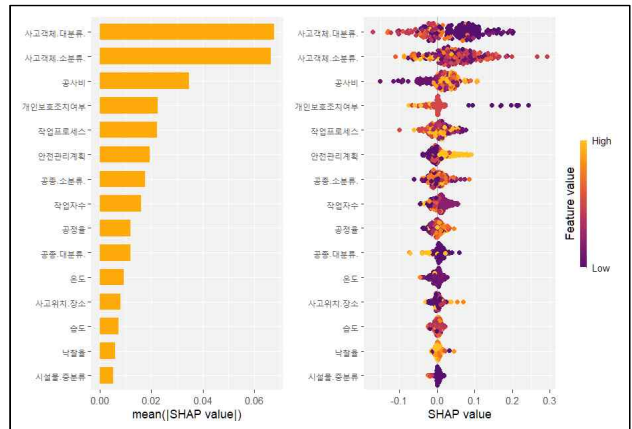
설 설치 및 해체 작업에서는 떨어짐, 넘어짐, 물체에 맞음이 주요 사고유형으로 동시에 나타났으며, 모델의 Top-K 예측에서도 유사한 사고유형이 상위 위험으로 제시되었다.

따라서, 건설사고는 특정 작업조건에서 복합적인 위험 구조를 가지며, Top-K 기반 접근이 이러한 다중 사고 가능성을 효과적으로 반영할 수 있음을 보여준다.

[표 2] Top-K 사고유형 패턴

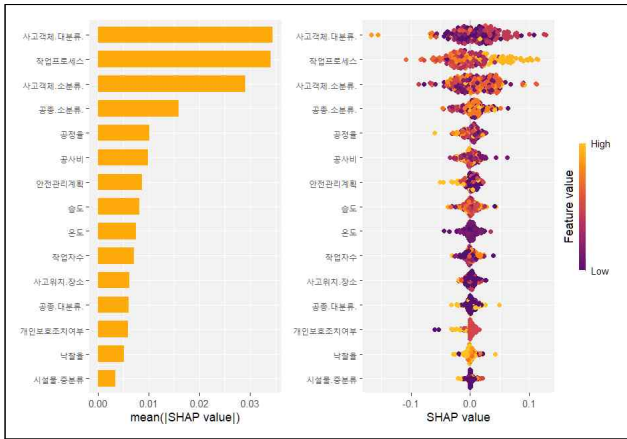
작업 유형	사고 객체	세부 객체	실제사고유형	Top-3 예측
설치 작업	가시 설	거푸 집	넘어짐(33.3%) / 끼임(22.2%) / 떨어짐(22.2%)	떨어짐(33.3%) / 넘어짐(29.6%) / 물체에 맞음(22.2%)
설치 작업	가시 설	거푸 집	넘어짐(33.3%) / 끼임(22.2%) / 떨어짐(22.2%)	넘어짐(33.3%) / 떨어짐(33.3%) / 물체에 맞음(20%)
설치 작업	가시 설	거푸 집	떨어짐(40%) / 물체에 맞음(40%) / 넘어짐(20%)	넘어짐(26.7%) / 떨어짐(26.7%) / 물체에 맞음(26.7%)
설치 작업	건설 자재	자재	넘어짐(40%) / 물체에 맞음(40%) / 찢림(20%)	넘어짐(33.3%) / 물체에 맞음(33.3%) / 끼임(26.7%)
이동	건설 자재	철근	넘어짐(60%) / 물체에 맞음(20%) / 찢림(20%)	넘어짐(33.3%) / 부딪힘(26.7%) / 찢림(20%)
해체 작업	가시 설	거푸 집	물체에 맞음(40%) / 부딪힘(40%) / 떨어짐(20%)	물체에 맞음(33.3%) / 부딪힘(26.7%) / 넘어짐(20%)
해체 작업	건설 자재	자재	물체에 맞음(60%) / 끼임(20%) / 떨어짐(20%)	넘어짐(33.3%) / 물체에 맞음(33.3%) / 부딪힘(33.3%)

그림 2는 떨어짐 사고유형에 대한 SHAP 분석 결과로, 사고객체, 소분류,와 사고객체,대분류,의 영향이 상대적으로 크게 나타나는 것을 확인할 수 있다. 또한 작업프로세스, 개인보호조치여부, 안전관리계획 등도 일정 수준 기여하는 것으로 나타났다. 이는 떨어짐 사고가 특정 작업조건 중에서도 특히 사고객체 관련 특성과 밀접하게 연관되는 경향이 나타난다.



[그림 2] 떨어짐사고유형 WISE-SHAP 분석 결과

그림 3은 해당 사고유형에 대한 SHAP 분석 결과로, 사고객체,소분류,와 사고객체,대분류,의 영향이 상대적으로 크게 나타났으며, 작업프로세스 또한 주요 변수로 작용하는 것으로 확인되었다. 또한 SHAP 값의 분포가 양(+)과 음(-) 방향으로 모두 나타나는 경향을 보여, 동일 변수라도 작업조건에 따라 사고 발생 가능성에 미치는 영향이 달라질 수 있음을 시사한다.



[그림 3] 물체에 맞음 WISE-SHAP 분석 결과

이와 같이 Top-K 기반 패턴 분석에서는 사고유형 간 동시 발생 구조가 확인된 반면, WISE-SHAP 분석에서는 사고유형별로 서로 다른 영향요인 구조가 나타났다. 이는 동일한 작업조건에서도 사고유형이 단일 원인보다는 다양한 요인의 결합에 의해 복합적으로 발생함을 시사한다. 따라서 건설사고는 복합 위험 구조를 가지며, 이를 반영하기 위해서는 단일 사고유형 예측이 아닌 복수의 사고유형을 동시에 고려하는 Top-K 기반 접근이 필요함을 확인하였다.

#### 4. 결론

본 연구에서는 건설현장 작업조건을 기반으로 사고유형 발생 가능성을 예측하기 위해 Random Forest 기반 다중분류 모형을 구축하고, Top-K 평가 및 사고위험 특성 해석을 수행하였다. 분석 결과, 정확도 0.436, Hit@3 0.755, Hit@5 0.926을 나타내어 기준모델 대비 우수한 성능을 보였으며, 특히 Top-K 기준에서 성능이 크게 향상되는 것으로 확인되었다. 이는 건설사고 예측에서 단일 사고유형이 아닌 복수의 사고유형을 함께 고려할 필요가 있음을 보여준다.

또한 Top-K 기반 패턴 분석 결과, 특정 작업조건에서 반복적으로 나타나는 사고유형 조합이 확인되었다. 가시설 설치 및 해체 작업에서는 떨어짐, 넘어짐, 물체에 맞음, 건설공구 작업에서는 절단-베임, 끼임, 물체에 맞음이 대표적인 사고유형 패턴으로 나타났으며, 이는 건설사고가 복수의 사고유형이 동시에 발생 가능

한 복합 위험 구조를 가진다는 점을 보여준다.

WISE-SHAP 분석에서는 사고유형별 영향 구조의 차이가 확인되었다. 떨어짐 사고유형은 사고객체와 작업프로세스의 영향이 상대적으로 뚜렷하게 나타난 반면, 물체에 맞음 사고유형은 다양한 변수들이 복합적으로 작용하는 경향을 보였다. 이는 동일한 작업조건에서도 사고유형별 위험 구조가 상이하게 나타날 수 있음을 보여준다.

따라서 본 연구는 건설사고를 단일 정답 예측이 아닌 복수 사고유형 관점에서 해석하고, 작업조건별 복합 위험 구조를 정량적으로 확인하였다는 데 의의가 있다. 이러한 접근은 현장 관리자가 발생 가능성이 높은 사고유형을 사전에 인지하고, 복수 위험요인에 대응하는 예방적 안전관리 전략 수립에 활용될 수 있을 것으로 판단된다.

#### 감사의 글

본 연구는 한국연구재단 이공분야기초연구(NRF-2022 R11A3068942)에 의해 수행되었으며 본 연구의 지원에 감사드립니다.

#### 참고문헌

- [1] 채정식, “중대재해처벌법 실효성 확보를 위한 법적 연구”, 동아대학교 박사학위논문, 2025.
- [2] Chi, S.; Han, S., “Analyses of systems theory for construction accident prevention with specific reference to OSHA accident reports”, International Journal of Project Management, 31(7), pp. 1027-1041, 2013.
- [3] Kim, K.-n., Cho, D.-g., Lee, M.-j., “A Machine Learning Approach for Factor Analysis and Scenario-Based Prediction of Construction Accidents”, Buildings, 15, 4343, 2025.
- [4] Breiman, L., “Random Forests”, Machine Learning, 45(1), pp. 5-32, 2001.