Random Forest Analysis of Occupational Accident Reports among Roofing Contractors

Ikechukwu S. Onuchukwu[1] and Behzad Esmaeili, Ph.D.[2]

[1]Graduate Research Assistant, Sid and Reva Dewberry Dept. of Civil, Environmental, and Infrastructure Engineering, Volgenau School of Engineering, George Mason Univ., Fairfax, VA (corresponding author). Email: ionuchuk@gmu.edu
[2]Assistant Professor, Sid and Reva Dewberry Dept. of Civil, Environmental, and Infrastructure Engineering, Volgenau School of Engineering, George Mason Univ., Fairfax, VA.
Email: besmaeil@gmu.edu

ABSTRACT

One of the trades most susceptible to occupational hazards in the construction industry is roofing contractors, whose projects inherently involve exposure to falls from height. In response, this study used a random forest data mining technique to analyze the impacts of accidents' contributing factors on roofers' injuries. The analysis examined over 600 incidents obtained from the Occupational Safety and Health Administration's (OSHA's) database of fatal and nonfatal accident reports. Some of the contributing factors considered include source of injury, cause of injury, project cost, project end use, project type, injured body part, and day of injury. The results of validated random forest model revealed that the most important factor for predicting the nature of injury is injured body part followed by source of injury. The presented results can be used by managers, policymakers, and safety professionals to reduce the frequency and severity of incidents.

INTRODUCTION

On average, a roofing contractor is about three times more likely to be exposed to fatal injuries than other construction workers (Moore & Wanger, 2014). This risk level can be attributed to environmental exposures and the dynamic nature of work in the construction industry. Besides regularly performing work at heights, additional factors increase these workers' risks on jobsites, including the type of work, the composition of workers, and defiance of work-safety regulations (Choi et. al., 2006). Since a roofing contractor's job depends on balance control, a reasonable proportion of these risks manifest in the nature of the work, as it is challenging to maintain good balance control when working on inclined surfaces, and a loss of balance can easily lead to a fall injury. These challenges are especially compounded when workers handle heavy and bulky material—as seen in roofing projects. Exposure to adverse weather conditions only makes the situation worse and riskier. All of these add up to make roofing one of the most hazardous trades in the construction industry (Fredericks et al., 2005; Dong et al., 2013).

In a report by Center for Construction Research and Training (CPWR, 2013), between 2008 and 2010, specialty trade contractors experienced the highest number of fatal falls in the construction industry, amounting to 579 deaths. Unlike other specialty trade contractors that have recently experienced a declining record of injuries and fatalities, roofing contractors have experienced a gradual rise in their number of fall fatalities between 2011 and 2017 (Bureau of Labor Statistics, 2018). Furthermore, out of all fatalities recorded in residential building construction, 48.7% (135 deaths) were the result of falls. Even among nonfatal accidents, injuries can be costly and debilitating. On the average, the rate of nonfatal occupational injuries among roofing contractors were 1.1 to 1.8 times more than that of other construction workers for the period of 1992 to 2001. For the same period, the rate of fatal occupational injuries among roofing contractors were 1.6 to 2.8 times more than that of other construction workers (Sa et al., 2009). In 2005, the Bureau of Labor Statistics (BLS, 2007) reported that the injury cases among roofers is about 2 times more than that of other construction workers. Moreover, according to BLS (2006), the risk of nonfatal occupational injuries is still higher among roofing contractors than other construction workers on the average. These records highlight the need to investigate the causal factors of occupational incidents among roofing contractors.

In response, after building and validating a random forest (RF) classifier, this study analyzed both fatal and nonfatal accidents among roofing contractors to investigate the impacts of accidents' contributing factors on roofers' injuries. The RF algorithm was used in this study because of its renowned ability to identify dominant predictor variables for predicting the target variable (the nature of injury in this study). Using descriptive statistics, the roles of these predictor variables may be unseen or underexplored, particularly when the patterns are non-linear. Additionally, given that most of the variables considered here are categorical, this supervised data mining technique was employed because it is popularly used to uncover hidden patterns in categorical data. The results provide practitioners with insights into the nature of injury among roofing contractors, opportunities for designing specific training schemes, high-risk factors, and the relevance of incorporating safety during design to mitigate and manage risk for these workers.

BACKGROUND

Workers' injuries and fatalities in the construction industry are mainly caused by fall accidents (e.g., Stern et al. 2000; Dong et al. 2013). A study conducted by Stern et al. (2000) and involving 11,144 members of the UURWAW (The United Union of Roofers, Waterproofers, and Allied Workers) identified that fatalities significantly occur as a result of falls. Furthermore, Occupational Safety and Health Administration (OSHA, 2017) confirmed falls were the main cause of fatalities among the 'Construction's Fatal Four'—falls, struck-by object, electrocutions, and caught-in/between.

Kaskutas et al. (2009) discovered that certain fall protection systems were not used by small construction firms, and Olbina et al. (2011) confirmed that small construction firms also did not provide a safe work environment for their workers. In an analysis of records from the Census of Fatal Occupational Injuries (CFOI) involving 20,498 deaths in the construction industry for the period of 1992 to 2009, Dong et al. (2013) discovered that one-third of the fatal falls were roof fatalities, and 67% of them occurred within small construction establishments that have between 1 to 10 employees. The findings of their study can provide practitioners with relevant information on fall protection strategies and potential risk factors. However, their study did not focus on roofing as a specialty trade but instead focused on fatal falls from roofs of construction projects. It also focused on fatal injuries only, without considering incidents with nonfatal injuries, which are more common among roofers.

Many of these studies focused on fall accidents without considering other types of incidents that roofing contractors face, such as struck-by, caught-in/between, and electrocution—many of which may also contribute to falls. Additionally, the data used in these studies were from all trades in the construction industry, including large general contractors and big residential project construction. They were not specific to roofing contractors and hence may produce results that are not relevant to such small specialty trades as roofing contractors. Therefore, it is necessary to analyze the nature of injury with data specific to roofing contractors.

A limited number of studies have analyzed the nature of accidents among roofing contractors. A study by Fredericks et al. (2005)—based on the BLS's safety and health statistics database—explored the relevance of certain injury factors, including causes of injuries, types of injury, and event types. The study focused on roofing contractors' incident reports for the period of 1999 to 2000. Though valuable, the study faced limitations in that (1) the results were not based on the actual accident history but on survey responses; (2) the survey was limited to participants in Michigan; and (3) the study did not apply inferential statistics but used only descriptive statistics.

Hung et al. (2009) conducted a study to examine the training needs and preferences specific to small roofing contractors by distributing surveys among 20 roofing subcontractors. In their study, they applied a criterion sampling technique for choosing quality samples in the survey, and they recommended essential training needs for roofers, particularly raising the importance of sufficient worker's training, fall hazard awareness, and the use of suitable safety training techniques. The study was expanded by Hung et al. (2013) to involve 29 semi-structured interviews with residential roofing subcontractors to provide insights into the development of fall protection training. The study recognized the importance of supplementing informal jobsite safety training with formalized safety training, and it provided suggestions on fall protection adapted design.

Using survey information obtained from 252 roofers in the Midwest (Michigan, Indiana, Illinois, Wisconsin, and Iowa), Sa et al. (2009) compared fall-from-height accidents experienced by commercial roofers with those experienced by residential roofers. Their result showed that residential roofers were more likely to be involved in a fall accident than commercial roofers due to the lack of use of fall protection devices and their enforcement.

The major limitation faced by these studies is that they only applied descriptive statistics in their investigation without supplementing such analyses with inferential statistics. Since the relationships between accident-causation factors can be nonlinear and can include higher-order interactions, descriptive statistical techniques would fail to reveal hidden patterns in the data. Therefore, highly sophisticated inferential statistical techniques are required to handle large amounts of high-dimensional data to reveal hidden patterns that may inform safety practitioners' decisions on how to mitigate the risk of injuries and fatalities.

Random Forest Data Mining for Statistical Analysis

One beneficial inferential statistical technique is random forest, which is a supervised, simple, and classical (Izquierdo-Verdiguier & Zurita-Milla, 2020) ensemble learning algorithm renowned for its wide use in various applications (Breiman, 2001). The popularity of the algorithm can be mainly attributed to its ability to efficiently solve nonlinear classification problems and data imbalances in different classes, particularly in large datasets (Liu, 2009).

The RF classifier requires only two parameters to develop a prediction model. These parameters are: 1) the number of trees (N_{trees}) in the forest; and 2) the number of predictor variables (m_{try}) used at each node to develop the tree. Hence, to predict a class of new data, the data is classified by aggregating the prediction of N_{trees} number of trees with the use of a userdefined constant number, m_{try}, of randomly selected predictor variables to determine the split criteria in each node. The new data are assigned to the class having the highest number of votes from each of the N_{trees} trees in the forest. For classification, the default value of m_{try} is the square root of the number of predictor variables and it is recommended to try the default, half of the default, and twice the default value of m_{try} and then pick the best result (Liaw & Wiener, 2002). The number of trees, N_{trees}, is normally set to a few hundred trees (e.g., 500 trees as in this study), since more trees may not necessarily result in a better performance and may only slow down the processing time (Izquierdo-Verdiguier & Zurita-Milla, 2020). RF does not lead to overfitting as the number of trees increases because additional trees produce a limiting value of the generalization error (Breiman, 2001). Hence, more trees may not improve the result beyond a certain limit.

The RF algorithm (for both classification and regression) carries out three steps (Liaw & Wiener, 2002). First, it generates the N_{trees} bootstrap sample of the original training dataset. Second, it grows the N_{trees} number of unpruned trees with each of the bootstrap samples. Additionally, instead of choosing the best node-splitting variable from all the predictor variables in each bootstrap sample, the algorithm randomly selects m_{try} number of variables from all the predictor variables and determines the best splitter from this m_{try} shortlist of predictor variables. The variable chosen as the best node splitter and the split threshold are easily and quickly determined with the use of the Gini Index (GI), which minimizes the probability of misclassification by:

$$GI = 1 - \sum_{j=1}^{k} (P_j)^2 \tag{1}$$

where k is the number of classes and P_j is the probability of class j (Izquierdo-Verdiguier & Zurita-Milla, 2020). Third, the RF algorithm predicts the class of new data by a majority vote of the aggregated prediction results of the N_{tree} individual trees in the forest. These provide insights users can apply to assess their data.

POINT OF DEPARTURE

Roofing is one of the riskiest trades in the construction industry, with the most frequent accident among roofing contractors being falling from a roof. In fact, roofers are more likely to be involved in a fall accident than other workers in the construction industry (Huang & Hinze 2003). It is therefore very important to analyze the contributing factors that influence the nature of roofing contractor's injury, especially via inferential statistical techniques that may reveal hidden patterns in the data.

With respect to the literature review, two major limitations were observed among previous studies relating to the occupational safety and health of roofing contractors: (1) some studies focused on fatal injuries only, without considering incidents involving nonfatal injuries; (2) many of the studies focused their analyses on fall accidents without considering other types of incidents roofing contractors experience industry-wide, as can be found in a large database. To address these limitations, this study departs from the current body of knowledge by focusing on all types of roofer's incidents and by applying data mining techniques to the dataset to identify and predict factors affecting accident outcomes.

RESEARCH METHODOLOGY

Incident Database

To analyze accidents among roofing contractors, this study used data about fatal and nonfatal injuries collected during the period of January 1, 2007 to December 31, 2013 in the Occupational Safety and Health Administration's (OSHA's) Integrated Management Information System (IMIS) accidents database. The factors or variables considered in this study included *source of injury*; *cause of injury*; *project cost*; *project end use*; *project type*; *day of injury*; and *injured body part*. As the target/dependent variable, the *nature of injury* included the following categories: bruises/contusions, burns, concussions, cuts/lacerations/puncture, electrocutions/electric-shocks, fractures, non-specified injuries/disorders, and others.

In the preprocessing stage, the categories adopted in this study followed the Occupational Injury and Illness Classification Manual (OIICM), developed by the U.S. Department of Labor Bureau of Labor Statistics (BLS, 2012). After processing the data, 699 accident reports were obtained and used for analysis. Consistent with similar studies involving similar data mining techniques, these 699 accident reports were split into training (80%, 560 observations) and testing/validation (20%, 139 observations) datasets (Mistikoglu et al., 2015; Gholizadeh et al., 2021).

The RF model in this study was built from the 560 accidents reports that formed the training dataset and from seven project features/information/variables/attributes, namely: *source of injury*, *cause of injury*, *project cost*, *project end use*, *project type*, *day of injury*, and *injured body part*. The *nature of injury* dependent variable broke into two categories: fracture and non-fracture. The *nature of injury* involving a fracture was labelled "fracture," and the rest were regarded as non-fractures. Injuries within the body system and head/neck injuries were considered fragile body parts, while the rest were regarded as non-fragile body parts.

The randomForest package (Liaw & Wiener, 2002) in R (R Core Team, 2013) was used to build a RF model for: (1) predicting the nature of injuries resulting from an accident during a roofing project; and (2) identifying the factors that are most important for predicting the nature of injuries of roofers.

Random Forest Feature Selection Method

Decision trees have a famous ability to select important variables, and this ability is inherent in RF, which is an ensemble of trees. Thus, the randomForest package (Liaw & Wiener, 2002) offers the variable importance option (Breiman, 2001) used to identify the most relevant predictor variables for a given problem. The algorithm estimates variable importance tree-by-tree as the RF is developed. It calculates how much the prediction error increases by permuting the data for that variable while other variables are left the same (Liaw & Wiener, 2002). According to Izquierdo-Verdiguier and Zurita-Milla (2020), the importance of predictor variable x_i can be defined as:

$$Importance_i = \frac{1}{N_{trees}} \sum_{v \in S} G(x_i\, , v) \tag{2}$$

where S represents the set of nodes in which the predictor variable, x_i is used for splitting the samples. $G(x_i, v)$ is the RF gain of x_i, and it is computed by means of impurity measures when splitting the samples at each node. The function $G(x_i, v)$ can be computed using:

$$G(x_i, v) = GI(x_i, v) - \omega_R GI(x_i, v_R) - \omega_L GI(x_i, v_L) \quad (3)$$

where GI represents the Gini index, and ω_R and ω_L are the sample proportions in each node. Generally, impurity measures the extent to which the samples are correctly classified in each node (Gholizadeh et al., 2021). Its lowest value indicates that the node only includes one class of independent data. This study capitalized on these values to build a predictive model that helps to identify important features for predicting the target variable which is the nature of injury.

RESULTS

Random Forest Model Accuracy

To confirm the RF-identified variables of interest were predictively accurate, the developed RF model was evaluated by using the model to predict the outcomes of the 139 accident reports reserved for model validation. The results of this evaluation can be seen in the confusion matrix in Table 1. The confusion matrix compares the actual classifications of the nature of injuries and the model's predicted nature of injuries based on the accident reports. The major diagonal values of the confusion matrix in Table 1 indicate that 101 accident reports were classified correctly. This means that out of the 139 roofing accident reports in the validation/testing dataset, 101 accident reports (amounting to 72.7%) were classified correctly.

Table 1. Confusion Matrix of the Testing Dataset.

	Actual/Reference	
Prediction	**Fracture**	**Non-fracture**
Fracture	48	18
Non-fracture	20	53

Hence, as seen in Table 2, the proposed RF model can predict the nature of injury (as fracture or non-fracture) of the roofers' accident reports in the validation/testing dataset with a prediction accuracy of 72.7%. Furthermore, the 95% confidence interval in Table 2 indicates that there is a 95% confidence that the true accuracy of this proposed RF model lies between 64.5% and 79.9%. On the other hand, 38 accident reports out of a total of 139 (amounting to 27.3%) were misclassified. This outcome results in a misclassification rate of 27.3%.

Table 2. Evaluation of the Confusion Matrix and Random Forest Accuracy.

Evaluation Statistics	Results
Accuracy	0.727
95% Confidence Interval	(0.645, 0.799)
Precision	0.708
Specificity	0.704
Sensitivity/Recall	0.750
Kappa	0.454

The confusion matrix in Table 1 specifies there were 48 true positives (i.e., accident reports that involved actual fracture related injuries and that were predicted as fracture related injuries), 53 true negatives (i.e., accident reports that involved actual non-fracture related injuries and that were predicted as non-fracture related injuries), 18 false positives (i.e., accident reports that involved actual non-fracture related injuries but that were predicted as fracture related injuries), and 20 false negatives (i.e., accident reports that involved actual fracture related injuries but that were predicted as non-fracture related injuries). Using these outcomes, one can compute the precision, specificity, and sensitivity/recall resulting from using the proposed RF model to predict the accident reports in the testing dataset, as shown in Table 2. The evaluation results (e.g., accuracy of 72.7%) in Table 2 indicate that the RF model in this study is reasonable and consistent with similar studies that used a similar data mining technique such as Classification and Regression Trees (CART) analysis (Gholizadeh et al., 2021), C5.0 algorithm, and Chi-squared Automatic Interaction Detection (CHAID) algorithm (Mistikoglu et al., 2015).

Another evaluation statistic of the model reported in Table 2 is the kappa statistic, which indicates how well the proposed RF model prediction matches the actual classifications. The kappa value of the proposed RF model is 0.454, which represents a moderate score (McHugh, 2012).

Evaluation of Variable Importance

Table 3 lists the order of importance of the project variables/information/features/factors used in this analysis, which the authors computed by estimating how much the prediction accuracy decreased when permuting the data for a particular variable/information/feature/factor while holding other variables unchanged (Liaw & Wiener, 2002)—in other words, by estimating the impact on the prediction accuracy of excluding only the data for a particular variable. Table 3 demonstrates that the first-level attribute—or most important attribute—is the *injured body part* while the second-level attribute is the *source of injury*. The least important attribute for predicting the *nature of injury* in roofing projects is the *day of injury*.

DISCUSSIONS

The RF model in this study was developed to predict the nature of injury among roofing contractors and to determine the most important factors/features for the prediction. These contributing factors provide safety managers with the information needed to strategically assign scarce safety resources, especially in small firms. The RF algorithm was used to: group the accident reports of roofing contractors into the two classes of the dependent variable (nature of injury), namely fracture and non-fracture; and determine the level of importance of the contributing features for determining nature of injury.

Table 3. Variable Importance of the Proposed Random Forest Model.

Rank	Variables/Attributes	Mean Decrease in Accuracy
1	Injured body part	46.0
2	Source of injury	12.4
3	Project type	9.2
4	Cause of injury	8.5
5	Project cost	5.9
6	Project end use	2.9
7	Day of injury	0.5

As seen in Table 3, the *injured body part* is the most essential factor for predicting the nature of injury of roofing contractors. It can also be seen that if the data for *injured body part* are permuted while other features remain the same, the mean decrease in the percentage of the accuracy would be about 46 percent. In other words, with the exclusion of only the factor labelled *injured body part*, the percentage of the accuracy of the proposed RF model will be decreased by 46 percent on the average. Therefore, the feature/factor/variable labelled *injured body part* has a very high level of importance in terms of contributing to the accuracy of the proposed RF model. Thus, almost half of the predictive strength of the RF model is contributed by the project information/factor/feature labelled as *injured body part*, while the remaining half is contributed by all other project information/factors/features considered in this study. Such information is helpful for practitioners as this study's outcomes suggest that during workers' site meetings and roofing contractors' safety trainings, it is essential to emphasize the use of personal protective equipment that would sufficiently and effectively protect the body from injuries on roofing jobsites. Highlighting the importance of protective coverings and how to use them to safeguard roofers on jobsites cannot be overemphasized.

Furthermore, the contribution of the second most important feature/factor—*source of injury*—to prediction accuracy is also relatively significant. This contribution is because when the data for *source of injury* is permuted while other features remain the same, the mean decrease in accuracy would be 12.4 percent, as seen in Table 3. Hence, the *source of injury* (including machinery, parts/materials, structures/surfaces, tools/instruments/equipment, and vehicles) appears to be the second most important factor for roofers' accidents and therefore necessitates constant emphasis during site meetings and safety trainings for roofers.

The variables that are towards the bottom of Table 3 are not so important as they have minimal impact on the accuracy of the RF model and therefore have minimal contributions to the predictive strength of the model. Therefore, the variables at the bottom of Table 3 are less important compared to the variables at the top since their exclusion has minimal impact on the accuracy of the model. Table 3 indicates that the *day of injury* has the least impact on the predictive accuracy of the proposed RF model with respect to its contribution to the predictive strength of the model. In other words, with the exclusion of only the factor labelled *day of injury*, the percentage of the accuracy of the RF model will be decreased by 0.5 percent on the average. The information derived from the feature importance list in Table 3 could help safety managers understand where to better allocate limited resources to optimally mitigate the risk of injuries in roofing jobsites.

CONCLUSIONS

The work environment of roofing contractors in the construction industry is highly hazardous in nature and hence very prone to a lot of serious injuries and fatalities. This study proposed a RF model for predicting the nature of injuries resulting from an accident during a roofing project and identifying the factors that are most important in predicting the nature of injuries of roofing projects. It explored the possibility of using an RF algorithm to predict the nature of injury (as fracture or non-fracture) of roofing contractors based on such project information as *source of injury*, *cause of injury*, *project cost*, *project end use*, *project type*, *day of injury*, and *injured body part*. The research results identify *injured body part* as the most important factor for predicting the nature of roofers' injuries. Future studies could examine other factors—such as time of the accident, and age and sex of the employee—to explore additional patterns visible in accident outcomes.

Future extensions of this research question may consider other machine learning tools, such as decision trees, as well as other, more recent incident reports to search for other contributing factors affecting roofers or to assess recent developments within the industry. One limitation of the study is that an RF model does not give detailed information regarding the various categories or constituents that make up the variables (e.g., the categorical variables, such as *injured body part*, as in this study). For example, by applying CART to classify the variables, one can obtain a single comprehensive tree-like pictorial display that provides a streamlined result that is easy to interpret. Such continuing similar studies would help practitioners better understand the factors leading to injuries among roofing contractors to minimize the severity, frequency, and risk of incidents.

REFERENCES

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
BLS (Bureau of Labor Statistics). (2006). Fatal occupational injuries by industry and event or exposure. <http://www.bls.gov/iif/oshcfoi1.htm>, (June 7, 2021).
BLS (Bureau of Labor Statistics). (2007). Roofers. <https://www.bls.gov/oco/ocos212.htm> (May 7, 2021).
BLS (Bureau of Labor Statistics). (2012). Occupational Injury and Illness Classification Manual (OIICM). <https://www.bls.gov/iif/oiics_manual_2010.pdf>, (March 5, 2019).
BLS (Bureau of Labor Statistics). (2018). Census of fatal occupational injuries. <https://www.bls.gov/news.release/pdf/cfoi.pdf> (April 18, 2021).
Choi, S. D., Griinke, D., and Lederer, M. (2006). Fall protection equipment effects on productivity and safety in residential roofing construction. *Journal of Construction Research*, 7:149–157.
CPWR - The Center for Construction Research and Training. (2013). Fatal and Nonfatal Injuries from Falls in Construction. The Construction Chart Book. <https://stopconstructionfalls.com/wp-content/uploads/2013/07/Fatal-and-Nonfatal-Injuries-from-Falls-in-Construction-2013-update.pdf>.
Dong, X. S., Choi, S. D., Borchardt, J. G., Wang, X., and Largay, J. A. (2013). Fatal falls from roofs among U.S. construction workers. *Journal of Safety Research* 44, 17–24.
Fredericks, T. K., Abudayyeh, O., Choi, S. D., Wiersma, M., and Charles, M. (2005) Occupational Injuries and Fatalities in the Roofing Contracting Industry. *Journal of Construction Engineering and Management*, 131(11): 1233-1240.
Gholizadeh, P., Onuchukwu, I. S., and Esmaeili, B. (2021). Trends in Catastrophic Occupational Incidents among Electrical Contractors, 2007–2013. *International Journal of Environmental Research and Public Health*, 18(10), 5126.
Hinze, J., Huang, X., and McGlothlin, J. D. (2002). Review of fall accidents in construction. In *Proceedings of the 10th International Symposium on Construction Innovation and Global Competitiveness* (Eds.: B. Uwakweh and I. Minkarah). FL: Boca Raton, CRC press, 1117–1132.
Hung, Y. H., Winchester, W. W., III, Smith-Jackson, T. L., Kleiner, B. M., Babski-Reeves, K. L., and Mills, T. H., III. (2013). Identifying fall-protection training needs for residential roofing subcontractors. *Applied Ergonomics*, 44(3), 372-380.

Hung, Y. H., Winchester, W., Smith-Jackson, T., Mills, T., Kleiner, B., and Babski-Reeves, K. (2009). Fall Protection Training: Needs Analysis for Small Residential Roofing Subcontractors. *Human Factors and Ergonomics Society Annual Meeting Proceedings* 53(20), 1632-1636.

Izquierdo-Verdiguier, E., and Zurita-Milla, R. (2020). An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 88,1-13.

Kaskutas, V., Dale, A. M., Nolan, J., Patterson, D., Lipscomb, H. J., and Evanoff, B. (2009). Fall hazard control observed on residential construction sites. *American Journal of Industrial Medicine*, 52(6), 491–499.

Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18-22.

Liu, X.-Y., Wu, J., and Zhou, Z. H. (2009). Exploratory undersampling for class-imbalance learning. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, 39(2), 539–550.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica (Zagreb)*; 22(3):276-282. PMID: 23092060; PMCID: PMC3900052.

Mistikoglu, G., Gerek, I. H., Erdis, E., Usmen, P. E. M., Cakan, H., and Kazan, E. E. (2015). Decision tree analysis of construction fall accidents involving roofers. *Expert Systems with Applications*, 42, 2256–2263.

Moore, J. R., and Wanger, J. P. (2014). Fatal events in residential roofing. *Safety Science*, 70, 262–269.

OSHA (Occupational Safety and Health Administration). (2017). Worker Deaths in 2017. Commonly Used Statistics <https://www.osha.gov/oshstats/commonstats.html>.

Olbina, S., Hinze, J., and Ruben, M. (2011). Safety in roofing: Practices of contractors that employ Hispanic workers. *Professional Safety*, 56(4), 44–52.

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Sa, J., Seo, D. C., and Choi, S. D. (2009). Comparison of risk factors for falls from height between commercial and residential roofers. *Journal of Safety Research*, 40(1), 1–6.

Stern, F. B., Ruder, A. M., and Chen, M. D. (2000). Proportionate Mortality Among Unionized Roofers and Waterproofers. *American Journal of Industrial Medicine*, 37, 478-492.