

## **Applying decision tree algorithm to explore occupational injuries in the Turkish construction industry**

**DOI:10.36909/jer.12209**

Yilmaz Ogunc Tetik<sup>\*</sup>, Ozge Akboga Kale<sup>\*\*</sup>, Irem Bayram<sup>\*\*\*</sup> Selim Baradan<sup>\*\*\*</sup>

<sup>\*</sup>Department of Civil Engineering, Mugla Sitki Kocman University, Mugla, 48000, Turkey.

<sup>\*\*</sup>Department of Civil Engineering, Izmir Demokrasi University, Izmir, 35140, Turkey.

<sup>\*\*\*</sup>Department of Civil Engineering, Ege University, Izmir, 35040, Turkey.

<sup>\*</sup>Corresponding Author: yotetik@gmail.com, yilmazogunctetik@mu.edu.tr

### **ABSTRACT**

Occupational injuries and fatalities are one of the most significant issues in the construction industry. Variables, such as workers' behavior, age, worksite condition, and type of activity, play key roles in the occurrence of construction accidents. In recent years, data mining techniques have been successfully used not only in health, economy, and social sciences but also in construction-related fields. In this study, C5.0 decision tree algorithm was used to analyze the accident data obtained from the Social Security Institution of Turkey. A classification tree model was created to discover the associations between the attributes. The results show the relationship between the injury status of workers and the attributes, and the accuracy rate of the model was 70.26%. Meanwhile, according to findings, unsafe conditions, type of accident, and activity type were the most important attributes in the model. Furthermore, the predictor importance of the attributes was compared, and several outcomes were discovered; for instance, the workers' educational background has greater predictive power than age. On the other hand, the branches of the decision tree pointed out several attribute sequences due to their high rated serious/fatal injury rates. The results of this study can be used in the prevention and mitigation strategies for construction accidents.

**Keywords:** accident analysis; decision tree; C5.0 algorithm; construction safety.

## INTRODUCTION

According to the International Labor Organization (ILO), an estimated number of 2.3 million workers die annually due to occupational accidents, including nearly 360,000 fatal accidents (Sarkar et al., 2019). It is also reported that 3.2% of the total number of workers in Europe have accidents in their working places annually (EEC, 2010). The construction industry, which has a considerable high rate of fatal occupational accidents, is one of the most hazardous industries. At least 60,000 people die in construction sites, and several others have serious injuries or occupational diseases (Cheng et al., 2012). To sustain safer worksite conditions in the construction sector, safety management decisions should be made and safety systems should be implemented based on the potential hazards in each activity of the construction process (Qammaz and AlMaian, 2020). For instance, to mitigate falling incidents, Occupational Safety and Health Administration (OSHA) suggested several precautions, such as fall protection systems, positioning of device systems, and fall restraint systems (OSHA, 2015). Moreover, many studies used various analytical techniques to determine the causes of fall from height accidents and take precautions (Hsiao, 2014).

Hence, the knowledge of the major occupational hazards, as well as their causes and factors, is necessary for improving the safety conditions of workers. Occupational accident datasets are often large and complex, and they require specialized approaches that can handle large amounts of information for analysis. One such approach is data mining (DM), which can examine the genesis of occupational accidents and extract rules or behavioral patterns of injuries. Decision trees and classification algorithms are popular and useful tools for solving such problems because they are human-comprehensible techniques (Cao and Rockett, 2015).

The aim of data mining is to detect and predict qualitative and quantitative patterns in occupational accident data, thereby leading to new information (Shirali et al., 2018).

Several different method and algorithms, such as artificial neural networks (ANN) (Kim et al., 2017),

support vector machines (SVM) (Zhou et al., 2017), genetic algorithms (GA) (Sarkar et al., 2016), decision trees (DT) (Cho et al., 2017), and a combination of different methods (Sarkar et al., 2019), in other words hybrid models (Katanalp and Eren, 2020) are employed in data mining to perform data classification and prediction tasks. Among the aforementioned techniques, the decision tree is a very powerful and comprehensible approach because of its intuitive binary “if-then” rule-based structure. Researchers are preferred DT for analyzing accident data in the construction sector (Cheng et al., 2012; Amiri et al. 2016; Mistikoglu et al., 2015).

This study is conducted to understand the causes of accidents associated with construction industry through the analysis of 1373 occupational accidents in the Turkey between 2010 and 2012 and to develop intervention strategies to identify the causes of accidents and review of the literature. The objective of this study is to identify the factors that are most strongly associated with accidents and injuries in the construction sector by applying the C5.0 decision tree algorithm. The main factors that influence the critical attributes of occupational accidents are identified to improve prevention policies, minimize risks, and reduce injuries and deaths in construction sites. Consequently, the findings of this study will give an opportunity to provide further precautions for preventing future accidents. It is expected that the findings obtained in the study will guide the companies in the construction industry/occupational safety departments, inform the occupational safety experts and inspire the researchers.

### **DATA ACQUISITION AND DEMOGRAPHICS**

The data used in this study was acquired from the Social Security Institution (SSI) database. The data comprises fatal and nonfatal injuries in construction sites and production areas, such as ready-mix concrete, prefabricated structural concrete, and steel production. Overall, 1,373 construction-related cases occurred between 2010 and 2012 were obtained

from the database and statistically analyzed. A new taxonomy was created for statistical analysis. Additional categorical variables were also required to be added to the new database by using the information extracted from open-ended questions. Table 1 summarizes the taxonomy of the

study. The variables with missing information are presented with their valid percentages.

Injury severity score (ISS) was used to describe the degree of injury of each case. The ISS for each case was estimated by using the information on the type of injury and injured body part obtained from the accident report forms in the SSI workplace. The ISS for each casualty, ranging from 0 to 75, was calculated and then classified into two groups: minor/moderate ( $3 \leq \text{ISS} \leq 8$ ) and serious/fatal ( $\text{ISS} \geq 9$ ). The injury types were categorized according to the International Classification of Diseases (ICD) codes (WHO, 2010). Furthermore, the variables of unsafe actions and conditions were deduced from the open-ended questions in the workplace accident report forms based on the ILO's accident cause theory (ILO, 2014) and Hill's definitions (Hill, 2004). The remaining variables (project type, construction purpose, activity type, work experience, construction trade, age of worker, educational background) were directly obtained from the workplace accident reports and categorized.

**Table 1** Data layout and frequencies

| Variable                  | Categories                                       | Code | Count | Percentage (%) |
|---------------------------|--|------|-------|----------------|
| Degree of injury (DoI)    | Minor/Moderate                                   | 1    | 860   | 62.6           |
|                           | Serious/Fatal                                    | 2    | 513   | 37.4           |
| Project type (PT)         | New project                                      | 1    | 1331  | 96.9           |
|                           | Maintenance/Repair projects                      | 2    | 42    | 3.1            |
| Construction purpose (CP) | Residential                                      | 1    | 1134  | 82.6           |
|                           | Institutional/Commercial                         | 2    | 204   | 14.9           |
|                           | Infrastructure/Heavy construction                | 3    | 35    | 2.5            |
| Type of accident (TA)     | Fall from height                                 | 1    | 347   | 25.3           |
|                           | Fall at the same level                           | 2    | 157   | 11.4           |
|                           | Bite/Sting/Scratch                               | 3    | 198   | 14.4           |
|                           | Struck by falling object                         | 4    | 416   | 30.3           |
|                           | Caught in or between objects                     | 5    | 184   | 13.4           |
|                           | Vehicle accident/Explosion/ Electric shocks/Fire | 6    | 71    | 5.2            |
| Unsafe act (UA)           | Inappropriate position for task                  | 1    | 207   | 15.1           |

|                         |  |    |     |      |
|-------------------------|--|----|-----|------|
|                         | Unsafe work practices                        | 2  | 419 | 30.5 |
|                         | No Personal Protective Equipment (PPE) usage | 3  | 488 | 35.5 |
|                         | Unsafe act by a third party                  | 4  | 145 | 10.6 |
|                         | Defective/Inappropriate equipment usage      | 5  | 114 | 8.3  |
| Unsafe condition (UC)   | Insufficient/Lack of written work procedures | 1  | 442 | 32.2 |
|                         | Poor housekeeping                            | 2  | 324 | 23.6 |
|                         | No PPE provided                              | 3  | 214 | 15.6 |
|                         | Faulty tool/equipment/machinery              | 4  | 121 | 8.8  |
|                         | No collective protection systems             | 5  | 272 | 19.8 |
| Activity type (AT)      | Formwork                                     | 1  | 408 | 29.7 |
|                         | Material handling                            | 2  | 246 | 17.9 |
|                         | Cutting works                                | 3  | 43  | 3.1  |
|                         | Plaster/Paint& insulation                    | 4  | 165 | 12.0 |
|                         | Tile setting                                 | 5  | 12  | 0.9  |
|                         | Excavation/Filling                           | 6  | 10  | 0.7  |
|                         | Concrete works                               | 7  | 68  | 5.0  |
|                         | Assembly/Disassembly& welding                | 8  | 107 | 7.8  |
|                         | Commuting                                    | 9  | 48  | 3.5  |
|                         | Work breaks                                  | 10 | 63  | 4.6  |
|                         | Rebar tie                                    | 11 | 124 | 9.0  |
|                         | Field inspection                             | 12 | 35  | 2.5  |
|                         | Maintenance/Repair                           | 13 | 44  | 3.2  |
| Work experience (WE)    | First day                                    | 1  | 124 | 9.0  |
|                         | First week                                   | 2  | 126 | 9.2  |
|                         | 2-4 weeks                                    | 3  | 334 | 24.3 |
|                         | 1-3 months                                   | 4  | 428 | 31.2 |
|                         | 3-6 months                                   | 5  | 203 | 14.8 |
|                         | 6-12 months                                  | 6  | 98  | 7.1  |
|                         | >12 months                                   | 7  | 60  | 4.4  |
| Construction trade (CT) | Form worker                                  | 1  | 426 | 31.0 |
|                         | Unskilled worker                             | 2  | 369 | 26.9 |
|                         | Foreman                                      | 3  | 94  | 6.8  |
|                         | Plasterer/Painter                            | 4  | 152 | 11.1 |
|                         | Heavy equipment operator                     | 5  | 35  | 2.5  |
|                         | Assembly/Installation/Maintenance/Welder     | 6  | 76  | 5.5  |
|                         | Ironworker                                   | 7  | 151 | 11.0 |
|                         | Finishing works/Cleaning                     | 8  | 70  | 5.1  |
| Age of worker (WA)      | ≤18  | 1  | 19  | 1.5  |
|                         | 19-24  | 2  | 251 | 20.0 |
|                         | 25-29  | 3  | 246 | 19.6 |
|                         | 30-34  | 4  | 237 | 18.9 |
|                         | 35-39  | 5  | 168 | 13.4 |
|                         | 40-44  | 6  | 152 | 12.1 |
|                         | 45-49  | 7  | 115 | 9.2  |
|                         | ≥50  | 8  | 66  | 5.3  |

|                             |                   |   |     |      |
|-----------------------------|-------------------|---|-----|------|
| Educational background (EB) | Literate          | 1 | 70  | 9.5  |
|                             | Illiterate        | 2 | 18  | 2.5  |
|                             | Elementary school | 3 | 411 | 56.1 |
|                             | Middle school     | 4 | 213 | 29.1 |
|                             | ≥ High school     | 5 | 21  | 2.9  |

## METHODOLOGY

In this study, C5.0 DT algorithm was preferred and SPSS Modeler 18.0 software was used to classify and discover the relationships between the attributes. The C5.0 decision tree algorithm is an improved and modified form of the C4.5 and Quinlan's ID3 algorithms (Quinlan, 1993). The advantages of the C5.0 algorithm is that it is faster, it is easy to read and interpret, once the variables have been created, there is less data cleaning required, cases of missing values and outliers have less significance on the decision tree's data, has better memory usage, generates compact decision tree, handles missing values, provides more accuracy, cross-validation, boosting, and winnowing for large datasets. C5.0 algorithm has a better accuracy rate compared to other algorithms, such as CART, CHAID, and QUEST, for decision processes (Elsayad and Elsalamony, 2013). As an entropy-based algorithm, C5.0 uses information measure for selection. First, it selects the attribute that provides the highest amount of information in the data as the root node (Otero et. al.2012). The data are then split into leaves based on the root attribute splitting criterion. If most of the case data in a branch belongs to the same class, the node is labeled with that class, and further splitting in this branch is terminated. The algorithm is then recursively applied to all branches of the tree until all cases in each branch result in a leaf node or reach some designated stopping criteria (Konda,2010). The algorithm works by calculating the information gain of the entropy. In Eq (1), gain ratio is determined for F- attributes of the dataset.

$$Gain\ Ratio(F) = Information\ Gain(F)/Split\ info(F) \quad (1)$$

In Eq (2) Eq (3) and Eq (4), S denotes the sample of the given data, n refers to the number of different class types, pi refers to the proportion of class level frequency to the total

frequency.  $Info_F(S)$  refers to the information required to classify  $S$  after using  $F$  to split it to  $m$  partitions.

Continuously, gained information value is calculated as  $Gain(F)$  based on  $tuple(F)$ .

$$Information\ Gain(F) = Entropy(S) - Info_F(S) \quad (2)$$

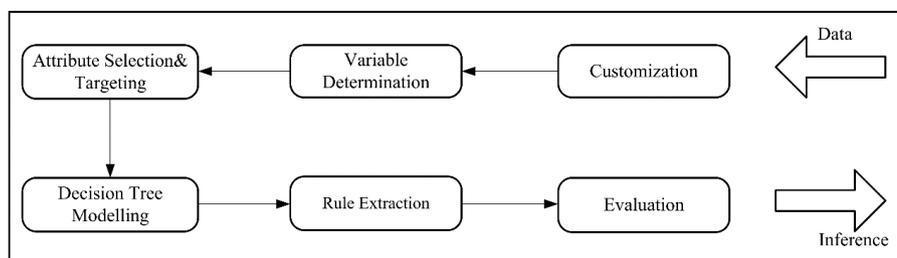
$$Entropy(S) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (3)$$

$$Info_F(S) = \sum_{j=1}^m |S_j|/|S| \times Info(S_j) \quad (4)$$

In Eq (5), split info value, evaluated as the normalization of the information gain value. The splitting mechanism is repeated until no information gain is yielded or each node contains only observations from a single class.

$$Split\ info(F) = - \sum_{j=1}^m |S_j|/|S| \times \log_2(|S_j|/|S|) \quad (5)$$

There is a pruning process which prevents tree growth and helps classifies the training set correctly. The C5.0 uses error-based pruning to remove artificial features that are not supported by the data (DeFries and Chan, 2000). The purpose of the pruning step is to reduce the complexity of the final classifier and improve the predictive accuracy at the same time by removing noisy data and reducing overfitting. In this study, the data were divided into two outcomes and a single attribute at each node. As part of the learning process, 70% of the data were used for training and 30% were used for testing. Along with the selection of the target attribute, decision tree algorithm was applied to data and the tree model was generated. The extracted rules of decision tree were evaluated and results were revealed (Figure 1).



**Figure 1** Process flowchart

## RESULTS AND DISCUSSION

The generated decision tree has a three-level depth with 24 rules distributed as branches. Since the decision tree wasn't small enough to be displayed here properly the pruned version of the tree shown in Figure 2. The DoI rate in the splitting nodes can be determined from the

percentage calculated by dividing number of cases to total cases. For instance, 50% (nine observed cases) of the workers were seriously injured or dead because PPE was not provided to prevent "fall from height" accidents. This situation was split into three leaves based on the educational background attribute (Rule Number: XV, XVI, XVII). From Table 2, the number of cases is 9 (1+8+0) and the total number of cases is 18 (5+13+0). Therefore, the percentage value was calculated as 50%.

On the other hand, results as attribute rules in accidents with serious/ fatal injuries (injury code: 2) are depicted in Table 2. Zero DoI indicates that no observation was recorded for those respective attributes. The relationships between significant attributes (Table 2) that result in minor/moderate or serious/ fatal injuries are given below:

- From Rule Number I, due to insufficient/lack of work procedures, about 19% of workers had serious/ fatal injuries (Tree Level (TL): 1, Rule Number (RN): I).
- The "faulty tool/equipment/machinery and "no collective protection systems" attributes resulted in 50.55% and 64.83% serious/ fatal injuries respectively. This indicates that they have a significant impact on high rates among other categories in accidents (TL: 1).
- In Rule Number II, about 19% of accidents occurred because of inappropriate positioning in tasks and unsafe actions by a third party. Moreover, 33% of these accidents resulted in serious/ fatal injuries (TL: 2, RN: II).
- In Tree Level 2 and Rule Number IV, 30% of workers who were struck by falling objects, bitten/stung/scratched, and caught in or between objects had serious/ fatal injuries because PPE was not provided. In a similar study, "being struck by an object" was the most observed

and fatal accident type that generally occurred during operations (Ayhan et. al., 2020).

- In workplaces with poor housekeeping, 77% of the accidents resulted in serious/ fatal injuries for formworkers, plasterers/painters, and the foreman who did not use PPE. On the other hand, in the same situation, 90% of unskilled workers, assembly/installation/welders/ maintenance, and ironworkers had minor/moderate injuries (TL: 3, RN: XII-XIII).

- In Rule Numbers XXI and XXII, due to faulty tool/equipment/machinery, 73% of workers who were 34 years old and below had “fall from height” accidents that resulted in minor/moderate injuries. However, 82% of workers who were 35 years old and above had serious/ fatal injuries (Tree Level: 3, Rule Number: XXI and XXII). These numbers indicate that increasing age rate is highly efficient for predicting occupational accidents. Similarly, Cheng et al. (2012) discovered that older workers (>55 years old) were the most likely to be involved in fatal accidents. In past studies, most accidents that resulted in fatalities were observed in the age group of 50 years or above (Amiri et. al., 2016).

- When collective protection systems were not used, and there were unsafe work practices, no PPE usage, unsafe action by a third party, and usage of defective/inappropriate equipment,

the accidents became a significant issue. In the “fall from height” accidents cases mentioned above, 70% of workers had serious/fatal injuries (TL: 3, RN: XXIV).

- Fall from heights was the most significant accident type. Meanwhile, 56.25% of serious/ fatal injuries were associated with faulty tool/equipment/machinery and 66.86% with no collective prevention systems.

The importance values of the attributes are shown in Figure 3. High-valued predictor importance indicates the predictive power of the attribute. “Unsafe condition” is apparently the most important attribute for the DoI. Accidents resulting from unsafe conditions are more severe due to human errors, no PPE provided, no PPE usage, ignoring necessary precautions,

etc. (Kang and Ryu, 2019). The “type of accident” and “activity type” have the second and third most significant predictor importance, respectively. “Construction trade”,

“educational background”, and “unsafe action” are the remaining attributes in the prediction process.

Although the educational status of workers seemed to be significant in the accidents, attributes such as unsafe condition or activity type are more important. Moreover, educational background is less significant in occupational accidents compared to the construction trade.

**Table 2** Tree rules for the degree of injury

| Tree Level | Rule Number            | Attribute Rule                         | Degree of Injury (%) - Code:2 | Total Cases/Number of Cases |
|------------|------------------------|--|-------------------------------|-----------------------------|
| 1          | I                      | UC=1                                   | 19.21                         | 307/59                      |
| 2          | II                     | UC=2& UA=1,4                           | 33.51                         | 185/62                      |
|            | III                    | UC=2& UA=5                             | 100.0                         | 4/4                         |
|            | IV                     | UC=3& TA=3,4,5                         | 30.08                         | 123/37                      |
|            | V                      | UC=3& TA=6                             | 100.0                         | 4/4                         |
|            | VI                     | UC=4& TA=2,3,4,                        | 36.17                         | 47/17                       |
|            | VII                    | UC=4& TA=5,6                           | 71.43                         | 28/20                       |
|            | VIII                   | UC=5& TA=2,3,5                         | 27.27                         | 11/3                        |
|            | IX                     | UC=5& TA=4,6                           | 100.0                         | 2/2                         |
|            | 3                      | X                                      | UC=2& UA=2& WA=1,2,3,4        | 19.64                       |
| XI         |                        | UC=2& UA=2& WA=5,6,7,8                 | 88.28                         | 9/8                         |
| XII        |                        | UC=2& UA=3& CT=1,3,4                   | 77.77                         | 9/7                         |
| XIII       |                        | UC=2& UA=3& CT=2,6,7                   | 9.09                          | 11/1                        |
| XIV        |                        | UC=2& UA=3& CT=5,8                     | 0.0                           | 0/0                         |
| XV         |                        | UC=3& TA=1& EB=1,2                     | 11.11                         | 5/1                         |
| XVI        |                        | UC=3& TA=1& EB=3,4                     | 66.67                         | 13/8                        |
| XVII       |                        | UC=3& TA=1& EB=5                       | 0.0                           | 0/0                         |
| XVIII      |                        | UC=3& TA=2& AT=1,3,4,5,6,9,10,11,12,13 | 0.0                           | 0/0                         |
| XIX        |                        | UC=3& TA=2& AT=2                       | 100.0                         | 2/2                         |
| XX         |                        | UC=3& TA=2& AT=7,8                     | 25.0                          | 4/1                         |
| XXI        |                        | UC=4& TA=1& WA=1,2,3,4                 | 26.78                         | 7/2                         |
| XXII       |                        | UC=4& TA=1& WA=5,6,7,8                 | 82.03                         | 9/7                         |
| XIII       |                        | UC=5& TA=1& UA=1                       | 18.18                         | 11/2                        |
| XXIV       | UC=5& TA=1& UA=2,3,4,5 | 70.25                                  | 158/111                       |                             |

When the educational background is not considered, the activity, i.e., the characteristic of the task, is more important for predicting accidents (Akboga Kale and Baradan, 2020). The “project type”, “construction purpose”, and “work experience” attributes were evaluated as

insufficient predictors. Besides, it was revealed that the “worker’s age” is a less significant attribute in the model. Table 3 summarizes the correct and incorrect predictions of the training and test results, as well as the prediction accuracy. 74.16% and 70.26% accuracy rates were obtained for the training and test data, respectively. Similar rates were calculated in other studies and regarded as satisfactory (Kang and Ryu, 2019).

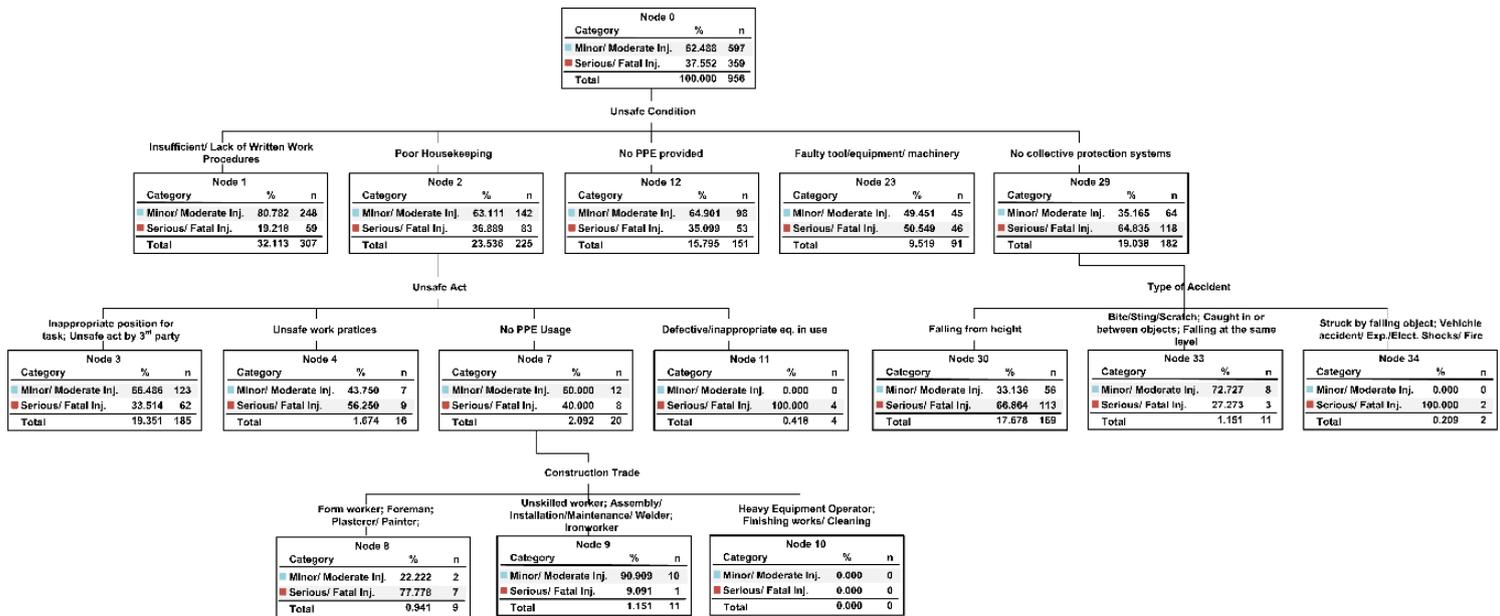


Figure 2 Pruned C5.0 decision tree algorithm

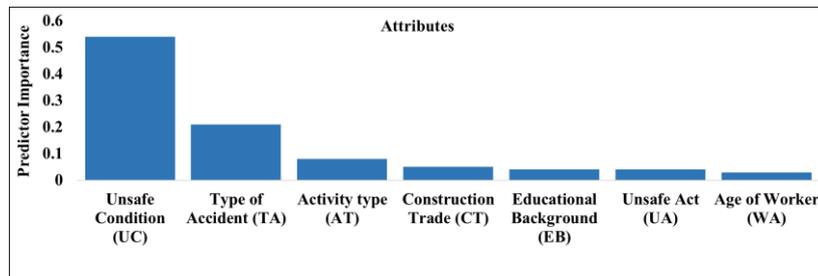


Figure 3 Importance values of attributes

Table 3 Training and testing results of the C5.0 algorithm

| Partition   | Training Data   |              | Test Data       |              |
|-------------|-----------------|--------------|-----------------|--------------|
|             | Number of Cases | Accuracy (%) | Number of Cases | Accuracy (%) |
| Correct     | 709             | 74.16        | 293             | 70.26        |
| Incorrect   | 247             | 25.84        | 124             | 29.74        |
| Total Cases | 956             |              | 417             |              |

## CONCLUSION

Occupational accidents are still a significant issue in the construction industry, and the fatality rate in Turkey is one of the highest among other countries (SSI, 2017). Many construction workers suffer from injuries, which breeds work interruptions and incapacity problems. It is important to discover root causes of accidents to reduce fatality rates and avoid

all types of injuries. Therefore, the relationships between the accident variables were investigated to help accident prevention effort. Interdependent results may be regarded as a mitigation strategy by OHS professionals for safety management. Based on the results, the accuracy rate of the model is acceptable, and regard as efficient.

This research showed that “unsafe condition” is the most significant variable among other attributes, whereas the worker’s age is the least significant one. When the root causes of accidents are discovered, prevention methods can be applied before the accident occurs. Thus, injury and fatality rates can be reduced as part of safety management. The decision tree demonstrated that accidents involving fall from heights have the greatest ratio in lack of collective protection systems, which resulted in serious/ fatal injuries. This indicates that accidents caused by unsafe actions can be reduced by increasing fall prevention and protection efforts measures. This study also emphasized the importance of providing PPE. According to OHS professionals’ observations, contractors in Turkey usually complain about the workers not using PPE despite being provided. However, results of the decision tree analysis revealed the opposite: lack of providing PPE had more importance than its lack of usage. Moreover, the worker’s age is a significant attribute in unsafe work practices, and serious/ fatal injury rates are considerably high when the workers are above 35 years old. Therefore, it is recommended to avoid assigning older (35+ years old) workers to activities that have the potential of becoming unsafe work practices.

In this study, ISS and DT methods, which were not used together before, were applied together. This use has taken a practical approach to digitize the severity of past occupational

accidents and to examine the root causes and to determine the precautions to be taken to prevent accidents. Limitations of this study include the occurrence on SSI accident reports in Turkey, unavailability of consistent data, and unfilled information categories in accident reports. The data used covers the years 2010-2012 due to the difficulty of obtaining the data. In addition, it is worth noting that the frequency distributions are specific to the data set

studied. In different data sets, it is possible that the frequency distribution of variables such as project type, construction purpose, age of the victim, educational status of the victim, work experience varies in percentage. However, it is quite possible that the results obtained can be used as a mitigation strategy to prevent repetitive accidents. As a future study, other decision tree algorithms or classification techniques such as Random Forest can be applied, metaheuristic optimization methods such as Particle Swarm Optimization can be embedded in C5.0, or combinations of them can be used to improve the accuracy rate of the model.

#### **ACKNOWLEDGMENTS**

We are grateful to Ege University Planning and Monitoring Coordination of Organizational Development and Directorate of Library and Documentation for their support in editing.

## REFERENCES

- Sarkar, S., Raj, R., Vinay, S., Maiti, J. & Pratihar, D.K. 2019.** An optimization-based decision tree approach for predicting slip-trip-fall accidents at work. *Safety Science* 18:57–69.
- Eurostat European Commission (EEC), 2010.** Eurostat statistical book. Health and safety at work in Europe (1999-2007)- A statistical portrait. 2010 ed.: 8. <https://ec.europa.eu/eurostat/documents/3217494/5718905/KS-31-09-290-EN.PDF/88eef9f7-c229-40de-b1cd-43126bc4a946>
- Cheng, C.W., Leu, S.S., Cheng, Y.M., Wu, T.C. & Lin, C.C. 2012.** Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry. *Accident Analysis and Prevention* 48: 214-222.
- Qammaz, A.S.B. & AlMaian, R.Y. 2020.** A critical success factors model for effective implementation of risk management process in the construction projects. *Journal of Engineering Research* 8 (3): 50-70. <https://doi.org/10.36909/jer.v8i3.7877>
- OSHA technical manual, 2015.** Fall protection in construction. Sec. V, Chapter 4. [https://www.osha.gov/dts/osta/otm/otm\\_v/otm\\_v\\_4.html](https://www.osha.gov/dts/osta/otm/otm_v/otm_v_4.html)
- Hsiao, H. 2014.** Fall prevention research and practice: A total worker safety approach. *Industrial Health* 52 (5): 381-392.
- Kim, Y.C., Yoo, W.S. & Shin, Y. 2017.** Application of artificial neural networks to prediction of construction safety accidents. *Korean Society of Hazard Mitigation* 17 (1): 7–14.
- Zhou, Y., Su, W., Ding, L., Luo, H. & Love, P.E.D. 2017.** Predicting safety risks in deep foundation pits in subway infrastructure projects: Support vector machine approach. *Journal of Computing in Civil Engineering* 31 (5).
- Sarkar, S., Patel, A., Madaan, S. & Maiti, J. 2016.** Prediction of occupational accidents using decision tree approach. *Proceeding of the 2016 IEEE Annual India Conference (INDICON)*, Bangalore, India: 1–6.

- Cho, Y., Kim, Y.C. & Shin, Y. 2017.** Prediction model of construction safety accidents using decision tree technique. *The Journal of the Korea Institute of Building Construction* 17 (3): 294–303
- Katanalp, B.Y. & Eren, E. 2020.** The novel approaches to classify cyclist accident injury-severity: Hybrid fuzzy decision mechanisms. *Accident Analysis and Prevention* **144**: 105590.
- Amiri, M., Ardeshir, A., Fazel Zarandi, M.H.F. & Soltanaghaei, E. 2016.** Pattern extraction for high-risk accidents in the construction industry: A data-mining approach. *International Journal of Injury Control and Safety Promotion* 23 (3): 264-276.
- Mistikoglu, G., Gerek, I.H., Erdis, E., Usmen, P.E.M., Cakan, H. & Kazan, E.E. 2015.** Decision tree analysis of construction fall accidents involving roofers. *Expert Systems with Applications* 42 (4): 2256–2263.
- WHO World Health Organization, 2010.** External causes of morbidity and mortality, other external causes of accidental injury. *International Statistical Classification of Diseases and Related Health Problems (ICD), 10th Revision, ed. 2010, Chapter XX.*  
[https://www.who.int/classifications/icd/ICD10Volume2\\_en\\_2010.pdf](https://www.who.int/classifications/icd/ICD10Volume2_en_2010.pdf).
- International Labor Organization (ILO), 2014.** Theory of accident causes.  
<https://www.iloencyclopaedia.org/part-viii-12633/accident-prevention>.
- Hill, D.C. 2004.** *Construction Safety Management and Engineering*, American Society of Safety Engineers: 707. Des Plaines, Illinois, U.S.A.
- Quinlan J.R. 1993.** *C4.5: Programs for Machine Learning*, Morgan Kaufmann, Los Altos.
- Elsayad, A.M. & Elsalamony, H.A. 2013.** Diagnosis of breast cancer using decision tree models and SVM. *International Journal of Computer Applications* 83 (5): 19-29.
- Otero, F.E.B., Freitas, A.A. & Johnson, C.G. 2012.** Inducing decision trees with an ant colony optimization algorithm. *Applied Soft Computing*, 12 (11): 3615-3626.

- Konda, R. 2010.** Predicting machining rate in non-traditional machining using decision tree inductive learning (PhD. Thesis). Graduate School of Computer and Information Sciences, Nova Southeastern University, FL., U.S.A.
- DeFries, R.S. & Chan, J.C. 2000.** Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment* 74 (3): 503-515.
- Ayhan, B.U., Doğan, N.B. & Tokdemir, O.B. 2020.** An association rule mining model for the assessment of the correlations between the attributes of severe accidents. *Journal of Civil Engineering and Management* 26 (4): 315–330.
- Kang, K. & Ryu, H. 2019.** Predicting types of occupational accidents at construction sites in Korea using random forest model. *Safety Science* 120: 226–236.
- Akboga Kale, Ö. & Baradan, S. 2020.** Identifying factors that contribute to severity of construction injuries using logistic regression model. *Teknik Dergi* 31 (2): 9919-9940.
- SSI Social Security Institution of Turkey, 2017.** Annual of Statistics 2017. [http://www.sgk.gov.tr/wps/portal/sgk/tr/kurumsal/istatistik/sgk\\_istatistik\\_yilliklari](http://www.sgk.gov.tr/wps/portal/sgk/tr/kurumsal/istatistik/sgk_istatistik_yilliklari)