ANALYZING ACCIDENTS AMONG SPECIALTY CONTRACTORS: A DATA MINING APPROACH

by

Pouya Gholizadeh
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Civil, Environmental, and Infrastructure Engineering

Committee:

___________________________________	Dr. Behzad Esmaeili, Dissertation Director
___________________________________	Dr. Girum Urgessa, Committee Member
___________________________________	Dr. David Lattanzi, Committee Member
___________________________________	Dr. Wenying Ji, Committee Member
___________________________________	Dr. Ping Deng, Committee Member
___________________________________	Dr. Sam Salem, Department Chair
Date:______________________________	Fall Semester 2022 George Mason University Fairfax, VA

Analyzing Accidents Among Specialty Contractors: a Data Mining Approach

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Pouya Gholizadeh
Master of Science
University of Nebraska-Lincoln, 2017
Master of Science
Shahid Beheshti University, 2013
Bachelor of Science
Iran University of Science and Technology, 2010

Director: Dr. Behzad Esmaeili, Assistant Professor
Department of Civil, Infrastructure and Environmental Engineering

Fall Semester 2022
George Mason University
Fairfax, VA

DEDICATION

To my Everything, Soha!

To my parents, Mohammad and Soudabeh, for their unconditional love!

ACKNOWLEDGEMENTS

First and foremost, I am grateful to my advisor, Dr. Behzad Esmaeili, for everything he has done for me. When I started my Ph.D., I had very little research experience and analytical knowledge. However, Dr. Esmaeili and I had a very pleasant conversion before he offered me to work in his new research group at the University of Nebraska-Lincoln. He encouraged me to take statistical and machine learning courses through which I could find my passion for conducting analytical works both in my academic and professional endeavors. Dr. Esmaeili was very understanding of my numerous experiments on the ideas and always provided thorough feedback on my work. I am also appreciative for the several assistantship opportunities Dr. Esmaeili granted me which allowed me to concentrate on my work while gaining invaluable research and teaching skills.

I also would like to express my sincere gratitude to my committee members, Dr. Urgessa, Dr. Lattanzi, Dr. Ji, and Dr. Deng for making my defense an enjoyable event with their positive and constructive feedbacks. Thank you!

I have not seen my parents or my only brother, since that very cold night in December 2013 when my wife and I arrived in the U.S. But every time I call them, they give me the love and encouragement I need to keep going and have helped me in whatever manner they were able to. I'm forever grateful for my loving and supportive family, who can lift my spirits and bring a smile to my face even though they are thousands of miles away!

I would also like to thank all of my friends who were there for me over the past eight years. I feel very fortunate to have found so many wonderful friends who I hold close to my heart.

I also want to thank all the nice individuals at George Mason University and the University of Nebraska who assisted me with various administrative tasks, filling out numerous forms, and reviewing hundreds of documents. I appreciate their friendliness and professional help!

Last but certainly not least, I want to thank Soha Rezaei, my high school sweetheart, best friend, my partner and the woman I am incredibly lucky to call my wife. None of this would have been possible if Soha had not been by my side from the beginning, when I started preparing for college entrance exam around 20 years ago! Yes, she has been giving up her time for that long in order to help me succeed in my academic career. She is a brilliant artist with a creative mind, and I have no doubt that we will soon see some of her work in fantastic animated films. Thank you Soha, I love you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

ANALYZING ACCIDENTS AMONG SPECIALTY CONTRACTORS: A DATA MINING APPROACH

Pouya Gholizadeh, Ph.D.

George Mason University, 2022

Dissertation Director: Dr. Behzad Esmaeili

Despite technological and regulatory improvements and plentiful research in occupational safety, construction has remained one of the most dangerous industries in the U.S. and around the world. This is mainly due to many relatively small employers with limited safety personnel and budget, multi-employer worksites, the presence of numerous hazards, and a highly mobile workforce. The uncertainty behind these conditions, combined with the limited personal experience of safety practitioners, can lead to poor safety decisions. Together, such factors ultimately contribute to the high number of fatal and non-fatal injuries in the industry, and the loss of millions of dollars each year. Analyzing historical incidents to understand the causes and consequences of them has been one of the main ideas in safety research to reduce the quantity and severity of occupational injuries. Indeed, the significant amount of safety data being collected on construction sites—e.g., as accident reports—provides a valuable source of information

for researchers seeking to better understand construction accidents. Recent developments in advanced analytical methods and computational tools can further improve previous efforts and provide a more data-driven objective approach toward construction safety. To test this approach, three objectives are defined in this research. The first objective is to evaluate the cost of the injuries (a main consequence of accidents) among various scenarios to quantify and compare their financial impact on companies and society. This objective can help contractors better quantify the risks of a construction project/task by estimating the severity of potential accidents in monetary values. Furthermore, the proposed methods contribute to the current body of safety knowledge by assessing alternative hypothesis testing practices that do not require specific assumptions. The second objective is to utilize statistical tests and models to identify the most influential factors contributing to construction accidents. The proposed analysis/modeling approach can be applied among all specialty contracting companies to identify and prioritize more hazardous situations within specific trades. The proposed model development process also provides a framework for codifying data from accident reports and analyzing them through a multivariate logistic regression model. The last objective is to investigate the potential correlations among accident outcomes and propose a novel way to incorporate such correlations through building multi-label machine learning models. The results indicate that knowing the value of one accident outcome can significantly increase the probability of a correct prediction for another outcome. The results further show that a particular multi-label method (i.e., classifier chains) can capture these latent relationships among accident outcomes during model training and significantly improve the performance of the predictive models. This research employs robust data and analytical models to predict the outcomes of accident scenarios, reliably, using variables available on construction sites. It is expected that the findings of this study will provide valuable insight into accident patterns and consequences to safety practitioners and transform the way machine learning models are being utilized in safety studies.

CHAPTER ONE: INTRODUCTION

1.1 Motivation and Research Objectives

Construction is one of the most hazardous industries in the US and it continues to be responsible for a disproportionate number of work-related illnesses, injuries, and fatalities in the United States. Based on the North American Industry Classification System (NAICS), the construction industry can be divided into three main categories: construction of buildings; heavy and civil engineering construction; and specialty trade contractors. Among these categories, specialty trade contractors suffer the highest number of fatal injuries. According to the Census of Fatal Occupational Injuries (CFOI) from the U.S. Bureau of Labor Statistics (BLS) database, 7,501 fatalities occurred from 2012 to 2019 in the construction industry, of which approximately 60% (4,661) were related to specialty contractors (BLS, 2019). NAICS has further divided specialty contractors into 18 subcategories. Notably, between 2012 and 2019, 43% of all fatal injuries among specialty contractors occurred in only three of these subcategories: roofing, site preparation, and electrical trades. It is essential to recognize the most crucial procedures that support the health and welfare of workers in these three trades given the influence they have on the business and the high rate of injuries among them.

A myriad of research has been done on construction safety as a result of this alarmingly poor performance. However, the majority of safety studies rely on subjective, secondary, or compiled data (Prades Villanova 2014). On the other hand, over the past few decades, large construction companies and federal agencies such as Occupational Health and Safety Administration (OSHA) have amassed extensive databases of digital injury-related incidents, which offer a wealth of empirical data (Tixier 2015). By leveraging abundant historical accident data along with statistical and machine learning techniques, the goal of this study is to increase safety performance among specialty contractors in the construction industry.

To achieve this goal, the following objectives are defined:

Objective 1: quantify the impact of individual accident factors (e.g., accident type, project end-use) by statistically comparing the monetary cost of injuries associated to them. Despite the fact that several studies have estimated the cost of injuries across various trades, demographics, event types, injury sources, and injury natures, the majority of these studies only provided descriptive statistics without analyzing any inferential statistics regarding differences between various groups.

Objective 2: employ methodologies of descriptive and quantitative statistics (i.e., chi-square test of independence and Cramer’s V test) to identify the contributing factors most affecting occupational accident outcomes among specialty trades and propose a multi-variate logistic regression model to determine the more severe accident scenarios.

Objective 3: investigate the potential impact that accident outcomes have on one another in a machine learning context and implement multi-label algorithms to determine if this impact can improve the prediction performance of models in construction accident studies.

There are various connections between these three objectives. First, the required analyses to accomplish these objectives are carried out using the same data and variables. The ability to use the results of one objective as inputs to another objective can be facilitated by the existence of this shared database. Second, the links between accident attributes and the outcomes of construction accidents serve as the foundation upon which all objectives are established. In order to more accurately characterize the consequences of construction accidents, objective 1 incorporates a financial component to quantify the impact of accident attributes. Utilizing this enhanced understanding of accident outcomes, objective 2 looks into the substantial correlations between accident attributes and outcomes. The impact of the combinations of such significant accident factors on an accident outcome is then investigated using logistic regression and decision tree models. The statistical tests and models, developed for the second objective, shape the framework for the final objective. By exploiting the correlations between accident outcomes that were discovered in objective 2, the final objective can more specifically address the drawbacks of binary models in safety studies.

To meet these objectives, the research was divided into four complimentary segments. Each segment is discussed briefly in this chapter and in details is the next chapters of this dissertation.

1.2 Dissertation Organization

This project investigates historical accident reports through statistical tests and machine learning models to find patterns among construction accidents to improve decision-making among safety practitioners and ultimately reduce the risk of injuries in the construction industry. The following is an outline of how this dissertation is structured.

Chapter 1 gives an overview of construction safety and background to the study. The motivation and research objectives derived from the current safety practices and need for an objective and data-driven approach to investigate occupational accident factors and consequences among specialty contractors are also presented.

The financial impact of various accident types on construction enterprises and society is examined in Chapter 2. This chapter also provides robust methods of hypothesis testing on data available from construction sites.

Chapter 3 analyzes the relationship between several accident attributes and the degree of injury as the main outcome of occupational accidents. Furthermore, data mining methods such as decision trees are tested to predict the nature of injuries.

Chapter 4 investigates the application of logistic regression modeling in describing various accident scenarios.

Chapter 5 studies the correlations among accident outcomes and proposes a multi-label machine learning method to benefit from such correlations. Several ways to train machine learning methods and evaluate their performance are also discussed.

Chapter 6 summarizes the research findings and concludes the dissertation. It also discusses future research extensions and opportunities, as well as the limitations of the research.

CHAPTER TWO: COST OF OCCUPATIONAL INCIDENTS FOR ELECTRICAL CONTRACTORS: A COMPARISON USING ROBUST FACTORIAL ANALYSIS OF VARIANCE

2.1 Introduction

Occupational incidents are costly (Leigh et al., 1997; Leigh et al., 2000). According to a report published by Liberty Mutual, workplace injuries that caused workers to miss six or more days of work cost U.S. employers around $59.9 billion in 2014 (Liberty Mutual, 2018). Within the construction industry, even less-severe occupational incidents typically require medical treatment, disrupt business operations, and negatively affect productivity, morale, and value-added activities related to long-term coordination and planning. These factors often combine to lower profit margins and decrease firms' competitiveness (Goetsch 2013; Argilés-Bosch et al., 2014; Feng et al., 2015). In fact, previous studies have consistently estimated the cost of occupational incidents in the construction industry to amount to more than 5% of total project costs, which is a great burden not only for contractors but also for owners, users, and society as a whole (Everett and Frank, 1996).

Quantifying the cost of an occupational injury in construction is a major step toward understanding the impact of an incident on project performance (Hinze and Applegate, 1991). Numerous studies have investigated the cost of occupational injuries and illnesses by statistically estimating the cost of injuries for different construction trades or demographics (Hinze and Applegate, 1991; Miller and Galbraith, 1995; Leigh et al., 1997; National Safety Council, 1999; Tang et al., 2004; Hinze et al., 2006; Leigh et al., 2006; Feng et al., 2015). While such studies yield valuable insights, their implications for contractors within trades are limited, since these studies do not provide statistical comparisons of the cost of injuries according to the type of accident (e.g., fall, struck-by) or project (e.g., building or non-building). As safety training can influence the potential for and outcomes of different accidents, understanding the costs associated with different types of accidents better prepares decision makers to allocate limited training and safety resources to those activities that might result in severe incidents. Such information would especially strengthen safety and financial outcomes among small contractors, whose narrower profit margins drive them to hire temporary workers with inadequate training or supervisors with poor safety-management skills (Cheng et al., 2010). Thus, a statistical cost analysis by accident type that can disaggregate risk factors into meaningful categories would foreseeably improve both project and safety performance by delivering decision-point data that can shape company policies.

In spite of the obvious benefits of such a statistical analysis, one reason there have been so few statistical comparisons of the cost of injuries according to accident type is that data about accident costs do not always disaggregate enough to meet traditional statistical assumptions. For example, common methods for comparing the mean of various samples often assume the normality of the samples' distributions and an equal variance among samples. Consequently, when these assumptions are violated, the analysis may be negatively affected (Keselman et al., 2002; Wilcox, 2012). Within the arena of construction safety, the non-normal distributions and the variability within and across construction project-samples often mean that these data do not satisfy statistical assumptions, making traditional statistical approaches to risk analysis ineffective at delivering sufficiently robust information to derive meaningful conclusions.

Beneficially, outside of the construction sector, statisticians have devised different procedures to manage non-normal populations and/or heteroscedastic variances in order to address such structural limitations within samples (White, 1980; Brunner et al., 1997; Long and Ervin, 2000; Wilcox and Keselman, 2003; Cribbie et al., 2007). These solutions range from transforming data and handling the outliers manually (Berry, 1987; Osborne, 2010) to considering more robust measures of location and scale to compare samples and better control for Type-I errors (Keselman et al., 1998; Luh and Guo, 2001). None of these techniques have ever been applied to analyze the cost of occupational injuries in the construction industry.

Accordingly, this study used more robust statistical methods (e.g., extensions to Welch and Yuen methods, and percentile bootstrapping) to disaggregate and compare the individual factors influencing the cost of different injury events within a single construction trade. To select a trade for analysis, this study has considered such primary criteria as the trade's safety performance and its portion of the construction industry. To measure safety performance, the authors evaluated the specialty trade contractors' classifications within the North American Industry Classification System or NAICS (Office of Management and Budget, 2017). Among the 19 categories of specialty contractors in this classification, roofing, site preparation, and electrical contractors accounted for more than 61% of all fatalities from 2011 to 2015 (Bureau of Labor Statistics, 2019). However, electrical contractors faced the largest increase in the number of fatalities in four years—from 44 in 2011 to 83 in 2015. Moreover, based on the Statistics of U.S. Businesses (SUBS), more than 144,000 establishments and firms were registered as electrical contractors in 2016, which is the second highest number of businesses among all specialty contractors (U.S. Census Bureau, 2016); comparatively, roofing and site preparation contractors were 37,108 and 68,651, respectively. Given the scope of the contractor pool as well as its large number of employees (i.e., more than 800,000 in 2016), the research team deemed this population a prime candidate for analysis. Furthermore, after reviewing OSHA's accident reports from 2007 to 2013, the research team noted that electrical contractors face a wide range of accident types such as electrocution, fall, struck-by, and caught in/between. Having a fair share of different accident types that can represent multiple accident scenarios in the data is crucial in quantifying the impact of each type. Thus, for all these reasons, this study selected electrical contractors as a case study for its robust statistical analysis of the costs of nonfatal injuries. Fatal injuries are excluded from this analysis for two reasons. First, as mentioned by Waehrer et al. (2007a), a fatality is estimated to cost $4 million whereas a non-fatal accident resulting in days away from work is estimated to cost only $42 thousand on average (almost 100 times less). Having one or two more fatalities in one category than the others would significantly shift average costs and thereby mislead conclusions about the severity level of each category. In statistical words, in a cost analysis, fatalities can act as outliers within the data set, impeding an ANOVA-type cost analysis, since these events happen much less frequently than non-fatal accidents and cost much more than any other injuries. This statistical concern in no way means that we should not consider fatal accidents when planning for safety; only, this consideration indicates why, in comparing cost of injuries among accident types, it is more reasonable to exclude the outlier-effect that fatalities might bring to the analysis. Second, the cost estimates used in this study were based on lost-time workers' compensation insurance claims, which are estimated based on non-fatal injuries (more on this in section 02.3.2 Estimating the cost of injuries).

The main objective of this study is to investigate the impact of common event types on the cost of an occupational accident among electrical contractors. In other words, the study aims to answer these questions: does the type of an accident play a role in determining the cost of the subsequent injury? More importantly, if there is a significant difference, does it exist among all event types or only among some specific pairs? Beyond event types, the authors were interested in analyzing two other factors—project end-use and project budget—because these two variables have been mentioned in similar studies and were found to have significant impacts on the outcome of construction accidents. For instance, when investigating the occupational injuries at small enterprises in Taiwan, Cheng et al. (2010) found that the type of project (e.g., building, road, bridge) is highly correlated with the type of accident. Furthermore, the authors determined that the number of occupational accidents is significantly higher in building projects than civil engineering projects in both low-budget (i.e., less than 5 million New Taiwan dollar) and high-budget (i.e., 5 to 50 million New Taiwan dollar) projects. As for the project budget, Feng et al. (2015) discussed that project characteristics (e.g., project size, contractor size, involvement of sub-contractors) can influence the indirect costs of accidents in construction building projects. Furthermore, they concluded that project size (i.e., the contract sum of a project) have a significant positive impact on indirect costs of accidents. Thus, these two factors have been shown to influence accident types and costs, though the interactions among all of these factors remains unknown. The factors and levels that have been used for the analysis are shown in Figure 2.1. Historical data about electrical contractor injuries were collected from the Occupational Safety and Health Administration (OSHA) Integrated Management Information System (IMIS) database, and the events were cross-referenced with cost-estimates of the average cost of lost-time workers' compensation insurance claims, which stood in for accident-cost data. These data were then coded during a content analysis. Subsequently, the research team applied three statistical approaches—Welch-type procedure, an extension of Yuen's method, and percentile bootstrapping—to disaggregate and compare the factors associated with the different accident types. These results yielded insights both in terms of the effectiveness of the analytical framework and the costs associated with different accident factors.

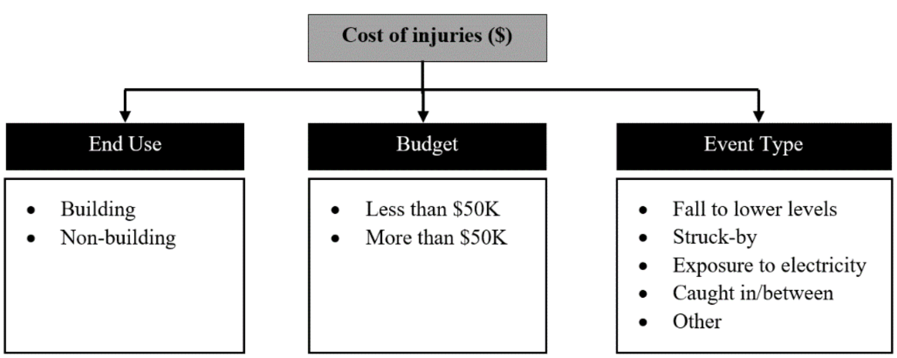


Figure 2.1. Factors and their levels

The success of this study provides three-fold benefits: The results of the analytical framework proposed in this study demonstrate the effectiveness of statistical methodologies for analyzing the cost of injuries within trades; such insights will aid researchers seeking to better address project performance within construction. Additionally, this study supports practitioners since the cost of injuries can be used as an indicator of potential risk associated with different activities. Trades, safety managers, project managers, or policy makers can use such statistically-derived cost knowledge to allocate limited safety resources more efficiently. Lastly, employers may use the outcomes of this study to better understand the financial implications of occupational accidents to make decisions that keep workers safe and companies profitable.

2.2 Background

2.2.1 Costs of Injuries

The costs of injuries are higher in construction than in other industries (Silverstein et al., 1998; CPWR, 2002). Because such costs are not value-added to businesses, construction companies are interested in measuring or predicting the cost of injuries for their projects (LaBelle, 2000; Rikhardsson and Impgaard, 2004; Sun et al., 2006). Understanding the costs of injuries helps these companies make better-informed decisions about their investments in occupational safety programs and/or in planning contingencies for projects based on certain characteristics (type, size, and complexity).

To efficiently allocate limited resources to various safety practices, construction companies need reliable information about the true cost of construction accidents. As a result, a number of studies have focused on the cost of injuries (Durbin, 1993; Miller and

Galbraith, 1995; Miller, 1997; Sun et al., 2006; Jallon et al., 2011a, b). A comprehensive literature review was conducted by the research team to identify the existing knowledge gap. The salient results of the relevant studies are provided here.

2.2.2 Direct versus indirect costs of injuries

Before one can analyze the costs of injuries, one must first understand the nature and relationship of injuries' direct and indirect costs. Direct costs, also known as insurable costs, refer to costs related to treatment of the injury and any compensation paid to workers as a result of being injured (Everett and Frank, 1996; Hinze, 1997; Feng et al., 2015). However, not all costs incurred as a result of an injury are insurable. Such costs include regular wages paid to an injured worker while his/her productivity is reduced; transportation costs to medical centers; loss of productivity due to activity disruption or distractions to other crew members; administrative efforts related to investigating and reporting the incident, hiring and training a replacement worker; and damage to materials or machinery (Hinze, 1997; Paez et al., 2006). All of these uninsurable costs due to an injury are called indirect costs, and while it is possible to calculate direct costs with reasonable accuracy, indirect costs are much more difficult to quantify.

In one of the earliest attempts to find a ratio between direct and indirect costs, Heinrich (1931) claimed that indirect costs can be approximately four times higher than direct costs. Since then, several large studies have been conducted to further examine the relationship between direct and indirect costs (e.g., Paez et al., 2006; Manuele, 2011). One such study in Great Britain surveyed 1,858 companies to create a database of injuries' costs to quantify the burden of injuries on contractors; this study found that the insured costs (total annual employers' liability premiums of £83.1M) can be much higher than uninsured costs (£7.3M). Even for companies that had experienced an accident, the premium costs on average were still more than four times greater than all the other costs incurred due to the accident (Leopold and Leonard, 1987). Such an outcome demonstrates that a straight ratio between direct and indirect costs may not be universal.

Other studies sought to determine the ratio between direct and indirect costs based on the severity level of the injuries. In their study, Hinze and Appelgate (1991) collected 573 surveys from construction companies and separated events according to those demanding medical/doctor cases and those demanding restrictive activity/lost workday cases. Using unweighted averages, they found that the ratio of indirect to direct costs was 4.2 in medical cases and 20.3 in restrictive activity/lost workday cases. Their observation emphasized the necessary distinction between injuries based on their severity: the ratio in medical cases was in accordance with Heinrich's 4-to-1 rule (Heinrich, 1931), but indirect costs were much higher in cases with more severe injuries. Notably, these ratios included claim costs as indirect costs, which had a great impact on the magnitude of the results; without claim costs, the ratios would be 2.94 and 2.53 for medical and restrictive activity cases, respectively. In another study that considered workers' compensation insurance (WCI) and experience modification ratings (EMR), Everett and Frank (1996) found that indirect costs can be 1.65 times greater than direct costs. These results underscore the interaction between the type of cost (direct and indirect) and the severity of an injury: while more severe injuries would lead to higher direct and indirect costs, the effect is stronger for indirect costs.

Some researchers went a step further and tried to quantify the indirect costs of additional resources that society would bear due to a construction accident (e.g., government subsidies paid during hospitalization, solicitor fees, idle machinery and equipment, and costs incurred by government organizations such as fire service and police). Specifically, by investigating 119 construction projects with 1,414 accidents (6 fatal and 426 resulting in permanent disability), Tang et al. (2004) quantified the social costs of accidents and examined the relationship between these types of costs and social safety investments. They found that for each dollar invested in social safety, social costs due to construction accidents could be reduced by $2.27.

As evidenced by these outcomes, calculating indirect costs involves making several assumptions, and to date, there is a lack of consensus among researchers about the exact ratio between direct and indirect costs (Health and Safety Executive, 1993; Monnery, 1998). Accordingly, for the analysis executed within this study, the research team decided to focus on medical costs and lost wages from each incident (i.e., WCI), which combine to represent the direct costs associated with an injury.

2.2.3 Cost of injuries among different event types, sources, and injury characteristics

To determine the cost of treating occupational injuries, researchers have historically examined workers' compensation claims (Durbin, 1993; Dement and Lipscomb, 1999; Silverstein et al., 1998; Islam et al., 2001). Dement and Lipscomb (1999) studied workers' compensation claims among North Carolina Homebuilders Association (NCHA) members and their subcontractors between 1986 to 1994; their study found that injuries resulting from ‘struck by an object,’ ‘lifting/movement,’ and ‘falls from a different level’ accounted for the highest rates for incidents that involved medical costs or paid lost time. When Dement and Lipscomb (1999) subsequently examined body parts, they determined that ‘back/shoulders,’ ‘fingers,’ and ‘leg/knee’ were the most frequent parts that resulted in medical costs or lost–work time cases.

In another study, Hinze et al. (2006) investigated medical injury records of non-significant injuries (i.e., those that would not result in days away from work) to explore the effects of these incidents on the construction industry in terms of medical costs and frequency. After collecting data from more than 135,000 construction accidents from 2001 to 2003, their analysis showed that laceration, lumbar spine, upper extremity, and eye injuries were together responsible for about 59% of all injuries. The results also showed that lacerations, eye injuries, and fractures occurred more often, on average, in the construction industry than in all other industries combined. The average cost of these small injuries was $565 (equivalent of $800 in 2019), with shoulders/humerus ($933; equivalent of $1,321 in 2019), lumbar spine ($887; equivalent of $1,256 in 2019), cervical spine ($866; equivalent of $1,225 in 2019), and knee ($818; equivalent of $1,157 in 2019) having the highest costs per patient. Such higher costs of shoulder injuries were attributable to the fact that many of these injuries were followed by surgeries and physical therapy services. One should note that these numbers are much lower than the averages provided by the National Council on Compensation Insurance (NCCI). An explanation for this difference is that the study by Hinze et al. (2006) investigated only small accidents.

Other studies compared classes of injuries. According to Leigh et al. (2006), while ‘back sprains and strains,’ ‘other sprains and strains,’ and ‘fractures, crushing, and dislocations (except head and neck)’ are the top three injuries in total costs (together accounting for more than $31.5 billion), the share of fatalities in these costs is close to zero. Lipscomb et al. (2006) investigated occupational injuries occurring during the construction of the Denver International Airport and found that while only 18% of all construction injuries were preceded by a slip or trip, these injuries comprised 25% of workers’ compensation payments. At the same time, insulation work had the lowest payment rates per $100 of payroll and roofing had the highest (Lipscomb et al., 2006).

Waehrer et al. (2007b) analyzed cost injuries for different event types and sources of injuries. In terms of event types, the highest total costs for days-away-from-work injuries were associated with ‘falls to a lower level’ and ‘overexertion.’ For per-days-away-from-work cases, those with the highest costs were ‘unspecified bodily conditions’ ($136,222), ‘contact with electrical current’ ($86,829), ‘repetitive motion’ ($75,254), and ‘caught in or compressed by equipment or objects’ ($69,041) (the study excluded pedestrian, non-passenger struck-by vehicle incidents, mobile equipment cases, and assaults and violent acts by person). Similar investigations of sources of injuries revealed that injuries due to floor surfaces (1 in every 5 sources) and building materials resulted in the highest total-days-away-from-work costs—together, these sources were responsible for 35% of injures and 39% of costs—while ‘unspecified machinery,’ ‘unspecified other sources,’ and ‘unspecified structures’ had the highest per-days-away-from-work costs.

One major limitation of the previous studies was that they simply reported descriptive statistics without analyzing any inferential statistics to determine whether there are statistical differences between average costs. The current study aims to address this limitation. Besides, none of the previous studies have examined project characteristics (e.g., type and budget) that can influence the direct cost of construction injuries.

2.2.4 Construction trades

Understanding differences between cost of injuries among different trades is important for policy makers and insurance companies because it helps them determine insurance premiums. Safety managers also can benefit from these data to conduct more accurate risk assessments of construction sites.

In response to this important need, Lipscomb et al. (2003) evaluated workers' compensation records for residential contractors to compare costs of injuries among union carpenters between 1995 and 2000. They found that the costs of injuries resulting from falls, raising framed walls, setting steel I-beams, and pneumatic nail guns is higher than other incidents. In another study, Waehrer et al. (2007a) analyzed Bureau of Labor Statistics (BLS) and Current Population Survey data from 2000 to 2002 to estimate the cost of fatal injuries and those nonfatal incidents that resulted in days-away-from-work based on worker occupation. On average, injuries in the construction industry accounted for a $12.7 billion loss, with $4 million for each fatality and $42,000 for each nonfatal injury resulting in days-away-from-work. The occupation-specific costs revealed that construction laborers, followed by carpenters, roofers, and electricians, have the highest costs due to fatalities. Among these groups, electrician apprentices have the highest cost per fatality, at $5.3 million. In the nonfatal category, mining machine operators and earth drillers account for the highest cost per days-away-from-work. Occupations experiencing the highest total cost of injuries—with a combined share of more than 50% of costs—are, in order: construction laborers; carpenters; electricians; plumbers, pipefitters, and steamfitters; and roofers.

In a similar study, Waehrer et al. (2007b) used the standard industrial classification (SIC) system to compare injury costs among different trades within the construction industry. The results showed that in 2002, the construction industry accounted for 15% of injury costs while representing only about 5% of the workforce of the U.S. The construction industry's share of fatal costs was around 40%, virtually double the 21% found by Leigh et al. (2006) for all industries. More than half of the injury costs in the construction industry were found to occur to only five trades: miscellaneous special trade contractors (SIC 179); plumbing, heating, and air-conditioning (SIC 171); electrical work (SIC 173); heavy construction except highway (SIC 162); and residential building construction (SIC 152). In construction, on average, per-case costs of fatalities, nonfatal days away injuries, lifetime wage losses, restricted-work cases, and no-lost-work cases were estimated at $4 million, $42,000, $21,600, $618, and $777, respectively. The roofing, siding, and sheet metal work industry (SIC 176) was on top in terms of the rate of accidents, with 9.5 cases per 100 full-time equivalent employees.

As one can see, previous studies have examined overall cost of injuries among different trades; however, little is known about the cost of injuries in a specific trade (e.g., electrical contractors). Each trade faces specific accident scenarios that might result in considerably different combinations of injuries. Even similar events can lead to varied levels of severity in contrasting trades—for instance, struck-by accidents can yield more severe injuries for on-site preparation contractors than for electrical contractors as the injuries among on-site preparation contractors result from heavy machinery and vehicles. Thus, the current study addresses this limitation by investigating several event characteristics within a single trade.

2.2.5 Analytical framework: Factorial Analysis of Variance (ANOVA)

While statistical analyses help researchers and practitioners interpret vast amounts of data, generally, such analyses assume sufficiently large samples to enable isolating variables of interest and thereby make relational statements about the isolated variables. However, not all data support such analyses: in the context of estimating the costs of construction injuries by accident type, assorted factors such as project size/budget, project end-use, and/or the nature of the injury can all inhibit isolating and comparing a single parameter to determine risk. Accordingly, alternative statistical tools must be applied to make comparisons across accident types meaningful for decision makers.

One approach to such comparisons is Analysis of Variance (ANOVA). When the response variable is continuous but all predictor variables (e.g., accident factors such as project budget or accident type) are *categorical* and manifest two or more *levels* (e.g., a budget of >$50k vs. a budget of <$50k), one can use ANOVA to compare multiple group means (Lee and Ahn, 2003; Fox and Weisberg, 2011; Field et al., 2012). In its simplest form—one-way ANOVA—this approach can compare multiple factors with only one level. In cases with more than one level, a factorial (or multiway) ANOVA design should be implemented (Fox and Weisberg, 2011). Factorial design helps one consider the interactions between the conditions at play in the dynamic—for instance, whether the effect of one categorical level on the response variable changes according to the level of the second factor. In this way, factorial ANOVA reveals opportunities for deriving meaning across multi-dimensional data sets.

Factorial ANOVA enables testing the main effects of one variable even within unbalanced samples. Consider a situation in which factors F1 and F2 each has two levels (e.g., ‘a’ and ‘b’ for F1 and ‘c’ and ‘d’ for F2) and the number of observations (for a, b, c, and d) are not equal. When the hypothesis demands comparing the main effect of F1’s levels ‘a’ vs. ‘b,’ the choice between one-way ANOVA and factorial ANVOA becomes important. Within a factorial framework, one can use *unweighted* means in which the mean of each level of a factor would be the average of two means without considering the sample sizes; this characteristic is especially beneficial when one level may be rare but the impact of its occurrence is very large—so the analyst would not want a weighted value to wash out the variable’s importance and hence they consider the hypothesis to be independent of the sample size (Quinn and Keough, 2002). In our example of ‘a’ vs. ‘b,’ to obtain the mean value of ‘a,’ one would calculate the average for ‘a’ under two conditions, ‘c’ and ‘d’—and, thus, having the information from F2 for the values of ‘c’ and ‘d’ would provide important data. Within a one-way ANOVA framework, to compare ‘a’ and ‘b,’ one would use the overall means of ‘a’ and ‘b’ without considering the second variable, F2. When the samples are unbalanced, the *unweighted* means for ‘a’ and ‘b’ would be different than the *weighted* ones, as would be their difference and the main effects’ sizes. While one-way ANOVA would only compare the *weighted* means, factorial ANOVA can compare the un*weighted* means, which leads to more reliable results when testing the main effect of each factor on the dependent variable (Dien and Santuzzi, 2004). Thus, factorial ANOVA can execute more accurate comparisons for unbalanced designs (i.e. unequal sample sizes).

Both one-way and factorial ANOVA have been successfully applied to various situations. In their review study, Keselman et al. (1998) mentioned ANOVA as the most common method of analysis in educational research. Dien and Santuzzi (2004) investigated the application of the repeated measures ANOVA on event-related potential data and concluded that multivariate (i.e., factorial) ANOVA can provide higher statistical power to discussions. While the literature supports applications of this type of analysis—especially in experimental contexts, where researchers have more control on the data-collection process—caution must be exercised when using this approach in the context of observational data, such as those found in the context of this study.

Though ANOVA is a popular method in many research fields (Keselman et al., 1998; Kenny et al., 1998), the conventional ANOVA F-test requires many assumptions that may be challenging to satisfy—such as independence of sample distributions, normality, and homogeneity of variances (i.e., constant variance) among sample groups. Failing to meet these assumptions results in higher Type-I error rates—those wherein one rejects the null hypothesis when it is true—and lower statistical power (Keselman et al., 1998; Keselman et al., 2002; Cribbie et al., 2012; Wilcox, 2012). Having a group of unbalanced samples can further complicate the analysis process (Montgomery, 2001, p. 600) and worsen the inflated Type-I error-rates problem. Thus, Lix and Keselman (1998) concluded that the use of conventional estimators (i.e., group means and variances) is not recommended in these situations. For instance, Wilcox (2012) reported that in the case of unequal variances, the actual Type-I errors could rise to 0.09 at the 0.05 level when testing on data with four groups with equal observations. With unequal sample sizes and under non-normality, however, controlling Type-I errors can be even harder. Moreover, Wilcox (1995) concluded that the effects of violating assumptions and/or the presence of outliers are also significant for Type-II errors (i.e., not rejecting null hypothesis when in fact it is false). Problematically, having equal sample sizes and meeting the assumptions are unlikely in observational studies such as cost-of-injury investigations.

In response, some classical statistical textbooks have suggested transforming data and subjectively removing outliers. However, studies have identified the limitations of these approaches. For instance, Wilcox (2012) reported that beyond the problems caused by interpreting the results of transformed data, transformation cannot always lead to less skewed distributions. Removing outliers and then using conventional methods is also not recommended, as the remaining observations are not independent anymore and can result in the wrong standard errors (Wu, 2002).

To this end, some studies have introduced a different idea: using more robust measures of location and scale—such as trimmed means and Winsorized variances—instead of means and variances to improve the control over Type-I errors (Keselman et al., 1998; Luh and Guo, 2001). To adopt these measures, previous studies have tested several heteroscedastic (i.e., those with no constant variance assumption) methods to determine the method that produces the best performance based on the data characteristics and the test design. For example, to compare the Type-I error and power-rates of the F-test among several robust methods (i.e., Welch's heteroscedastic F-test with and without trimmed means, James's second-order test, parametric bootstrap procedure, and trimmed parametric bootstrap procedure), Cribbie et al. (2012) conducted Monte Carlo simulations using multiple ANOVA designs. They suggested that using Welch test or parametric bootstrapping methods with trimmed means and Winsorized variances could control Type-I errors and provide power better than conventional F-tests using means and variances. In another study, Luh and Guo (2001) suggested a combination of Johnson's transformation and trimmed mean when ANOVA assumptions are violated, especially in cases with highly skewed distributions. Their proposal, however, was only tested on a specific design (i.e., two-way fixed-effect ANOVA). Lix and Keselman (1998) also compared six well-known methods to test the equality of location among different groups in the presence of heterogeneous variances. They reported that, although none of these methods were able to control the Type-I errors in all simulated conditions, the methods introduced by Alexander and Govern (1994), Box (1954), James (1951), and, to some extent, Welch (1951) could provide satisfactory results when using trimmed means and Winsorized variances. Wilcox (2012, p. 301) has referred to the work of Lix and Keselman (1998) and argued that while their results are valid in most situations, they still have limitations when the sample sizes are small; thus, Wilcox (2012) described a bootstrap method that could deal with this problem. Bootstrap methods have been recommended by some studies as approaches that could be used in conjunction with test statistics; such bootstrapping would be based on trimmed means and would consequently produce more reliable results under non-normality and heterogeneity of variance (Wilcox et al., 1998; Wilcox, 2012). Methods that are deemed most appropriate to analyze injury costs are discussed in more detail later in this section.

2.2.6 Robust measures of location and scale

A measure of location (e.g., mean, median) and/or scale (e.g., range, variance) is considered robust whenever "slight changes in a distribution have a relatively small effect on the values of measures" (Wilcox, 2012, p. 23). Since the sample mean and variance are not robust and could be affected significantly by outliers in skewed distributions (especially heavy-tailed ones), replacing these parameters with robust measures such as trimmed means and Winsorized variance could help one to control Type-I errors and achieve higher power even when the assumptions of the F-test are violated (Huber, 1981; Wilcox, 2005; Cribbie et al., 2012; Wilcox, 2012). Since these two measures and how they are calculated are critical for understanding the results of this study, a brief description is provided here.

<u>*Trimmed Mean*</u>*:* Let $\lambda = [\kappa n]$, where $[\kappa n]$ is the largest integer less than or equal to 'κn'. 'κ' represents the proportion of observations trimmed from each tail of the distribution and 'n' is the sample size. Then, $h = n - 2\lambda$ represents the effective sample size (i.e., the sample size after trimming). The sample trimmed mean for $x_i, (i = \lambda + 1, \dots, n - \lambda)$, then could be measured as: $\bar{X}_t = \frac{1}{h}\sum_{i=\lambda+1}^{n-\lambda} \mathrm{X_i}$

Winsorized variance: Subsequently, sample Winsorized variance will be calculated using the following equation: $s_{win}^2 = \frac{\sum_i (\mathrm{Y_i} - \bar{X}_{win})^2}{n-1}$

Where $\bar{X}_{win}$ is the sample Winsorized mean, which represents the sample mean after replacing the trimmed observations in the lower and upper tails with the lowest and highest untrimmed observations respectively: $\bar{X}_{win} = \frac{1}{n}\sum_i \mathrm{Y_i}$; where

$$Y_i = \begin{cases} X_{(\lambda+1)} & if\ X_i \leq X_{(\lambda+1)} \\ X_i & if\ X_{(\lambda+1)} < X_i < X_{(n-\lambda)} \\ X_{(n-\lambda)} & if\ X_i \geq X_{(n-\lambda)} \end{cases}$$

As mentioned before, several methods have been developed to adopt these robust measures. Based on the fact that some groups with relatively small sizes were presented in this study, the authors have decided to adopt a combination of robust methods to test the hypotheses. These methods are as follows.

A Welch-type procedure

To address the violation of 'constant variance' assumption when testing $H_0: \mu_1 = \cdots = \mu_j; (for\ j = 1, \ldots, J)$, Welch (1951) suggested an alternative test statistic (W) as:

$$W = (J + 1)\, \Sigma[w_j\big(A\bar{x}_j - \sum w_j \bar{x}_j\big)^2]/[A^2(J^2 - 1) + 2(J - 2)B],$$

Where J is the number of groups. To calculate w_j, A, and B, the user lets s_j equate to the estimated standard deviation of group j, and calculate the values as below:

$$w_j = \frac{n_j}{s_j^2}; \qquad A = \sum w_j; \qquad B = \sum\left[\frac{\big(A - w_j\big)^2}{n_j - 1}\right]$$

These choices enable users to compare the values of W with critical F values and reject the hypothesis whenever W is larger than critical values of F. In this approach, the degrees of freedom are $(J\text{-}1)$ and $A^2 \times (J^2 - 1)/3B$. However, one may notice that this method can only address the one-way designs. To extend this test to two-way and three-way designs, Algina and Olejnik (1984) proposed a generalization to the work of Johansen (1980) wherein a matrix formula is proposed to calculate the critical values.

While Welch test (and its extensions to factorial designs) can address the violations of equal variances, studies have shown that when the sample distributions are not normal (e.g., heavy-tailed samples), the power of the test can be low (Wilcox, 2012). Consequently, replacing $\bar{x}_j$ with trimmed means ($\bar{x}_{tj}$) and s_j^2 with sample Winsorized variance (s_{winj}^2) can lead to more powerful tests that can better control Type-I errors.

Though this method can test the global main effects and interactions within a factorial design, it cannot perform the post-hoc hypothesis testing. Therefore, the next two methods have been proposed to test the pairwise comparisons.

An extension of Yuen's method to trimmed means

A very important effect of using trimmed samples is that the remaining observations are no longer independent from each other. Therefore, the variance of the trimmed sample cannot be used to obtain the standard errors that are needed for the t-test. To address this problem, when comparing only two groups, Yuen (1974) suggested that one can use Winsorized variances to calculate the standard error for the trimmed means as:

$$d_j = \frac{(n_j - 1)s_{winj}^2}{h_j(h_j - 1)}, (for\ j = 1, 2)$$

Therefore, to perform the pairwise comparisons between two groups with trimmed means (i.e., $\mu_{t1} - \mu_{t2}$) at the significance level α, the confidence intervals can be calculated as: $(\bar{X}_{t1} - \bar{X}_{t2}) \pm t\sqrt{d_1 + d_2}$, where t is the $1 - \alpha/2$ quantile of the Student's t-distribution with $\hat{v} = [(d_1 + d_2)^2] / \left[\frac{d_1^2}{h_1 - 1} + \frac{d_2^2}{h_2 - 1}\right]$ degrees of freedom. If, instead of one test, the goal was to use linear contrast codes to investigate multiple pairwise tests, one can update the hypothesis as: $H_0: \psi = \sum_{j=1}^{J} c_j \mu_{tj} = 0$, The estimate for the squared standard error can then be rewritten as: $d_j = \frac{c_j^2 (n_j - 1) s_{winj}^2}{h_j (h_j - 1)}$, and when we let $D = \sum d_j$, an estimate of the confidence intervals for ψ would be: $\hat{\psi} \pm t \sqrt{D}$.

One last note is that, here, with C contrast codes, t is the $1 - \alpha$ percentage point of the C-variate Studentized maximum modulus distribution (Dunnet, 1980). This extension would allow multiple pairwise tests when using a trimmed mean as the measure of location.

Percentile bootstrapping

Efron and Tibshirani (1994) defined the bootstrap method as "a computer-based method for assigning measures of accuracy to statistical estimates" (p. 10). All bootstrap methods are based on a simple idea that instead of assuming an underlying distribution for our samples, one can use the data itself and generate estimates solely based on the data. Krishnamoorthy et al. (2007) have applied a parametric bootstrap method on a data set using 5,000 simulations and concluded that this approach could result in better control over Type-I error rates and greater power than the Welch test, especially when there are many groups and the sample sizes are small. Among different bootstrapping methods, percentile bootstrapping, with a trimming amount of 20%, generally yields better results (Wilcox, 2012, p. 304).

By assuming that $X_1, \ldots, X_n$ are the n observations in the sample, a percentile bootstrapping would begin by randomly resampling n observation with replacement from the original sample, yielding $X_1^*, \ldots, X_n^*$. In order to estimate the trimmed mean of the original sample ($\bar{X}_t$), one can calculate the trimmed mean of the bootstrap sample ($\bar{X}_t$*). The process of obtaining bootstrap samples could continue for T times, leading to T estimates as $\bar{X}_{t1}^*, \ldots, \bar{X}_{tT}^*$. A final step toward finding $1 - \alpha$ confidence intervals includes ordering these estimates in an ascending format such as: $\bar{X}_{t(1)}^*$, , …, $\bar{X}_{t(T)}^*$, and finding the right quantiles wherein $L = \alpha T/2$ (rounded to the nearest integer) and $u = T\text{-}L$. Thus, the bootstrap confidence interval for the trimmed mean would be ($\bar{X}_{t(L+1)}^*, \bar{X}_{t(u)}^*$).

In the case of using the linear contrast codes, one can find the estimate of psihat ($\hat{\psi} = \sum_{j=1}^{J} c_j \bar{X}_{tj}$) in which $\bar{X}_{tj}$ is calculated from the true trimmed means of J independent groups and c_js are constants with the sum of zero—and subsequently compute the p-values associated with each test. An example could clarify this process: suppose there are 5 groups, and the hypothesis is to compare groups 2 and 3. One could code c_2 as +1, c_3 as -1 and the rest as 0 and test whether $H_0: \psi = \mu_{t2} - \mu_{t3} = 0$. To test this hypothesis, one could obtain a T bootstrap sample as explained previously for each group (second and third groups in this example), find the estimate of $\hat{\psi}$, order them, and find the right quantiles. The hypothesis could be rejected whenever the confidence interval of $\hat{\psi}$ does not include zero and when the p-values are less than the significance level (α). Percentile bootstrapping could be considered as an alternative method to perform multiple pairwise tests while using robust measures of location and scale.

Controlling Type-I error rates in multiple pairwise comparisons

Testing multiple hypotheses can result in a multiplicity problem that causes false positive-rates (i.e., Type-I errors) to increase excessively (Benjamini and Hochberg, 1995). Different methods have been suggested to address this issue. The simplest adjustment, known as the Bonferroni method, posits that to test m hypotheses in such a way as to enable the probability of one or more Type-I errors to remain at most at the significance level (α), one can execute each test at α/m significance level. However, some studies have raised concerns about the reliability of this approach. Perneger (1998, page 1), for instance, claims that the Bonferroni correction could "create more problems than it solves" because it focuses on a condition in which all null hypotheses are true, which is not valid in all cases. Perneger also suggests that using very conservative adjustments for Type-I errors—such as the Bonferroni method—can lead to losing more power, which is not desirable. Therefore, several improvements have been suggested in previous studies to overcome these challenges (Hochberg, 1988; Rom, 1990; Benjamini and Hochberg, 1995; Wilcox, 2001).

This study has adopted the false discovery rate (FDR) method introduced by Benjamini and Hochberg (1995) as FDR (and its extensions) has been widely used in several studies to avoid inflated Type-I error rates (Weisberg et al., 2003; Storey and Tibshirani, 2003; and Anders and Huber, 2010). Specifically, by assuming that R is the number of hypotheses that are rejected from the total of m hypotheses, and Q is the proportion of R that are true but rejected in error (i.e., Q is the proportion of Type-I error), FDR is defined as the expected value of Q. Benjamini and Hochberg (1995) illustrate that the conventional familywise error rate (FWER)—the probability of having at least one erroneous rejection—can be replaced with FDR, which is "less stringent" and can therefore produce more powerful results.

The Benjamini-Hochberg procedure starts with listing all p-values in an increasing order such that $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$. Users let k equal the largest value of i for which $P_{(i)} \leq (i \times \alpha/m)$. One can then reject all hypotheses for $i = 1, 2, \ldots, k$. For instance, suppose five hypothesis tests' p-values, at significance level of 0.05, are: 0.005, 0.011, 0.015, 0.05, and 0.080. Bonferroni correction would suggest rejecting only tests with p-values less than 0.05/5 = 0.010. Only one test (0.005) would be rejected in this scenario. With the Benjamini-Hochberg method, however, the largest i that can pass the test is 3, and therefore three of five hypotheses could be rejected at $\alpha = 0.05$. This example shows how Benjamini-Hochberg method could provide a less-conservative metric while controlling Type-I error rates in multiple tests.

2.3 Research Methods

Based on the past studies and the established statistical methodologies discussed in the background section, the research team implemented three main activities to determine the cost of injuries across different event types within a single construction trade. First, a comprehensive content analysis was conducted on accident reports to collect potential factors and to develop a data set for analyzing the costs of injuries. Second, the costs of injuries were estimated based on the nature of injuries (e.g., fracture, burn). Third, the research hypotheses were developed and tested using robust methods (i.e., adopting trimmed mean and Winsorized variance in Welch, Yuen, and bootstrapping methods) to compare the cost of injuries among various event types, project end-uses, and project budgets and to investigate the possible interaction effects among these accident factors. The last activity also included examining the cost distributions of independent categories to gain more insights into accident data and—more importantly—to find appropriate methods of analysis. The detailed description of each activity is provided here.

2.3.1 Content analysis

Content analysis has been successfully used to detect accident causes and to predict the severity of accidents (Esmaeili et al., 2015a, b, Gholizadeh and Esmaeili, 2016). The research team built upon the success of these previous studies to develop a reliable safety database for further statistical analysis. First, the research team collected 708 accidents reports from the electrical contractors' category (North American Industry Classification System (NAICS) number: 238210) in the OSHA IMIS database; these reports were dated between 2007 and 2013. After downloading the accident reports, the research team screened the data and removed repeated reports of a single accident, reported accidents with no injuries, reported injuries whose causes were unrelated to work (e.g., heart attack, embolism), reports without enough information (e.g., unspecified source and nature of injury, unknown event type), and reported accidents that happened in a trade other than electrical contractors. This process reduced the number of accidents reports to 633. However, after excluding the fatal accidents from the reports, a total of 388 cases were left for cost analysis.

The content analysis then coded and documented the following variables: event (i.e., accident) type, project end-use and project budget. When using inspection/accident summaries to investigate construction incidents, it is common to encounter errors or inconsistencies in reporting various factors. Previous studies have acknowledged this limitation in different contexts. For instance, Seixas et al. (1998) designed an experiment and asked three trained inspectors to measure the hazard levels of a construction site; they concluded that, even in very simple situations, there would be high “interobserver” variations in rating hazards. To avoid such errors and subjective judgments, researchers often code the narratives manually based on some frameworks such as Haddon’s matrix (Haddon 1968, 1972), which was originally developed to classify factors contributing to motor vehicle injuries (Bondy et al. 2005). In this study, the content analysis was conducted manually and over multiple rounds by applying a common standard (Occupational Injury and Illness Classification Manual or OIICM), which was developed by the U.S. Department of Labor’s Bureau of Labor Statistics (Bureau of Labor Statistics, 2012). An example of the subjective reporting errors was seen when the authors were reviewing struck-by accidents. In 16 cases, when working with electrical parts, the worker was “struck-by an arc flash”; all these cases were reported as struck-by accidents by OSHA inspectors. However, the ‘event or exposure’ section of the OIICM clearly has classified such accidents as ‘exposure to electricity’ under code 51. Such standardization across the subjective reports enabled the team to derive quantitative data from the qualitative reports.

Fourteen event types had been reported initially in accident reports, including two large categories of ‘other’ (7%) and ‘missing’ (16%) and seven categories (i.e., absorption, fall on same level, repetitive motion, rubbed/abraded, cardiovascular/respiratory failure, ingestion, and inhalation) that only accounted for 6% of the data overall. All instances with ‘other’ and ‘missing’ values were reviewed and whether reclassified to a new category (e.g., fall or struck by) using OIICM standard or removed from further analysis (in case of missing values). The seven small categories were combined to form a new ‘other’ category, which is now more meaningful as its components are known. ‘Struck by’ and ‘struck against’ accidents were combined into one category. These modifications formed the final five categories of event types in Figure 2.1. This new classification allows the research team to compare the cost of injuries among common event types that are representative of the accidents that often occur to electrical workers.

In original OSHA reports, project budgets and project end-uses were reported in seven and seventeen categories, respectively. Such numerous categories would result in groups of accidents with only a few observations, which would degrade the comparisons into categories that are not meaningful and/or representative—for instance, end-uses such as ‘excavation, landfill,’ ‘bridge,’ and ‘pipeline’ had only two, one, and one cases, respectively, in the original dataset. The effect of having factors with too many levels would be even more noticeable in multiple ANOVA, where the number of combinations between levels can grow quickly and result in many groups with no observations. To avoid this problem, the authors decided to consider only two categories for each of these two factors. For the end-use, a logical choice was, first, to divide the projects into building (i.e, 298 accidents or 67%) and non-building (i.e., 90 accidents or 23%). The relatively low number of non-building projects suggests that dividing this factor into more categories would make very small, non-representative groups, which is not ideal for analysis.

For the project budget, the original categories (and their frequencies) were as follows: under \$50k (181), \$50k to \$250k (67), \$250k to \$500k (37), \$500k to \$1m (28), \$1m to \$5m (38), \$5m to \$20m (13), and more than \$20m (24). These groupings clearly show that accidents are not distributed evenly among different project sizes and that most of the accidents have occurred among the smallest projects. To reduce the number of categories, the authors decided to consider \$50k as the cut-off point to re-classify project budgets. Combining all projects with budgets more than \$50k would create two relatively equal final categories: projects with a budget under \$50k—which represents 47% of the data—and projects with more than \$50k—which include 53% of accidents. More importantly, selecting this cut-off point allows the research team to compare the cost of injuries among the smallest projects in the data to those with much higher budgets. One may note that these budgets indicate only the amount that was given to the electrical contractors and are not representing the total budget of projects.

2.3.2 Estimating the cost of injuries

Unfortunately, OSHA's accident summaries do not directly report the cost of injury. Therefore, to determine the costs for each accident, the research team used OSHA's safety pay program in which the cost of injuries is determined based on the nature or type of injuries. These estimates are derived from a dataset of the NCCI and represent the average cost of lost-time workers' compensation claims for policy years 2011 through 2013. Using this database, the research team could calculate the costs of injuries for each accident. For instance, if the nature of an injury was a fracture, its corresponding direct cost from the NCCI report would be $50,778. The research team decided to use only the direct cost of injuries for two main reasons. First, there are several assumptions involved in calculating the indirect cost, and as described in the background section of this paper, there is no consensus among researchers regarding the exact ratio between direct and indirect costs. Second, according to OSHA's safety pays program, when the cost of an injury is higher than $10,000, the ratio of indirect to direct costs will be the same (i.e., 1.1) regardless of the type of injury. In this study, all the costs were more than $10,000, which is reasonable as only catastrophic accidents are involved in OSHA's dataset. Since multiplying all the direct costs to the same constant (i.e., the ratio of indirect to direct costs) would not change our analysis, the research team decided to use only the direct costs. The cost of injuries for each nature of injury are presented in Table 2.1.

Using the information in Table 2.1, now each accident in the data would have an estimated cost. Furthermore, the average cost of injuries among different levels of the three factors in the study can be calculated for future comparisons. For instance, from the 388 non-fatal accidents, 47 included a struck-by event. Adding the cost of all different injuries in these 47 cases would add up to around $1,953,000, which results to an average of $41,553 for each struck by accident.

Table 2.1. Aggregated cost of different natures of injuries (all costs are in dollars)

Nature of injury	Cost ($)
Laceration	19,713
Hernia	22,313
Heat Prostration	23,495
Puncture	25,523
Contusion	27,511
Strain/sprain	31,565*
Burn	40,188
Fracture	50,778
Concussion	59,372
Multiple physical injuries	73,749
Amputation	77,995
Electric shock	93,858

* This cost is the average of strains ($33,140) and sprains ($29,989)

2.3.3 Design of the analysis

As mentioned earlier, many studies have used statistics to estimate costs and have reported costs according to such factors as demographics (race, age, and occupation of the workers), professional trades, and type and source of injury. However, no study has compared these factors by means of inferential statistical tools such as hypothesis testing. Without such analysis, the differences between the costs cannot be fully investigated.

To determine the effects of different event types, project end-uses, and project budgets on the costs of injuries, this study implemented two types of statistical tests: i) main effect tests and ii) post hoc tests. The list of null hypotheses that were tested in this study are presented here:

Main effects: Cost of injuries are equal among different event types/project end-uses/budgets.

Interactions: The effect of event types/end-uses on costs does not depend on the project budget and vice versa.

Pairwise comparisons among main effects: (applied only on event type as this category is the only factor with more than two levels): Cost of injuries is equal among two different event types (e.g., fall vs. exposure to electricity).

Pairwise comparisons among interactions (should be tested only if the overall interaction test is significant): The effect of two event types/end-uses on costs does not depend on the project budget and vice versa.

These hypotheses provided the testable questions that drove our statistical analysis.

ANOVA design

A main effect test examines whether measures of location (e.g., statistical means) vary among different levels of the variable. If the factor had only two levels (e.g., end-use and budget), a significant main effect test is sufficient to reject the null hypothesis and conclude that one level, on average, has higher values of a dependent variable than the other level. However, when there are more than two categorical levels influencing the null hypothesis (e.g., as occurs under the event-type factor), rejecting the null only implies that two or more means are different at a significance level (α), and it is not possible to exactly identify which categorical levels are significantly different. To address this limitation, one can perform multiple post hoc pairwise comparisons to test which categorical level under a factor caused the observed effect (Schumacker, 2014).

To check the assumptions of ANOVA, the research team assessed the distribution of costs across the various factors and levels (e.g., event type). Then, the sample sizes and variances were checked to determine whether a conventional F-test would be appropriate to check the main effect and pairwise hypothesis tests. When any of the necessary assumptions were not met, the research team applied robust methods. Beyond the main effects of the factors 'event type,' 'project end-use,' and 'project budget' on the cost of injuries, the interactions among these variables have also been of interest. Particularly, the authors were interested in testing whether the effects of "event type" on cost depend on projects' end-uses (i.e., building, and non-building) and projects' budgets (i.e., less than \$50,000 or more than \$50,000). One-way ANOVA could not answer these types of comparisons about the interactions between the levels underneath the high-level factors. Therefore, the team needed to design a factorial ANOVA framework to test the main effects, interaction effects, and post-hoc pairwise comparisons of the factors presented in Figure 2.1.

Since there were three factors — 'project end-use,' 'project budget,' and 'event type'—the authors proposed a three-way factorial ANOVA design in which the levels under the factors determined the dimensions of the analytical matrices (5 event types x 2 end-uses x 2 budget levels). This framework is appropriate to test the interactions between factors and also can provide more accurate main effect tests by using unweighted means.

Hypotheses testing

After selecting the design framework, one needs to find an appropriate test statistic to conduct the hypotheses testing within the proposed design. When the variances were not constant among levels and the distributions of the dependent variable were not normal, the research team applied a robust measure of location (i.e., 20% trimmed means) and scale (i.e., Winsorized variance) along with a combination of robust methods, as described in the background section. The Welch-type procedure is used to test the main hypotheses. For pairwise tests, the Yuen's extension was used as it accounts for non-independent issue of trimmed samples. Bootstrap methods also can be used in this situation. Several R functions (e.g., *t3wayv2, mcp3atm, bbbmcppb*) from the *WRS* package can perform this three-way analysis and tested all pairwise comparisons and their interactions. To control for Type-I error rates, the Benjamini-Hochberg method was selected by putting the value of *bhop =TRUE* in the R function. The significance level of $\alpha = 0.05$ was used in this study and 2,000 samples were generated to build the confidence intervals for the bootstrapping process.

One common practice is to test multiple hypotheses using linear contrasts to compare combinations (Davis, 2010). However, as there was no 'control' group among the variables in this study (i.e., no levels were of more interest than others), the authors decided to test all hypotheses *pairwise* since any type of event could cause serious injuries to workers. For a variable with J levels, the number of pairwise comparisons can be calculated as: $J(J\text{-}1)/2$. Thus, based on the 'event types' factor, ten pairwise tests (= 5 × 4 / 2) needed be generated, as shown in Table 2.2, and the *con3way( )* function from the

WRS package assigned these comparisons contrast codes. (The abbreviations from Table 2.3 can help to represent each group for the explanation here.) For instance, in order to test whether the cost of injuries for exposure to electricity and fall accidents are significantly different (i.e., $\mu_E - \mu_F$), four exposure groups (EBL, ENL, EBM, and ENM) were coded as +1, four groups of fall accidents (FBL, FNL, FBM, and FNM) were coded as -1, and the other twelve groups were coded as 0. Since ‘project end-use’ and ‘project budget’ each only have two levels, one contrast code would be enough for their analysis. For instance, in the case of project end-uses, all ten “building” projects could be coded as +1, and all “non-building” projects as -1. The null hypotheses then would be: $H_0: \mu_B - \mu_N = 0$.

Table 2.2. Contrast codes for pairwise comparisons among levels of event type

	Less than $50,000 (L)										More than $50,000 (M)									
	Building (B)					Non-building (N)					Building (B)					Non-building (N)				
Hypothesis	Caught in/b. (C)	Electricity (E)	Fall (F)	Other (O)	Struck-by (S)	Caught in/b. (C)	Electricity (E)	Fall (F)	Other (O)	Struck-by (S)	Caught in/b. (C)	Electricity (E)	Fall (F)	Other (O)	Struck-by (S)	Caught in/b. (C)	Electricity (E)	Fall (F)	Other (O)	Struck-by (S)
$\mu_C - \mu_E$	+1	-1	0	0	0	+1	-1	0	0	0	+1	-1	0	0	0	+1	-1	0	0	0
$\mu_C - \mu_F$	+1	0	-1	0	0	+1	0	-1	0	0	+1	0	-1	0	0	+1	0	-1	0	0
$\mu_C - \mu_O$	+1	0	0	-1	0	+1	0	0	-1	0	+1	0	0	-1	0	+1	0	0	-1	0
$\mu_C - \mu_S$	+1	0	0	0	-1	+1	0	0	0	-1	+1	0	0	0	-1	+1	0	0	0	-1
$\mu_E - \mu_F$	0	+1	-1	0	0	0	+1	-1	0	0	0	+1	-1	0	0	0	+1	-1	0	0
$\mu_E - \mu_O$	0	+1	0	-1	0	0	+1	0	-1	0	0	+1	0	-1	0	0	+1	0	-1	0
$\mu_E - \mu_S$	0	+1	0	0	-1	0	+1	0	0	-1	0	+1	0	0	-1	0	+1	0	0	-1
$\mu_F - \mu_O$	0	0	+1	-1	0	0	0	+1	-1	0	0	0	+1	-1	0	0	0	+1	-1	0
$\mu_F - \mu_S$	0	0	+1	0	-1	0	0	+1	0	-1	0	0	+1	0	-1	0	0	+1	0	-1
$\mu_O - \mu_S$	0	0	0	+1	-1	0	0	0	+1	-1	0	0	0	+1	-1	0	0	0	+1	-1

Since the research team also wanted to study interactions between factors, relevant contrast codes were developed for each comparison using the *con3way* function. For instance, to test whether 'project end-use' had any effect on the difference between the cost of injuries caused by 'exposure to electricity' and 'fall' accidents, the following null hypothesis was tested: $H_0: \mu_{EB} - \mu_{FB} - \mu_{EN} + \mu_{FN} = 0$.

2.4 Results

To determine the need to disaggregate the injury costs down to the levels under each factor—and thereby to determine whether the research design demanded one-way or factorial ANOVA—the research team examined the effect of event types. Table 2.3 shows the summary of costs of injuries among the five event types. Before testing the main-effect hypothesis (i.e., comparing the mean of the cost for each level), the research team observed that both the standard deviation (and hence the variance) and the number of observations differed among the groups. Such an observation signified, for instance, that 'exposure to electricity' leads to injuries with a more diverse range of costs compared to 'fall' accidents. While this observation may be offset by having more observations in the sample, the fact that these standard deviations diverged indicated that a simple one-way ANOVA design was not appropriate for this study.

Table 2.3. Cost of injuries among five event types (in thousand dollars)

	Fall to lower level (F)	Struck-by (S)	Electricity (E)	Caught in/between (C)	Other (O)
Frequency	146	47	141	24	30
Mean	$47.89	$41.55	$61.49	$68.67	$45.63
Standard Deviation	$10.64	$17.29	$26.32	$12.81	$13.28

The results of a Levene's test (test statistic = 132.98, p-value < 2.2e-16) further illustrated that the variances were not equal, and checking the estimated distribution of the residuals showed that the normality assumption was also violated. Figure 2.2 visualizes these violations and their effects on the confidence intervals of the estimated means. Instead of just showing the pointwise estimates (i.e., means), Figure 2.2 (left) displays the 95% confidence intervals for each mean. If the sample size and variance were equal, the length of these intervals would be the same, thereby providing a fair context for using a conventional F-test to test the hypothesis: categories with no overlap in confidence intervals would be considered significantly different. Furthermore, the estimated distribution of residuals in Figure 2.2 (right) indicates that the normality cannot be assumed for these five categories. Therefore, the one-way ANOVA assumptions could not be met, which would conceivably inflate the Type-I error rates of the F-test beyond the 0.05 level and thereby weaken the findings, as discussed in the background section.

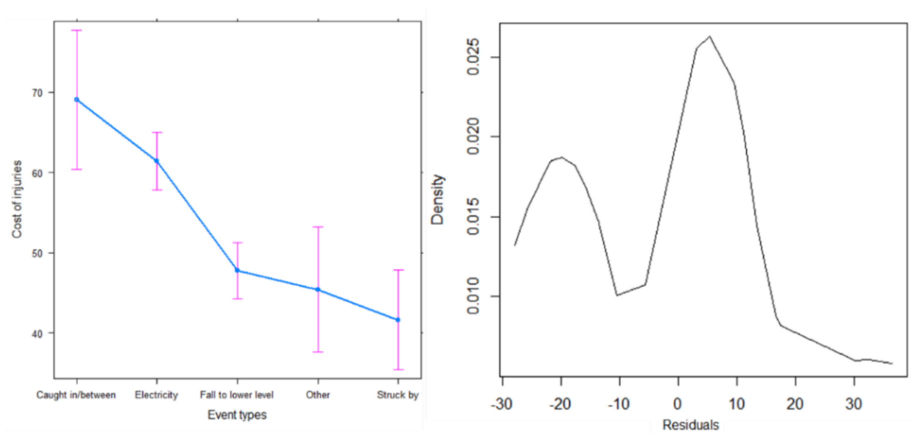


Figure 2.2. (a) Mean estimates and confidence intervals of injury costs (in thousand dollars) among five event types; and (b) estimated distribution of the residuals

To provide more clarity, the research team applied both the conventional F-test and the robust methods with trimmed means to show how selecting a wrong method can produce misleading results. Both omnibus tests indicated that the cost of injuries differed among event types and therefore that the null hypothesis can be rejected (F-test statistic was 19.53, with a p-value of 1.23e-14; Welch-type test statistic with 20% trimmed means was 10.11, with a p-value of 5.67e-06). However, the pairwise comparisons revealed the differences between the two methods. Concretely, while the F-test found that the cost of 'exposure to electricity' is significantly higher than 'falls' (confidence interval for the difference: [7.51, 19.77]), the confidence interval of the Yuen's method (i.e., [-0.45, 18.04]) indicated that this difference is not significant. In another test based on a bootstrap version of the Yuen's method, there is enough evidence to believe that 'fall' and 'struck-by' accidents would result in significantly different costs (confidence interval for the difference: [0.18, 18.10]) even when the F-test cannot reject the hypothesis that these two categories are equal (confidence interval for the difference: [-2.52, 14.89]).

Based on the discussions in the background section, given these data, the research team found the results of the more robust tests more reliable. Therefore, for the rest of this study, the findings were made using the robust methods. As a reminder, a Welch-type method with trimmed means was used to test the main effects, and two methods (i.e., an extension to Yuen's method and percentile bootstrapping) were selected for the pairwise comparisons because of the non-normal population, heteroscedastic variances, and un-equal sample sizes. Using two different techniques to test pairwise comparisons enabled the research team to confirm the reliability of the findings.

2.4.1 Three-way design: main effects and interactions

As mentioned before, to study the potential main effects of event types, end-use, and project budget as well as their interactions, a three-way design with 20 distinct groups was proposed. Table 2.4 lists the average cost of injury for all 20 possible event–end-use–budget combinations as well as the number of accidents in each sample and their standard deviations—all costs are in $1,000 to simplify the representation of data. For instance, in building projects with budgets less than $50,000, 62 accidents happened due to exposure to electricity, which cost, on average, $54,780. One may note that the "other" factor consists of seven event types, including "falls on the same level"; "exposure to temperature extremes"; "struck-against object or equipment"; "explosions"; "fires"; "crushed in collapsing structure, equipment, and material"; and "exposure to harmful substances."

Table 2.4. Descriptive statistics of twenty distinct groups

			Fall (F)	Struck-by (S)	Electricity (E)	Caught in/between (C)	Other (O)
Building Projects (B)	Less than $50,000 (L)	Mean	45.81	42.69	54.78	75.02	45.03
		Size	50	12	62	8	9
		SD	11.82	18.10	24.43	10.16	6.02
	More than $50,000 (M)	Mean	47.98	38.33	61.82	60.45	52.04
		Size	70	23	52	5	7
		SD	11.00	17.35	26.30	13.93	21.18
Non-building Projects (N)	Less than $50,000 (L)	Mean	50.06	41.98	72.04	64.52	50.41
		Size	11	6	15	4	4
		SD	8.21	16.36	26.97	15.75	7.26
	More than $50,000 (M)	Mean	51.57	51.42	80.57	70.90	39.04
		Size	15	6	12	7	10
		SD	2.56	17.10	24.89	13.39	12.54

To better illustrate how end-use and project budget can affect the injury costs among different event types, Figure 2.3 shows the estimated means (and their confidence intervals) for the five event types under the four different project contexts. In general, ‘caught in/between’ and ‘exposure to electricity’ led to higher injury costs than the other three accident types under all four contexts. In these two event types, building projects with low budgets were the only context in which ‘caught in/between’ accidents resulted in higher costs than ‘exposure to electricity’. On the other hand, among the other three categories, ‘struck-by’ only exceeded ‘other’ accidents in non-building projects with higher budgets. The significance of these interaction effects will be tested in the following sections.

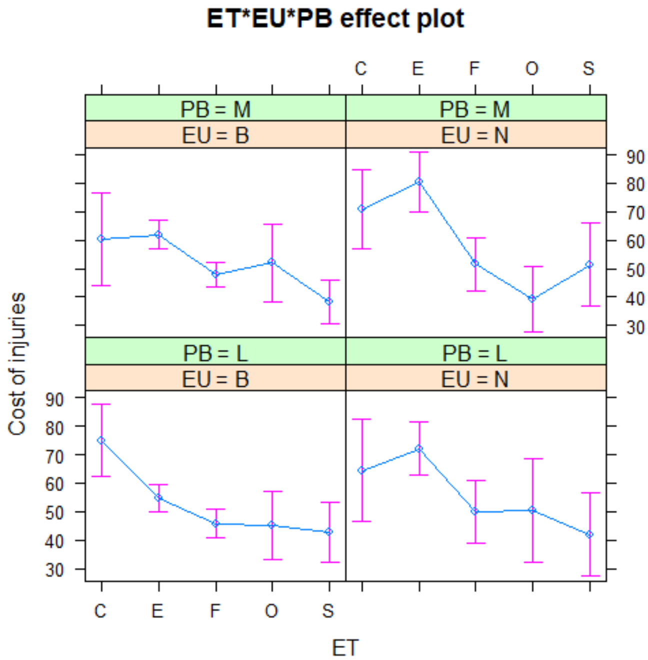


Figure 2.3. Cost of injuries among event types partitioned by end-use and project budget
PB = project budget; M = more than $50,000; L = less than $50,000; EU = end-use; B = building; N = nonbuilding; ET = event type; C = caught in/between; E = electricity; F = fall to lower levels; O = other; and S = struck-by.

To employ robust methods, one can apply a 20% trimming on each group. Table 2.5 shows the sample statistics, after 20% trimming, of a 5 (event types) × 2 (end-use) × 2 (project budget) factorial ANOVA design on cost of injuries among electrical contractors.

Table 2.5. Descriptive statistics of twenty distinct groups after trimming

			Fall to lower level (F)	Struck-by (S)	Electricity (E)	Caught in/between (C)	Other (O)
Building Projects (B)	Less than $50,000 (L)	Mean	49.64	45.05	47.35	78.35	41.69
		Size	30	7	38	6	8
		SD*	8.26	5.87	23.87	1.65	12.08
	More than $50,000 (M)	Mean	50.79	51.59	58.61	58.67	37.66
		Size	42	5	32	3	15
		SD*	1.14	20.70	25.99	13.63	14.63
Non-building Projects (N)	Less than $50,000 (L)	Mean	50.77	50.41	75.49	64.52	44.54
		Size	7	4	9	4	4
		SD*	1.51	7.26	26.49	15.75	16.29
	More than $50,000 (M)	Mean	51.03	40.59	87.20	73.21	51.49
		Size	9	6	8	5	4
		SD*	0.62	12.11	24.53	13.05	0.58

* All the SDs are Winsorized standard deviations

2.4.2 Testing the hypotheses

First, a Welch-type procedure was applied to test the main effects and interactions, and the results appear in Table 2.6. Among the three factors, only the main effects of “event type” and “end-use” were found to be significant. Consequently, one can conclude (1) that different event types can cause different injury costs and (2) that the cost of injuries is not equal between buildings and non-building projects. However, this study did not find a significant difference among the cost of injuries that occur in projects with budgets ‘less than $50,000’ versus those with budgets ‘more than $50,000.’ In terms of interactions between the factors, no significant effect was detected. However, the interaction between event types and end-use had a relatively small p-value and therefore was investigated in more detail during the pairwise comparisons.

Table 2.6. The main effects and interactions among the three factors

		Test Statistic	Critical Value	p-value	Effect size
Main effects	Event Type (ET)	**41.43**	**12.76**	**2e-04**	0.56
	End-Use (EU)	**5.36**	**4.10**	**0.026**	0.26
	Project Budget (PB)	0.19	4.10	0.67	0.09
Interactions	ET:EU	11.24	12.76	0.07	0.63
	ET:PB	2.26	12.76	0.74	0.56
	EU:PB	0.56	4.10	0.46	0.22
	ET:EU:PB	5.67	12.76	0.32	0.64

While end-use has only two levels and, based on the previous results, one can conclude that non-building projects lead to higher injury costs than building projects, the specific level differences among event types were not apparent within the previous findings. Therefore, two robust methods based on the 20% trimmed means were used to investigate all pairwise comparisons among the five event types, and the findings appear in Table 2.7. One may note that all the values for *psihat* are positive, meaning that 'category1' has higher values than 'category2' in all tests. While the bootstrapping method generally generated smaller confidence intervals, both methods found the exact same tests to be significant:

- There was a significant difference in the cost of injuries between 'caught in/between' accidents and 'fall' accidents; the cost of injuries from 'caught in/between' accidents were significantly higher than 'fall' accidents.
- The same results occurred when comparing injury costs between 'caught in/between' and 'other' ($\hat{\psi}$=87.10), and between 'caught in/between' and 'struck-by' accidents ($\hat{\psi}$=99.38), demonstrating the higher costs of the 'caught in/between' event-type level than both 'other' and 'struck-by' events.

- Similar conclusions can be drawn for accidents caused by ‘exposure to electricity.’ Post-hoc tests indicate that ‘exposure to electricity’ can cause higher injury costs than ‘falls,’ ‘other,’ and ‘struck-by’ levels.

Table 2.7. The pairwise comparison of the main effects of event types

				Extension of Yuen’s method to trimmed means		Percentile bootstrap	
Category1	Categoriy2	Effect size	psihat	CI lower	CI upper	CI lower	CI upper
Caught in/between	Fall	0.73	72.52	**10.52**	**134.53**	**38.65**	**104.06**
Caught in/between	Other	0.76	87.10	**18.67**	**155.53**	**38.74**	**130.71**
Caught in/between	Struck-by	0.76	99.38	**31.98**	**166.77**	**51.13**	**144.71**
Caught in/between	Electricity	0.24	6.10	-66.37	78.58	-47.79	61.17
Electrical	Fall	0.25	66.42	**11.64**	**121.20**	**14.15**	**101.34**
Electrical	Other	0.34	81.00	**13.04**	**148.95**	**27.22**	**129.50**
Electrical	Struck-by	0.37	93.27	**27.00**	**159.55**	**31.03**	**146.74**
Fall	Other	0.39	14.58	-36.04	65.19	-16.68	43.95
Fall	Struck-by	0.36	26.85	-23.27	77.00	-5.22	64.16
Other	Struck-by	0.17	12.28	-48.90	73.46	-33.54	60.57

As mentioned before, only one interaction between event types and end-use was marginally significant. Hence, a set of ten pairwise comparisons were tested regarding this interaction. The results in Table 2.8 indicate two significant interactions when using the percentile bootstrap method. Based on these findings, one can conclude that nonfatal ‘exposure to electricity’ causes higher injury costs than nonfatal ‘falls,’ and this effect is even stronger in ‘non-building’ projects than ‘building’ projects. In the same manner, the effect of end-use on injury costs is even more significant when the accident has been caused by ‘exposure to electricity’ than the ‘other’ level. One may note that the second interaction was not found to be significant in Yuen's method and should be interpreted with more caution.

Table 2.8. The pairwise interactions between event type and project end-use

		Extension of Yuen's method to trimmed means		Percentile bootstrap	
Interactions	psihat	CI lower	CI upper	CI lower	CI upper
CB-CN-EB+EN	56.01	-16.46	128.49	-5.02	106.29
CB-CN-FB+FN	0.65	-61.35	62.66	-31.15	35.26
CB-CN-OB+ON	-6.35	-74.79	62.08	-51.47	39.13
CB-CN-SB+SN	15.97	-51.43	83.36	-33.37	62.84
EB-EN-FB+FN	-55.36	**-110.13**	**-0.58**	**-91.20**	**-7.23**
EB-EN-OB+ON	-62.37	-130.32	5.59	**-113.12**	**-8.06**
EB-EN-SB+SN	-40.05	-106.32	26.23	-92.45	15.16
FB-FN-OB+ON	-7.01	-57.62	43.61	-39.31	21.28
FB-FN-SB+SN	15.31	-34.84	65.46	-24.62	48.97
OB-ON-SB+SN	22.32	-38.86	83.50	-31.42	68.14

2.5 Discussion

2.5.1 Injury costs across accident type, project end-use, and project budget

The effects of distinct event types, project end-uses, and project budgets on the costs of occupational injuries have been analyzed to determine which accident types/end-uses/budget levels are costlier. One should note that, as mentioned in the introduction section, the analysis has only included the nonfatal injuries and therefore the interpretations should be limited to nonfatal cases. The main effect tests determined that the cost of injuries can vary significantly among different event types and project end-uses.

Regarding event types, while 'exposure to electricity' and 'fall' were the most common accident types among electrical contractors—and hence accumulated the most cost—one must focus on the average cost of each type of event to compare events. The exploratory results suggest that ‘caught in/between’ and ‘exposure to electricity,’ on average, can cause higher costs than ‘fall,’ ‘struck-by,’ or ‘other’ accidents. The robust hypothesis tests further confirmed the significance of these differences. After the omnibus test revealed that the costs of different event types are not equal, pairwise tests were conducted to indicate where the difference comes from. The pairwise comparisons revealed that ‘caught-in/between’ accidents, on average, cost significantly more than accidents caused by either a ‘fall to a lower level,’ ‘struck-by objects,’ or ‘other.’ These findings are consistent with the results of Waehrer et al. (2007b), who reported that caught-in accidents ranked higher (rank 6) than those caused by falls (rank 9) or struck-by objects (not among the first 10) when ranking events based on the average cost of days away from work. Such an outcome could be explained by the nature of injuries caused by different events. One reason for this could be that ‘caught-in/between’ accidents, would result in more severe and often permanent injuries (e.g., amputation, concussion) than ‘fall’ and ‘struck-by.’ For instance, 56% of ‘caught-in/between’ accidents have resulted in an amputation/concussion; this number is only 11% for ‘fall’ and ‘struck-by’ accidents. In one case, for instance, during a lifting operation, a worker’s hand was caught between a transformer and the sling, causing the index finger of the worker to be amputated. In comparison, fractures were the most common outcome of a nonfatal fall (74%) or struck-by (50%) accidents. MacKenzie et al. (2007) reported that the total direct cost (including hospitalizations, doctor visits, etc.) of lower extremity amputations could reach $91,000 (in 2002 dollars), which is substantially higher than the $39,000 cost for multiple fractures reported by Bonafede et al. (2013). ‘Exposure to electricity’ yielded similar results as ‘caught in/between’ when compared to ‘struck-by,’ ‘fall,’ and ‘other’ accidents. Almost all nonfatal ‘exposure to electricity’ accidents (98%) resulted in either an electrical burn or an electrocution. Electrocutions resulted in the highest injury costs reported by NCCI, which can explain the fact that exposure accidents had higher costs than ‘struck-by,’ ‘fall,’ and ‘other’ accidents. Waehrer et al. (2007b) ranked “contact with electricity” as the second costliest (on average) event. Fordyce et al. (2007) also acknowledged the high average cost of electrical injuries, especially when the whole-body system is injured ($228,830). To summarize the initial findings, one can conclude that for electrical contractors, ‘exposure to electricity’ and ‘caught-in/between’ events have higher injury costs than ‘fall to a lower level,’ ‘struck-by,’ and ‘other’ accidents.

For the test on end-uses, this study found that, while around 77% of the nonfatal accidents among electrical workers occur on building worksites, these accidents, on average, cost significantly less than injuries that occur on non-building projects. This result can be attributed to the fact that while electric shocks, amputations, and concussions (i.e., the costliest types of injury) account for only 21% of injuries in building projects, their share in non-building projects is 33%. Higher injury costs in non-building projects indicate that these projects expose workers to higher levels of hazard (e.g., working in open spaces with more machinery in workers’ vicinity), which could result in more severe accidents. Moreover, unlike building projects—which are all, to some extent, similar in nature—non-building projects include a diverse set of end-uses ranging from powerlines and refineries to streets and bridges. This variety in project characteristics can lead to more uncertainties and make it harder for small contractors to recognize all the hidden safety-risk factors on a construction site. Previous studies have investigated potential factors that can influence the rate of accidents among construction companies. For instance, Hinze and Gambatese (2003) pointed out that factors such as implementing employee drug testing, safety training, and minimizing worker turnover can positively affect the safety performance of specialty contractors. However, the effect of project end-use on cost of injuries has not been studied enough among safety researchers. Therefore, the authors believe that the findings of this study could encourage researchers and practitioners to consider end-use in future studies as an important factor in determining the hazard levels and risks that are involved in specific construction projects/tasks. Consequently, the outcomes of our study's analysis resonate with both the literature and anecdotal analysis.

The significant effect of project end-use on the cost of nonfatal injuries was found to be stronger among some event types. The findings show that when compared to 'fall' and 'other' accidents, 'exposure to electricity' caused much more severe injuries in non-building projects than in building projects. In other words, non-building projects are far more dangerous if a worker's task involves the possibility of an 'exposure to electricity' rather than a 'fall' or 'other' accident. Further investigation reveals that two-thirds of 'exposure to electricity' cases that occurred in non-building projects resulted in an electrocution. This quantity is almost two times the rate of electrocutions in building projects when electricity is involved: most of exposure to electricity in building projects resulted in much less severe burn injuries. The research team have reviewed the electrocution cases and found two main reasons behind their prevalence in non-building projects. First, in non-building contexts (e.g., powerlines, transmission lines, excavation, highways), electrical parts/components are usually energized to much higher voltages. The reports show several cases in which the employee, or part of the vehicle he/she was working on, was contacted with massive voltages of 1200, 7200, 13800, 69000, and in one case, 115000 volts. Such large voltages, which are not common in building environments, can explain the more severe and costlier injuries in non-building projects. Another factor that played a significant role in nonfatal electrocutions in non-building projects was the deenergizing process. For instance, in a roadway project, the electrical contractor was responsible for moving the base of a traffic light pole while another crew was responsible for deenergizing the wirings. A lack of communication between these crews resulted in an electric shock while an employee was using a jackhammer to break the concrete of the base. The authors would like to remind readers that these findings are based on only nonfatal incidents. The conclusions might be quite different in the context of fatal accidents.

2.5.2 Practical implications

As mentioned by Hallowell and Gambatese (2009), risk mitigation may be achieved in three ways: reducing exposure to hazards, reducing the frequency of accidents, and reducing the severity of accidents. Because estimating the amount of exposure can be very project-oriented, most studies examining safety programs have focused on estimating the frequency and severity of construction accidents. While the frequency of various accident types (e.g., falls, electrocutions) can be derived rather readily from historical data, estimating severity is not as straightforward, particularly among non-fatal injuries. One way to infer the severities of non-fatal injuries is by considering the cost they impose on workers, companies, and society (Tang et al. 2004). Therefore, this study has estimated the cost of non-fatal injuries for five categories of accident types, two categories of project end-uses, and two categories of project costs.

The average costs of injuries for each category, however, cannot be used to differentiate among categories until a formal statistical test confirms the difference between categories. For this reason, robust hypothesis tests have been conducted on cost estimates to reach statistically significant conclusions. The results indicate which accident type/project end-use have caused more costs among electrical contractors. Considering costlier (i.e., more severe) accidents along with more frequent event types can draw a more realistic picture of safety priorities among electrical contractors. For instance, looking only at the frequency of an accident type would suggest that 'caught in/between' events do not expose electrical workers to high risks and therefore can be neglected or at least not prioritized when assigning safety resources and/or designing safety interventions. However, considering the findings of this study, managers may want to emphasize 'caught in/between' cases, as such events demonstrate higher risks to both employees and companies, given the severity of their consequences. The findings also suggest that safety managers need to consider non-building projects as high-impact risk environments, particularly when the workers are exposed to sources of electricity. Safety practitioners could consider specialized training sessions for workers in non-building projects and educate them about unique sources of injuries in such environments.

Another potential beneficiary of the results of this study are insurance companies. The findings can help insurers in adjusting their premium rates not only based on a contractor's safety performance history but also based on the type of projects and activities that they have been involved in. For instance, knowing that an electrical contractor is mostly involved in building projects could result in some sort of reduction in their premiums.

2.5.3 Methodological advancement for construction-safety risk analysis

Beyond the value of our findings for practitioners trying to make cost-benefit decisions about different types of accident-prevention measures, a main contribution of this study is the demonstration of a more robust hypothesis-testing methodology for analysts within the construction-safety arena. The robust and nuanced approach here will help users disaggregate data in such a way to better derive meaningful insights in their risk analysis.

The authors wish to underscore that while many methods have been introduced, studied, and reviewed in statistics literature on the topic of the methods used in this paper, the fact that no single best method is agreed upon by the majority of statistical scholars shows that methods should be chosen carefully according to the sample data at hand. Therefore, this study introduced an approach in which one needs to investigate the samples before using a conventional method, such as an F-test, to examine the hypothesis. Visualizing the distribution of samples and the residuals alongside some statistical tests (e.g., Levene's test) could help researchers to understand the data and decide upon which common hypothesis-testing methods are appropriate.

The study also applied more robust measures of location and scale along with applicable test statistics that can be used especially when the underlying assumptions of other test statistics are not satisfied. As shown here, these tests can be done through available statistical packages and produce more reliable results when comparing the costs of two or more categories. Whenever a single method cannot be proved to be the best, the approach described in this study would serve as a good practice to yield several robustness checks and to enable researchers to compare their results and then interpret them. Such a best practice can determine the stronger main effects, as more tests would confirm any effects first found within the data.

When applying any statistical tests, the authors encourage researchers to examine confidence intervals very closely: When confidence intervals are very close to zero, one needs to interpret these tests more carefully as another test with higher significance level may fail to reject the hypothesis. On the other hand, a test may fail to reject the hypothesis when its confidence interval barely contains zero. One needs to be aware that the two categories might be still considered different at a slightly lower significance level, and therefore must consider confidence intervals alongside significance levels to derive meaningful conclusions.

2.6 Conclusions

One common way to measure the burden of an accident on a construction project is to estimate the consequences in monetary terms. Advancing our knowledge about the cost of injuries provides several benefits. In particular, the cost of injuries can be used as an indicator of potential risks associated with different activities and trades. This knowledge can help decision makers more effectively assign limited available safety resources (Kane, 1996) and focus on costlier accident types to prevent more severe injury types.

A large number of studies have investigated the cost of injuries; however, few studies have examined the cost of injuries among specialty trades such as electrical contractors. Consequently, this study analyzed occupational accidents among electrical contractors between 2007 and 2013. The variables that were considered in the cost analysis were accident types—and the corresponding nature of injuries—project end-uses, and project budgets. After scrutinizing the distribution of samples and reviewing the assumptions of conventional methods of analysis of variance, the authors applied robust methods of hypothesis testing to account for assumption violations. Different robust methods were tested on the data to gain the most reliable results. Ultimately, a Welch-type procedure (for main effects and interactions), an extension of Yuen's method, and a percentile bootstrap method with trimmed means and Winsorized variances (for pairwise comparisons) were selected to analyze the cost data. The results of pair-wise comparison indicated which event types and project end-uses could result in injuries with higher costs.

Some limitations related to this study are worth mentioning. First, the data analyzed in this study is biased toward more severe—though nonfatal—accidents, which may be perceived as a limitation because some small accidents with lower costs may be excluded from OSHA's database. Unfortunately, minor injuries are not usually reported accurately and therefore are not included in most public databases. However, the authors believe that the findings of this paper can still offer a reliable comparison among some of the more common accident types and their impact on the safety performance of electrical contractors. Second, the data did not provide the unique costs of each accident, and therefore, the average costs were assigned based on injury types (i.e., nature of injury such as fracture, burn, etc.). The authors recognize the sampling error that using the average cost data might bring to the analysis. However, considering the large quantity and the high quality of NCCI's database for injury costs, the sampling error is deemed to be marginal. Third, the authors acknowledge that incorporating other variables such as 'injured part of body' and the 'severity of injury' (e.g., the degree of a burn) would lead to more precise cost estimations and possibly stronger statistical results; we encourage future studies to incorporate these factors. Lastly, considering more predictor factors—such as safety budgets, experience and education levels of the workers, protection measures, and environmental and human factors—could have provided insightful information about the effects of such factors on the cost of injuries. Unfortunately, these factors were not available on OSHA's database and could not be examined in this study.

In future studies, researchers may benefit from considering other variables, such as the injured part of body, to make better cost estimations. Future studies also need to examine in more detail the effects of project context (e.g., project end-use) on severity and cost of occupational injuries, a topic which has not received much attention in construction. The authors recommend such pursuits to better aid decision makers and to better protect workers in the field.

CHAPTER THREE: TRENDS IN CATASTROPHIC OCCUPATIONAL INCIDENTS AMONG ELECTRICAL CONTRACTORS

3.1 Introduction

Preventing occupational incidents in the construction industry is challenging, since this industry involves a large number of relatively small employers, multi-employer worksites, numerous hazards, and a highly mobile workforce (Abudayyeh et al. 2003; Esmaeili and Hallowell 2012; Sousa et al. 2014). The risk of injury for construction workers especially escalates when these workers frequently interact with electricity, as is the case for electrical contractors: An examination of the National Traumatic Occupational Fatalities (NTOF) database revealed that, although construction laborers have the highest *number* of fatalities among the 38 different occupations included in the database, electrical power installers and repairers have the highest *rate* of fatalities among all occupations (Chen and Fosbroke 1998).

Many of these risks manifest in the nature of the work—electrical contractors' job entails installing and maintaining electrical systems as well as using a variety of hand tools (e.g., screwdrivers, pliers, knives, and hacksaws), power tools, and testing equipment (Abudayyeh et al. 2003). Complicating the occupational safety of this entire population is the fact that, because of the degree of skill required, electricians often spend most of their career in one of two categories (Robinson et al. 1999; Abudayyeh et al. 2003): indoor contractors, who install conduits, connect wires, test circuits, and install and maintain lighting systems; and outdoor contractors, who work with high voltage wiring in settings ranging from the community to the consumer. Problematically, unlike other specialty trades, who have experienced a decreasing number of injuries and fatalities in recent years, the number of fatal injuries across electrical contractors has increased rapidly between 2011 to 2016 (Bureau of Labor Statistics 2019).

To better understand the nature of these occupational incidents among electrical contractors—and therefore to discover effective safety interventions to prevent injuries (Xu and Xu 2021), especially fatalities—this study analyzed a large national database of occupational accidents to detect which statistically significant influential factors contribute to injuries and deaths among electrical contractors. This study has collected data from the Occupational Safety and Health Administration's (OSHA's) Integrated Management Information System (IMIS) accidents database from 2007 to 2013. Hence, the outcomes of this research work apply to the construction industry of the United States of America. Due to the categorical nature of accident factors (e.g., sources of injuries, event type), determining the correct values for these factors remains one of the main challenges of using historical accident data. Therefore, this study adopted a rigorous content-analysis method to ensure the reliability of final variables. The effects of influential accident factors on the fatality rates were then analyzed using the chi-square test of independence, Cramer's V tests, and Classification and Regression Trees (CART) analysis using decision trees (Elassad et al. 2020). Given the statistical significance of the variables identified during this study, these results will help practitioners to better understand nature of accidents, design specialized training programs, consider safety during design, choose alternative means and methods of construction, identify high-risk periods of a project, help small contractors better allocate their resources, and more strategically select injury-prevention practices.

The research team conducted an in-depth literature review related to occupational incidents among electrical contractors. Since electrical contractors are extensively exposed to electrical hazards, the research team also reviewed existing studies to better understand how workers get involved in electrical accidents. Salient results of the literature review are provided here.

3.2 Background

3.2.1 Occupational Incidents Among Electrical Contractors

There are limited number of studies that have investigated the nature of accidents among electrical contractors. In a 2003 study, Abudayyeh et al. (2003) developed a survey based on the Bureau of Labor Statistics (BLS) safety and health statistics database to identify tasks associated with injury, illness, and fatality trends in electrical contracting. Their study revealed information on important factors contributing to (mostly non-fatal) injuries—such as sources of injuries, event types, and nature of injuries—that occurred to electrical workers between 1992 to 1998. While the results were of interest to the current study, their study faced several limitations: (1) results were based on perceptions of contractors who responded to survey and not their actual incident history; (2) geographical distribution of respondents was limited to Michigan; and (3) sample size was very small—only ten contractors responded to the survey. However, comparing the magnitude of attributing factors in their study to more recent accident data could provide perceptive discussion on accident mechanisms among electrical workers.

To investigate nature and impact of burn-related injuries on electrical utility workers, Fordyce et al. (2007) reviewed 872 reports from the Electric Power Research Institute database between 1995 and 2004. The results indicated that while the numbers of burn-related accidents (including thermal, electrical, and chemical burns) were not high, such injuries resulted in a higher number of work days lost and more serious injuries compared to other injury types. Although burn-related injuries accounted for just 3.7% of all injuries, they seemed to be costlier, representing about 13% of all medical costs.

Some studies supported by the Electric Power Research Institute (EPRI) focused on the safety and health of electric utility and power industry workers. For example, Fordyce et al. (2010) analyzed neck injuries among electric utility workers from 1995 to 2007 and found higher rates of neck injuries in young males who had trade/craft worker experience. In another study, Fordyce et al. (2016) investigated fatal and non-fatal injuries and injury-severity factors among electric power industry workers between 1995 and 2013 and found that fatal injuries were most commonly associated with vehicle collisions and contact with electric current. They also found risk of fatalities to be higher among line workers and that line workers experienced the second highest risk for non-fatal severe injuries, after meter readers. More recently, Volberg et al. (2017) analyzed the EPRI occupational health and safety database to study injuries among electric power industry workers from 1995 to 2013. They found that while injury rates among electric power industry workers tended to decrease over the study period, rate of injuries remained high among certain high-risk workers: line workers, mechanics, young males, older welders and machinists, and female meter readers. Though each of these studies contributed to the overall body of knowledge, one limitation these studies shared was that their database only included EPRI member companies (i.e., 18 large electric power companies) and only covered a few occupations, both of which indicate that there may be selection bias in results.

Outside of the United States, Marhavilas et al. (2011) proposed a new hybrid risk assessment process for analyzing Greek public electrical power providers and identified high risk activities in this sector. Building on this initial success, Marhavilas and Koulouriotis (2012) combined stochastic and quantitative risk assessment methods to build a more realistic forecasting framework for the electric power provider industry. More recently, Castillo-Rosa et al. (2017) evaluated the impact of personal factors and consequences of electrical occupational accidents in the primary, secondary and tertiary sectors in Spain. They found that electrical accidents in all three sectors caused more severe consequences. Regarding personal factors, they found that workers' sex, age, experience, nationality and occupation significantly impacted type of accident. While contributions of these studies are significant, the data sources used in these studies are from outside of the United States, and there is a need to investigate accident reports among American electrical contractors.

3.2.2 Electrical Incidents on the Whole

Contact with electric current is a major cause of injury and death among construction workers (Ore and Casini 1996; Kisner and Casini 1998; Loomis et al. 1999; Taylor et al. 2002; McCann et al. 2003; Janicak 2008; Homce et al. 2008). Between

1994-2000, Census of Fatal Occupational Injuries (CFOI) data indicated that “contact with electric current” was the fourth leading cause of work-related deaths—after “falls,” “transportation incidents,” and “contact with objects and equipment” (McCann et al. 2003). Finding innovative ways to identify, assess, and mitigate electrocution hazards in the early stages of a project would save lives and prevent injuries.

Most studies investigating electrocution accidents have relied on reviews of accident reports. Jenkins et al. (1993) investigated all fatal occupational injuries in the U.S. from 1980 to 1989. Electrocution was reported to be the fifth most frequent cause of occupational deaths in all industries in the U.S. (responsible for 7% of all deaths); however, in the construction industry, electrocution accounted for more than 15% of fatalities, making electrocution the industry’s second most frequent cause of death after falling. Furthermore, the construction industry was the only industry for which electrocution was one of the top three causes of fatalities, and about 39% of all fatal electrocutions happened in the construction industry. According to Cawley and Homce (2008), electricians and their apprentices, followed by construction laborers and electrical power installers, were the most vulnerable groups to electrical fatal injuries.

McCann et al. (2003) studied construction fatalities between 1992 and 1998 using the CFOI database and injury reports. Categorizing workers into electrical and non-electrical trades, they conducted several statistical analyses to find significant differences between these groups. The results revealed that working on or near “live” electric current is a major cause of injury and death among electrical accidents. To reduce the risk of these kinds of accidents, they suggested a permission process for people working on live circuits, along with use of personal locks and training sessions. The 61 non-fatal electrical injuries detailed in this study were limited to one hospital and therefore might not reveal common sources of injuries for electrical contractor's trade.

In another study, Janicak (2008) analyzed CFOI data from 2003 to 2006 to identify influential variables involving electric current. He found that contact with overhead power lines was the most common injury event in both the construction industry (47.2%) and all other industries (43.2%). Other frequent electrocution events in the construction industry that caused fatalities included contact with wiring, transformers, or other electrical components (34.3%); and contact with the electric current of machines, tools, appliances, or light fixtures (12.4%). These were followed by some minor causes, including contact with electric current, unspecified (2.6%); struck by lightning (2.4%); and contact with underground, buried power lines (1.0%). Janicak (2008) also calculated proportionate mortality ratios (PMRs) and found that the construction industry had 20% more fatalities due to contact with wiring, transformers, or other electrical components than was expected statistically. The study concluded that contact with wiring, transformers, and other electrical components contributed to a higher proportion of fatalities in the construction industry compared to other industries. Notably, Janicak's study focused on fatal injuries and did not consider non-fatal scenarios, which are more prevalent among electrical workers.

In an attempt to develop a coding system that would facilitate the categorization of fatal electrocutions and selection of prevention strategies, Chi et al. (2009) examined 255 occupational electrical deaths from 1996 to 2002. They considered variables such as the cause of electrocution, performing task, source of injury, individual factors, and company size, and identified five main accident patterns for electrocution accidents: direct worker contact with an energized power line; boomed vehicle contact with an energized power line; conductive equipment contact with an energized power line; direct worker contact with energized equipment; and improperly installed or damaged equipment. The results of this study could help practitioners determine electrocution protection strategies according to specific characteristics related to accident patterns and variables that impact potential risk factors.

To create a safer environment near dangerous zones of power lines, Hesla (2009) analyzed the underlying reasons for accidents near energized power lines and found the main contributors to be distraction of crane operators and observers, unclear working zones, and inability of workers to indicate the location of power lines. To mitigate risk of such accidents, he suggested providing appropriate equipment, such as line guards, ball markers, cone shaped markers, and line conductor coils. Other researchers developed wearable electric field sensors to notify workers or their supervisors when a worker comes in proximity to, or in contact with, a live power circuit (Neitzel et al. 2001). In addition, anti-current devices that prevent transmission of electrical current from energized power lines to vehicle components (Neitzel et al. 2001) can be used to reduce risk of contact between a boomed vehicle and overhead power line. Alternatively, proximity and current warning devices can notify at-risk workers or operators to avoid potential contact instead of interrupting the transmission of electricity.

As mentioned earlier, these studies focus only on electrical hazards and do not study other types of incidents (e.g., falls) in which electrical contractors may also be involved. Moreover, the data for these studies were usually collected from all trades within the construction industry (e.g., large building construction and heavy civil companies), and therefore their findings might not be pertinent to small specialty (i.e., electrical) contractors. While results of these studies can help to reduce electrocution, to be effective, safety programs need to be designed and implemented for specific trades and based on characteristics of certain tasks and sequences. Therefore, there is a need to study accident patterns among electrical contractors.

3.3 Point of Departure

The results of the literature review indicated four limitations in previous studies related to the occupational health and safety of electrical contractors: (1) some studies focused only on fatalities and ignored other incident outcomes; (2) most of the studies investigated electrical incidents, with only a limited number of studies examining via a large database documented incidents among electrical contractors industry-wide; (3) most of these previous studies only reported descriptive data without using any inferential statistics or machine learning algorithms; and (4) not all of accident types that can happen to electrical workers have been investigated in previous studies.

To address these limitations, researchers need to analyze more recent incident report databases and employ more sophisticated statistical techniques to make inferences that can help practitioners better understand the nature of incidents among electrical contractors and mitigate the risk of injuries and fatalities (Gholizadeh and Esmaeili 2015). Therefore, this study has collected data from the Occupational Safety and Health Administration's (OSHA's) Integrated Management Information System (IMIS) accidents database to analyze accidents related to electrical contractors using a chi-square independence test. One advantage of using OSHA reports in this study is that the reported values for accident factors have been checked and modified based on available summaries, which serves as a major step in understanding the true mechanisms of electrical accidents. We will detail our approach in the methodology section that follows.

3.4 Methodology

To attain this study's research objectives, the analysis first requires reliable national data of incidents among electrical contractors. Consequently, the authors acquired data from OSHA and then conducted a thorough content analysis to (1) ensure the consistency of variables across the data, (2) reduce the ambiguity of reported values, and (3) prepare the data for statistical analysis. Previous studies have successfully used this approach to analyze construction accident databases (Esmaeili et al. 2015 a, b; Gholizadeh and Esmaeili 2015, 2016; Gholizadeh et al. 2018). To investigate and explain the relationship between contributing factors to accidents and the degree of accident injuries, chi-square, Cramer's V tests, and the data mining method known as Classification and Regression Trees (CART) were applied. The rest of this section has been devoted to explaining each of these steps.

3.4.1 Incident Database

Using the OSHA IMIS online database, the authors collected 621 accident reports about injuries involving electrical contractors between 2007 to 2013. While most of these accidents only involved one worker, some cases included multiple injuries; thus, in total, 689 electrical workers' occupational injuries were entered into the database during the seven-year period of this study. One should note that OSHA only requires documentation of 'catastrophic' accidents, wherein a work-related accident caused a fatality, in-patient hospitalization, amputation, or loss of an eye. Therefore, most of the reported accidents had serious outcomes, and only a small fraction of reports included non-hospitalized injuries. Other than non-hospitalized amputations/loss of an eye (which still need reporting), two reasons for the presence of non-hospitalized cases in data include: i) accident affected multiple employees and therefore was reported because some injuries were fatal or needed hospitalization, and ii) employer reported incident even without being required by law. It is also important to note that inclusion in OSHA's database inherently means an accident occurred. Thus, studying this database enables researchers to assess accidents that occurred historically rather assess or predict rates of accidents, which would require data outside the scope of this study.

Within each entry in the large database appears a summary of each accident, as reported by OSHA inspectors, and a limited number of variables used to describe the accident (e.g., event type, source, and cause of injury), its context (e.g., project end-use, type, cost), and its consequences (e.g., nature and degree of injuries, injured part of body). To process data, this study adopted categories found in the Occupational Injury and Illness Classification Manual (OIICM), developed by the U.S. Department of Labor Bureau of Labor Statistics (Bureau of Labor Statistics 2012). In total, 64 cases were omitted from further investigation due to insufficient or missing information, leaving 619 incidents for analysis.

3.4.2 Analysis Methodologies

Pearson Chi-Square Test of Independence and Cramer's V

When variables of the study are nominal, chi-square test can be used to determine significant associations among any pair of variables by calculating a test statistic (i.e., $\chi 2$), which approaches a chi-square distribution (McHugh 2013). Researchers have used this test for more than 100 years (Sharpe 2015): in psychology studies that were published in six journals in 2008 alone, the results of the chi-square test were reported more than 600 times (Bakker and Wicherts, 2011). In construction research, Mustapha and Naoum (1998) have utilized this test to show that the effectiveness of construction managers is related to their age, university degree, membership in professional institutes, overseas experience, and management style. Zuppa et al. (2009) have performed chi-square tests on survey data and found that building information modeling have strong positive impact on projects' success measures such as quality, cost, and schedule. Similar to these efforts, this study has adopted chi-square test of independence to identify accident factors with significant effect on the degree of an injury. Consider a contingency table with R rows, C columns and c cells, the test statistic is:

$$\chi 2 = \sum_{r=1}^{c} \frac{(O_r - E_r)^2}{E_r}$$

Where O_r is the number of observations in cell r and E_r is the expected count in cell r:

$$E_r = \frac{M_R \times M_C}{N}$$

Where N is the total number of observations, M_R is the row marginal for cell r, and M_C is the column marginal for that cell. Once the test statistic is known, it can be compared to a chi-square distribution with (R-1) × (C-1) degrees of freedom to acquire the p-value. A small p-value can reject the null hypothesis that the variables are independent.

To evaluate the strength of significant associations, chi-square tests usually are accompanied by the simple and widely used Cramer's V (Fan et al. 2006; McHugh 2013) test introduced by Cramer (1946):

$$V = \sqrt{\frac{\chi 2}{N[\min(R, C) - 1]}}$$

Higher values of *V* indicate a stronger association among two variables. Using Phi coefficient (i.e., Cramer's V with the sign of the effect), one can also measure the effects at each level of significant factors (Walker and Lev 1953). The effects of several contributing accident factors on the degree of injury would be tested through these three test statistics.

Decision tree learning

A decision tree is a supervised data mining methodology widely used to uncover hidden patterns in categorical data (Ripley 1996; Steinberg 2009; Mistikoglu et al. 2015; Shirali et al. 2018; Choi et al. 2020; Zhu et al. 2021) that can be visually represented by an inverted tree-like structure or diagram. The goal of most decision tree algorithms is to split data by minimizing the impurity of the final categories. Impurity is a general term to define how well a data set is classified in a node, and it is smallest when the node includes just one class of independent data (Breiman et al. 1984). The splitting of the new class is intended to group the data further into more similar sub-classes and hence improves the similarities between the variables within each successive class. The successive split process continues until stop condition is reached (Rivas et al. 2011). By traversing from a root node to a leaf node and fulfilling the split conditions along the way, one can form decision rules (Rivas et al. 2011) which reveal existing associations between predictor variables (Ospina-Mateus et al. 2021) and how they combine to predict the response variable. Since decision trees are easy to use and interpret (Kassambara 2018; Zhu et al. 2021), especially when studying the association between variables or factors (Mistikoglu et al. 2015), they have been applied in similar studies to analyze occupational accidents (Rivas et al. 2011; Cheng et al. 2012; Gholizadeh and Esmaeili 2016; Amiri et al. 2016; Gholizadeh et al. 2018; Shirali et al. 2018; Ospina-Mateus et al. 2021). Common decision tree algorithms include C4.5, classification and regression tree (CART), and Chi-square automatic interaction detection (CHAID). In this study, a CART technique (decision trees) was used because most of the variables considered here are categorical (Cheng et al. 2012).

Classification and regression trees (CART) algorithm

The CART algorithm seeks traits in predictor variables and splits the data (at the root node) into exactly two groups by means of recursive partitioning. These two classes called child nodes are formed through the algorithm’s binary process. The split condition is satisfied when the observations in a class are as homogenous as possible in terms of the dependent variable (Berk 2008). In the process of splitting a categorical variable (such as all the variables in the research work), there are (2^{k-1} -1) possible splits when there are k categories. The CART algorithm chooses the best split and continues the process by splitting the two current classes each into two other new classes until a certain set threshold is reached or until no more useful splits can be achieved. At these points where the splits terminate, these nodes that cannot be split further are referred to as terminal nodes. The sum of all the observations in all the terminal nodes add up to the total number of data points in the root node and data set. At each iteration, the best split chosen by the algorithm is the split that achieves the minimum impurity of the node and is defined as:

$$I(A) = \Phi[p(y = 1|A)]$$

where ϕ represents the measure of impurity, A represents any node of the tree, y is the independent/predictor variable with 2 classes, p represents the probability of the independent variable of class 1 in node A, and I(A) is impurity of node A. ϕ is non-negative and symmetrical and reaches its minimum value when all cases in node A are ones or zeros. When there are an equal number of ones and zeros in the node, ϕ has its maximum value. However, Φ can be defined in multiple ways including entropy and the Gini index. As defined below, the CART algorithm calculates the Gini index to find the impurity:

$$\Phi(p) = p(1 - p)$$

Various tree forms could be the outcome applying various measures of impurity to classify the same data. However, with just a number of cases, the Gini index tends to yield a relatively homogeneous node with low impurity since most of the data points would have similar values for the independent variable. On the other hand, with more cases, a relatively heterogenous node would yield higher impurity since different classes of the independent variables would be somewhat evenly mixed together. This outcome is preferable for data classification instead of the outcomes resulting from algorithms that apply entropy. This is because, in the latter case, nodes are closely alike in size and homogeneity (Berk 2008). After all the nodes have been formed, the algorithm would then assign classes to the terminal nodes by calculating the proportion of classes of independent variables. Hence, the node will be labelled with the class of which it has the highest proportion. For example, if the amount of 0s (in this study, non-electrical injuries) is greater than half of the total number of data points in that node, then the node will be labelled as 0 (non-electrical).

In this study, in the development of the CART, the *nature of injury* was the target (response) variable and has two categories: electrical and non-electrical. The electrical category is about 40% of the data and the remaining 60% represents the non-electrical category. On the other hand, the predictor (explanatory) variables include *the end-use, project type, project cost, source of injury, environmental factor, human factor, and cause of injury*. The data set was split into training and testing (validation) data set in the process of applying CART to the raw data set. Out of the 619 accident reports used in this analysis, a random selection of 496 (i.e. 80%) of the accident reports was used as the training data set and was trained with the CART algorithm in R (R Core Team 2013). The remaining 123 (i.e., 20%) was used for validation/testing. The classification and regression training (CARET) package (Kuhn 2015) and the recursive partitioning and regression trees (RPART) package (Therneau and Atkinson 2019) in R (Kuhn 2015) were used to develop the decision tree for: (1) forecasting the nature of injuries due to an accident during an electrical project; (2) identifying the factors that are most important in forecasting the nature of electrical project injuries.

Prediction Accuracy

In a decision tree analysis, we measure predictive accuracy by instances correctly classified. The Kappa statistic is a measure of a model's accuracy that estimates how well the predictions of the model and the actual classifications match or agree. It estimates the extent of the model's ability to predict better than any random classifier (e.g., predicting expected accuracy). Hence, the model's ability to observe and predict correctly yields the Kappa statistic. It uses the observed and predicted values for each class of independent variables to estimate the model's Kappa value which range from -1 to 1 (McHugh 2012). Values less than zero indicate no agreement, 0.01-0.20 indicate none to slight agreement, 0.21-0.40 indicate fair agreement, 0.41-0.60 indicate moderate agreement, 0.61-0.81 indicate substantial agreement, and 0.81-1.0 indicate perfect agreement (Sim and Wright 2005).

Precision

Precision is a measure of the proportion of the time you were right when you declared an instance (a positive). In relation to this study, the precision of the proposed decision tree model is the proportion of correct prediction of electrical injuries (true positives) to all accident reports that are predicted as electrical injuries. In other words, the precision of the model is the ratio of true positives (TP) to total number of cases predicted as positives (i.e., TP + FP).

Sensitivity

Sensitivity or recall measures the proportion of actual positives that you declared were positives. With respect to this study, the sensitivity or recall of the proposed decision tree model is the proportion of correct prediction of electrical injuries (true positives) to accident reports that are actual electrical injuries. In other words, the sensitivity or recall of the model is the ratio of true positives (TP) to total number of cases that are actual positives (i.e., TP + FN).

Specificity

The specificity of the proposed decision tree model is the proportion of correct prediction of non-electrical injuries (true negatives) to accident reports that are actual non-electrical injuries. In other words, the specificity of the model is the ratio of true negatives (TP) to total number of cases predicted as negatives (i.e., TN + FP).

Cross Validation Analysis

Cross validation is a technic that tests how well the results of a model will generalize to new data. It involves splitting the data set into a training set and a testing set. The model is given the training set on which the training is carried out. After the training, the set-aside independent testing data set that was unseen by the model is used to evaluate the results of the training. Cross validation is carried out to reduce bias and variance in the training data set in the development of the proposed model. A k-fold cross validation involves splitting the data into k folds, and then using 1 fold as the testing set and the remaining k-1 folds together as the training data set. Folds 1 through k individually gets used 1 time as the testing set and k-1 times as part of the training set in the k different fittings of the model. In this study, the accident reports making up the training data set were split into ten folds for cross validation. This is to improve the predictive capability of the model especially on new cases (Rivas et al. 2011). The training data set contained 496 accident reports and was randomly split into ten (k) folds. One of the ten folds was in turn set aside as the sub-testing set (for validation purpose) while the remaining nine (k-1) formed the sub-training set. In an iterative process, every one of the ten folds in turn had a chance of being the sub-testing set in only one fitting of the model and part of the sub-training set in the others. The ten-fold cross validation resulted in ten iterations/fittings/resamples, ten Kappa values, and ten (sub-testing) prediction accuracies as outlined in the results section. The average accuracy of these ten resamples was computed and used in the development of the decision tree model.

3.5 Results

The results are presented in two separate sections. First, the explanatory statistics of catastrophic accidents that have affected electrical contractors from 2007 to 2013 are presented. The emphasis is on the degree of injury as the most evident outcome of these incidents. Then, the associations between degree of injury and several variables—such as type of projects (i.e., project end-use, type, and cost), worker’s task (i.e., source and cause of injuries), and other outcomes of an accident (i.e., nature of injury, injured part of body, and event type)—are tested.

3.5.1 Exploratory Analysis

Within the accident reports in the database, in total, 226 (37%) of accidents resulted in a fatality. The remaining (non-fatal) injuries were filed into two categories: 343 (55%) hospitalized injuries and 50 (8%) non-hospitalized injuries. As mentioned in the research methods, eight variables were coded in the content analysis to better understand the nature of accidents. The salient results and the rates of 'degree of injury' for each category are presented in Table 3.1. One should note that a *fatality rate* in this study represents a specific quantity *given the occurrence of a catastrophic accident.* Hence, this fatality rate is the share of fatal injuries from the total number of *catastrophic injuries* in the dataset and should not be confused by estimated rates that are calculated using full-time equivalent workers. In other words, a fatality rate of 37% simply means that 37% of all injury events in the data led to a death; not that 37% of electrical workers would die on the job.

As far as end-use is concerned, electrical accidents occurred mostly in building projects (77%), with commercial buildings being the dominant end-use. In non-building projects, *utility systems* (particularly power and communication lines) were the primary environments. The results indicate that different environments have quite similar fatality rates.

Another variable that can describe the project condition is the project type. Project type, as opposed to end-use, implies the purpose of construction projects and not their context. The project types with the most accidents were *new project or new addition* (36%), *alteration or rehabilitation* (28%), and *maintenance or repair* (25%). Conditional on an accident occurring within these project types, the fatality rates for these accidents were 40%, 40%, and 35%, respectively, which are all close to the total fatality rate among electrical contractors (i.e., 37%). However, while *demolitions* represent only 2% of projects, their accidents' fatality rate (58%) is much higher than other types of projects.

Regarding project costs, a large proportion of accidents occurred during relatively small-budget projects: around 74% of projects in have *budgets less than $500,000*. While most of cost categories present fatality-rates close to the average, two categories with the highest budgets show different values: projects with *budgets between $5 to $20 million* have significantly higher fatality-rates (i.e., 52%) while projects with *more than $20 million budgets* have fatality rates of 24%, which is much less than the average.

As far as the sources of injuries were concerned, the largest category is *parts and materials*, as it involves all the electrical parts. Regarding the degree, though, *vehicles* and *machinery* caused higher fatality rates, with 58% and 44% of accidents being fatal, respectively.

With regards to causes, *installing equipment (HVAC and other)*, *interior plumbing, ducting, electrical work*, and *installing plumbing, lighting fixtures* were three individual causes with the most frequency. However, in terms of the severity, the *fencing, installing lights, signs, etc.* cases represent the highest fatality rate.

To explain the circumstances of an accident, OSHA reports the event type. While having high proportions of *exposure to electricity* with high fatality rates was expected, the findings also show that *fall* accidents are quite prevalent among electrical contractors.

For the nature of injuries consideration, only 3 categories represent 72% of all injuries: *fractures*, *electrocutions*, and *burns*. However, in terms of the degree of the injuries, *electrocutions* and *concussions* have given rise to the most severe injuries.

Regarding, different body parts that were injured in this data set, *upper extremities*, *head*, and *body system* were the parts with the most injuries, respectively. Injuries to the *body system* resulted in the highest fatality rate (64%), indicating how electricity can critically affect the whole body.

Table 3.1. Accident characteristics among electrical contractors

	Variables	Frequency (%[1])	Degree of Injury (%)		
			Fatality[1]	Hospitalized	Non-Hospitalized
End-use	Highway, street, and bridge	28 (5)	39	57	4
	Nonresidential building	406 (66)	36	55	9
	Other heavy and civil engineering	28 (5)	39	57	4
	Residential building	67 (11)	34	54	12
	Utility system	90 (15)	38	58	4
Project Type	Alteration or rehabilitation	173 (28)	40	52	8
	Demolition	12 (2)	58	42	0
	Maintenance or repair	155 (25)	35	57	8
	New project or new addition	222 (36)	40	50	10
	Other	57 (9)	14	83	4
Project cost	$50,000 and less	276 (45)	34	58	8
	$50,000–$250,000	115 (19)	39	56	5
	$250,000–$500,000	60 (10)	38	43	18
	$500,000–$1,000,000	45 (7)	38	56	7
	$1,000,000–$5,000,000	63 (10)	40	54	6
	$5,000,000–$20,000,000	27 (4)	52	37	11
	$20,000,000 and more	33 (5)	24	70	6
Source of injury	Machinery	45 (7)	44	56	0
	Parts and materials	286 (46)	39	52	9
	Structures and surfaces	97 (16)	31	62	7
	Tools, instruments, and equipment	118 (19)	24	66	10
	Vehicles	50 (8)	58	40	2
	Other sources	23 (4)	35	52	13
Causes	Fencing, installing lights, signs, etc.	30 (5)	53	43	3
	Installing equipment (HVAC and other)	121 (20)	40	55	6
	Installing plumbing, lighting fixtures	90 (15)	44	51	4

	Interior plumbing, ducting, electrical work	102 (17)	28	64	8
	Temporary work (building, facilities)	35 (6)	43	43	14
	Other	111 (18)	32	55	13
	Not reported	130 (21)	32	59	9
Event type	Caught in/between	38 (6)	34	58	8
	Exposure to electricity	253 (41)	44	47	8
	Fall	210 (34)	30	67	3
	Struck-by	78 (13)	37	46	17
	Other	40 (6)	23	63	15
Nature of injury	Amputations, avulsions, enucleations	21 (3)	0	71	29
	Bruises, contusions	19 (3)	16	37	47
	Concussions	44 (7)	59	41	0
	Cuts, lacerations	24 (4)	13	62	25
	Electrical burns	84 (14)	1	83	16
	Electrocutions, electric shocks	166 (27)	67	28	5
	Fractures	193 (31)	20	78	2
	Non-specified injuries and disorders	35 (6)	86	14	0
	Other	33 (5)	39	49	12
Injured part of body	Body system	110 (18)	64	32	4
	Head	141 (23)	45	48	7
	Lower extremities	60 (10)	0	93	7
	Multiple body parts	78 (13)	19	78	3
	Trunk	74 (12)	39	56	5
	Upper extremities	156 (25)	31	53	17

[1] The percentages were rounded to the closest integer and some cases might not add up to 100%.

Other variables such as human and environmental factors were also reported by OSHA inspectors. Table 3.2 reports some of the more common factors (i.e., each factor represents at least 5% of the frequency after excluding the missing cases) for each variable. The findings show that while *misjudgment* is by far the main human factor, problems with *lockout/tagout procedures*, *inappropriate position for task*, and neglecting necessary *safety devices* were more dangerous. For environmental factors, *work surface and facility layout condition* is the most common factor among electrical workers that leads to accidents. However, *material-handling equipment or method, overhead moving- or falling-object action*, and *squeeze-point action* caused higher fatality rates. The variables in Table 3.2, though, were excluded from the following statistical analyses due to the high number of missing values.

Table 3.2. Frequency and fatality rate for human and environmental factors

Main category	Subcategory	Frequency	Fatality rate (%)
Human factors	Misjudgment in hazardous situation	207	37
	Malfunction in lockout/tagout procedure	61	53
	Safety devices removed or used inappropriately	39	39
	Insufficient or lack of personal protective equipment	35	34
	Inappropriate equipment for operation	34	35
	Inappropriate position for task	26	42
	Malfunction in securing or warning operation	26	31
Environ mental factors	Work surface or facility layout condition	125	30
	Material-handling equipment or method	39	51
	Overhead moving- or falling-object action	39	49
	Temperature tolerance	18	11
	Squeeze-point action	17	47
	Flying-object action	16	25

3.5.2 Chi-Square Independence Test and Cramer's V

As mentioned earlier, the second objective of the study was to test whether the degree of injuries (i.e., fatality, hospitalized injury, and non-hospitalized injury) was associated with the type of projects, the worker's task, or other outcomes of an accident. Table 3.3 shows the results of the Chi-square test for eight variables.

Table 3.3. Associations between degree of injury (i.e., fatality, hospitalized injury, and non-hospitalized injury) and eight accident factors

Variables against degree of injury	Chi-square statistic	Degree of freedom (d.f.)	p-value	Cramer's V
Project end-use	4.86	8	0.77	-
Project type	23.48	8	0.00	0.14
Project cost	18.78	12	0.09	-
Sources of injury	27.69	10	0.00	0.15
Causes of injury	17.42	10	0.07	-
Event type	33.51	8	0.00	0.17
Nature of injury	273.41	16	0.00	0.47
Injured part of body	111.96	10	0.00	0.30

Non-hospitalized and hospitalized injuries can be combined into one "Non-fatal" category—as opposed to the "Fatal" injuries—to test the effect of accident factors on degree of injury more directly. Table 3.4 presents the results of this test for eight variables.

Table 3.4. Associations between degree of injury (i.e., fatality and non-fatal injury) and eight accident factors

Variables against degree of injury	Chi-square statistic	Degree of freedom (d.f.)	p-value	Cramer's V
Project end-use	0.40	4	0.98	-
Project type	16.86	4	0.00	0.17
Project cost	6.33	6	0.39	-
Sources of injury	21.49	5	0.00	0.19
Causes of injury	11.89	5	0.04	0.14
Event type	13.90	4	0.01	0.15
Nature of injury	201.18	8	0.00	0.57
Injured part of body	87.82	5	0.00	0.38

The p-values in Table 3.3 and Table 3.4 indicate that, at the significance level of 0.05, the degree of injury is significantly affected by the type of project, sources of injury, type of accident, nature of injury, and injured part of body. Cause of injury can be considered significant only when hospitalized and non-hospitalized injuries are combined (Table 3.4). This suggests that for this variable the result should be interpreted more carefully. For the rest of variables, it is safe to continue with 'fatal' versus 'non-fatal' scenario as they are significant in both cases. The results of both tables, however, show a lack of evidence to claim an association between degree of injuries and end-use, nor between degree of injuries and cost of projects. The Cramer's V values show the amount of association between the significant factors and the degree of injuries. Nature of injuries and part of body have the highest association with the degree of injury. To locate the effects among these two significant factors, the values of Phi coefficients are calculated for each level of nature and body parts (Table 3.5).

Table 3.5. The effect of different natures of injury and parts of body on degree of injury

Variable	Level	Phi Coefficient
Nature of injury	Amputations, avulsions, enucleations	-0.14
	Bruises, contusions	-0.08
	Concussions	0.13
	Cuts, lacerations	-0.10
	Electrical burns	-0.29
	Electrocutions, electric shocks	0.38
	Fractures	-0.23
Part of body	Body system	0.27
	Head	0.10
	Lower extremities	-0.25
	Multiple body parts	-0.14
	Trunk	0.02
	Upper extremities	-0.07

3.5.3 Decision Tree Analysis

Model Interpretation

The proposed decision tree model in Figure 3.1 displays the classification of the nature of injury of electrical contractors as observed in the accident reports analyzed in this experiment. This tree was generated from 496 accidents reports that make up the training data set and seven project information/features/attributes namely: end-use, project type, project cost, source of injury, environmental factor, human factor, and cause of injury. The decision tree model was built for predicting the target variable which is the nature of injury. In this experiment, the nature of injury has two categories: electrical and non-electrical. The nature of injuries involving electrical burns and Electrocution (electrical shocks) are regarded as electrical and the rest (such as amputations, avulsions, enucleations, bruises, contusions, concussions, etc.) are labelled non-electrical as shown in Table 3.1. The tree was pruned with R tuning parameters to avoid overfitting (Zhu et al. 2021) and the optimal model with the best accuracy was selected. It can been be seen from Figure 3.1 that the decision tree model has a total of thirteen nodes of which seven nodes are leaf nodes. Figure 3.2 provides some explanatory notes on the decision tree model.

By looking at Figure 3.1, it can be observed that out of the 496 accident reports in the training data set, a total of 463 reports (which accounts for 93.35%) were correctly classified and 33 reports (which accounts for 6.65%) were incorrectly classified. It can be seen that the target attribute in the root node is the nature of injury; the first level attribute is the source of injury as parts and material; the environmental factor labelled overhead moving- or falling-object action is the second level attribute; this is followed by the other environmental factor as the third level attribute; work surface or facility layout condition as an environmental factor is the fourth level attribute; the other cause of injury is the fifth level attribute; and lastly project type as new project or new addition is the fifth level attribute. The variable importance section below outlined the list of essential variables according to their order of importance in the development of this proposed decision tree model.

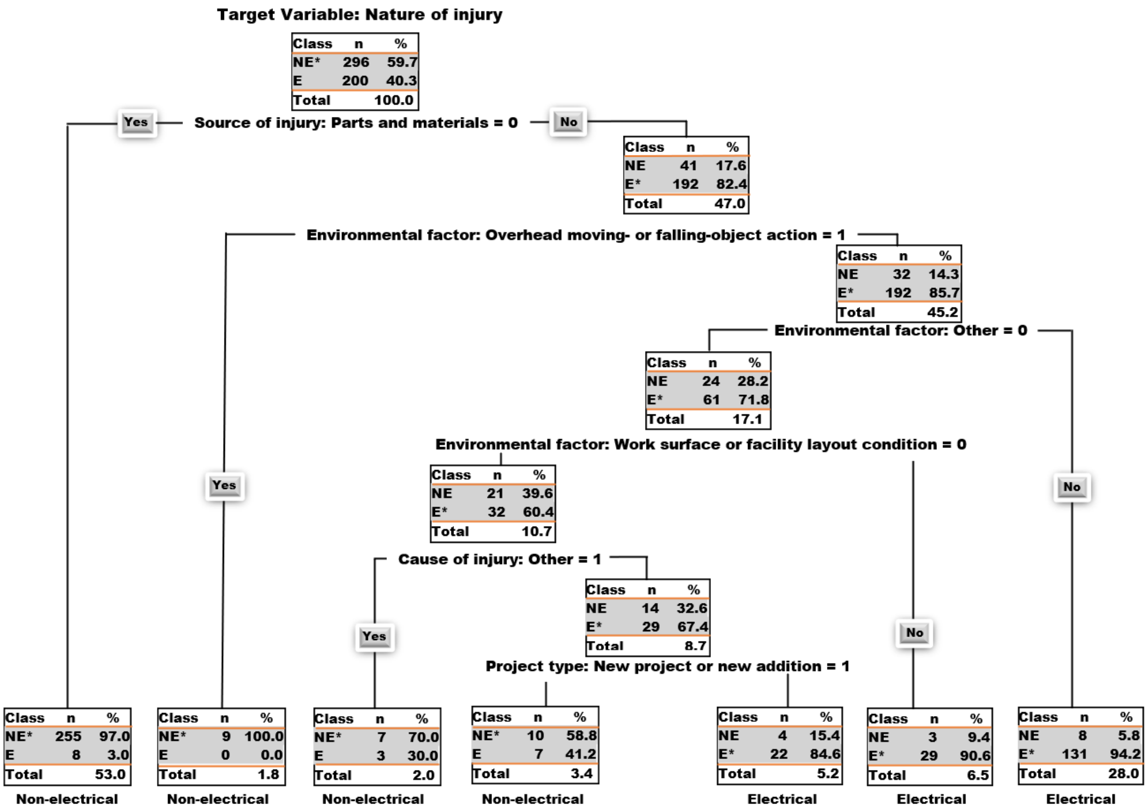


Figure 3.1. Decision tree for the prediction of Nature of Injury

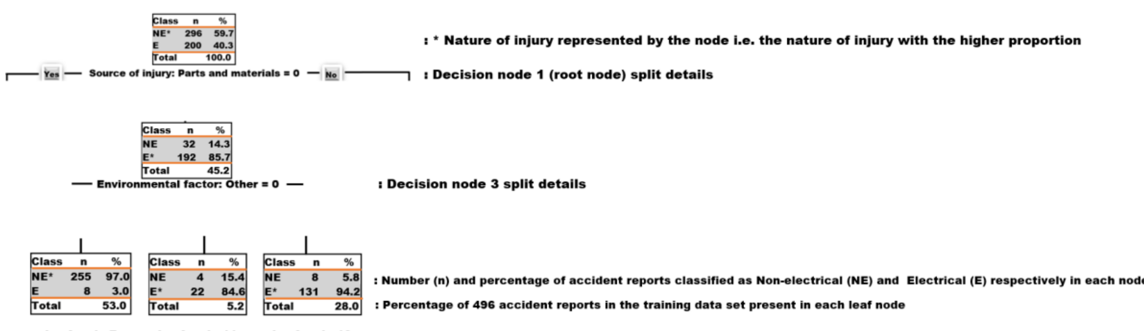


Figure 3.2. Explanatory notes on decision tree representations

In Figure 3.2 above, the nodes represent split points where the observations within that node are split into the two classes: non-electrical (NE) or electrical (E). The number of observations in each node classified as NE and E are listed under the number of observations (n) for NE and E respectively. These two numbers for these two classes add up to the total number of observations in that node. In the same way, the percentage of observations in each node classified as NE and E are listed under the percentage of observations (%) for NE and E respectively. These two percentages for these two classes add up to 100% of the total number of observations in that node. However, the percentage total in each node represents the percentage of 496 accident reports in the training data set that is present in that node. For instance, from start, node 1 in Figure 3.2 has 296 NE and 200 E data points which adds up to 496 accident reports which is a 100% of the total observations or accident reports in the training data set. Hence, node 1 is labelled non-electrical (NE*) since the NE class has a higher proportion of the observations in this node. Leaf node 11 in Figure 3.2 has 4 NE and 22 E data points which adds up to 26 observations which is about 5.2% of the total accident reports in the training data set (i.e. 5.2% of 496 = 26). Hence, Node 11 is labelled electrical (E*) since the E class has a higher proportion of the accident reports in this node. In Figure 3.2 above, source of injury: parts and materials = 0 could be read as "if source of injury is not parts and materials". Hence if this statement is true (Yes), you go left and if this statement is false (No) you go right. A left branch of the above tree in Figure 3.1 and Figure 3.2 is always a Yes-turn while a right branch is a No-turn. In this study, nodes were counted from top to bottom, left to right, from 1 to 13. On the other hand, as seen in Figure 3.1, environmental factor: overhead moving- or falling-object action = 1 can be read as "if environmental factor is overhead moving- or falling-object action". If this statement is true, you take a Yes-turn, and if this statement if false, you take No-turn. These terminologies were used in the development of the decision rules below.

Decision Rules

Decision rules are usually formed with each leaf node of the decision tree model. Table 3.6 shows list of unique decision rules developed for every leaf node of the decision tree in Figure 3.1. The proposed decision rules are discussed further in the discussion session.

Table 3.6. Decision rules derived from the proposed decision tree model

S/N	Node	Decision rules
1	7	If the source of injury is not parts and materials, then the nature of the injury is non-electrical
2	8	If the source of injury is parts and materials, and the environmental factor leading to accident is overhead moving- or falling-object action, then the nature of the injury is non-electrical
3	9	If the source of injury is parts and materials, the environmental factor leading to accident is neither overhead moving- or falling-object action, nor unknown, nor work surface or facility layout condition, and the cause of injury is unknown, then the nature of the injury is non-electrical
4	10	If the source of injury is parts and materials, the environmental factor leading to accident is neither overhead moving- or falling-object action, nor unknown, nor work surface or facility layout condition, the cause of injury is not unknown, and the project type is new project or new addition, then the nature of the injury is non-electrical
5	11	If the source of injury is parts and materials, the environmental factor leading to accident is neither overhead moving- or falling-object action, nor unknown, nor work surface or facility layout condition, the cause of injury is not unknown, and the project type is not new project nor new addition, then the nature of the injury is electrical
6	12	If the source of injury is parts and materials, the environmental factor leading to accident is neither overhead moving- or falling-object action, nor unknown, but work surface or facility layout condition, then the nature of the injury is electrical
7	13	If the source of injury is parts and materials, the environmental factor leading to accident is not overhead moving- or falling-object action, but unknown, then the nature of the injury is electrical

Decision Tree Model Accuracy

The evaluation of the proposed decision tree model involved applying the model to the testing data set which contains 123 accidents reports that were set aside for validation. The results of this test are presented in the confusion matrix in Table 3.7. The confusion matrix compares the model predictions of the nature of injuries versus the actual classifications of the nature of injuries of the accident reports. The shaded diagonal in Table 3.7 shows that one hundred and sixteen accident reports are correctly classified. Hence, the accuracy of the proposed decision model in predicting the nature of injury of the accidents reports in the testing data set is 94.3%. On the other hand, there are seven misclassified accident reports resulting in an error rate or misclassification rate of 5.7%.

Table 3.7. Confusion matrix of the testing data set

	Actual/Reference	
Prediction	Electrical	Non-electrical
Electrical	45	2
Non- electrical	5	71

Table 3.8. Evaluation of the confusion matrix and decision tree accuracy.

Evaluation Statistics	Results
Precision	0.957
Sensitivity/Recall	0.900
Specificity	0.973
Accuracy	0.9431
95% Confidence Interval	(0.886, 0.977)
No Information Rate	0.594
P-Value [Acc > NIR]	<2e-16
Kappa	0.881

In the confusion matrix in Table 3.7, one can see that true positives (i.e. accident reports involving actual electrical injuries and that are predicted to involve electrical injuries) are 45, true negatives (i.e. accident reports involving actual non-electrical injuries and that are predicted to involve non-electrical injuries) are 71, false positives (i.e. accident reports involving actual non-electrical injuries but were predicted to involve electrical injuries) are 2, and false negatives (accident reports involving actual electrical injuries but were predicted to involve non-electrical injuries) are 5. Using these values, one can calculate the precision, sensitivity (recall), and specificity resulting from applying the proposed decision tree model on the testing set as presented in Table 3.8. According to the evaluation results reported in Table 3.8, it can be seen that the proposed decision tree model in this study is dependable as it is at per with similar studies (Rivas et al., 2011; Mistikoglu et al., 2015) and as seen by an accuracy of 94.31% on new data, Kappa value of 0.881, and a P-Value (Acc > NIR) of 2e-16. Hence, with 94.31% accuracy (Table 3.8), it successfully predicted the nature of injuries of electrical contractors into electrical injuries and non-electrical injuries using the accident reports in the testing data set.

The value obtained for the no-information rate is 0.594 and it gives an idea of the proportion of accident reports (0.596) that involve non-electrical injuries in the training data set. This means that without the proposed decision tree model and prediction is done by guessing each accident report in the testing data set to be, say, non-electrical injuries, the accuracy of prediction would be the no-information rate of 0.594 (i.e., approximately equal to the probability (0.596) of occurrence of non-electrical injuries within the training set). Also, this no-information rate of 0.594 is less than the accuracy of the proposed decision tree model presented in this study 95% of the time (i.e., 95% confidence interval of 0.886, 0.977). This means that there is a 95% chance that the true accuracy of this proposed decision tree model lies between 88.63% and 97.68%. Hence, the 94.31% accuracy of the proposed decision tree model in this study is greater and better than the non-information rate of 59.4%. The model also has a significantly better performance than chance (i.e., accuracy > no information rate) as suggested by a p-value less than 2e-16. Hence it can be said that there is sufficient evidence that the accuracy of this model is greater than the no information rate with a p-value of < 2e-16. The kappa value of 0.88 obtained in this study indicates a perfect agreement between the classification predictions made from the proposed decision tree model and the actual classifications.

Evaluation of Variable Importance

Unlike linear regression, all important variables may not show up on the decision tree model as a node splitter. In CART, the contribution made by a predictor variable is determined by primary splits and surrogate splits. In the development of a tree, the variable that appears in the tree structure is the primary splitter, but CART also keeps track of surrogate splits and uses them as an alternative whenever the variable is missing. With the RPART package and the summary function in R, the evaluation of variable importance could be assessed as shown in Table 3.9. The rounded variable importance scores presented in Table 3.9 are scaled up to 100% as reported by the summary printout in R. Variables with an importance score less than 1 are omitted. In assigning importance to variables, the loss function (e.g., mean squared error) that can be attributed to each variable at each split is tabulated and summed. Hence, in evaluating variable importance, the goodness of split measure is summed up for each split where it is a primary variable and where it is a surrogate. Therefore, a variable that doesn't show up in a tree may be assigned a high variable importance depending on the measures obtained where it is a surrogate and primary splitter. Using this approach, variable masking and nonlinear correlation among variables could be revealed in the ranking of variable importance (Steinberg, 2009) Some variables in Table 3.9 such as "Source of injury: Structures and surfaces;" "Human factor: Malfunction in lockout/tagout procedure;" and "Environmental factor: Overhead moving- or falling-object action" have nonzero importance in the development of the tree but still did not show up in the tree structure. This means that they played an important role in the development of the tree in Figure 3.1 strictly by acting as surrogates to the other splitting variables that showed up on the tree.

Table 3.9. Variable importance of the proposed decision tree model

Variables/Attributes	Importance Score
Source of injury: Parts and materials	49
Source of injury: Tools, instruments and equipment	13
Source of injury: Structures and surfaces	11
Human factor: Malfunction in lockout/tagout procedure	9
Environmental factor: Other	6
Environmental factor: Overhead moving- or falling-object action	4
Cause of injury: Interior plumbing, ducting, and electrical work	4
Environmental factor: Work surface or facility layout condition	2
Project type: New project or new addition	1
Cause of injury: Other	1

By looking at Figure 3.1 and Table 3.9, one can see that *parts and materials* (root node) followed by *tools, instruments, and equipment* as sources of injury are the most relevant variables for the prediction of the nature of injury. It can also be observed that the two least important variables for predicting the nature of injury in electrical projects are *new project or new addition* project type and *other* cause of injury.

Cross Validation results

The cross-validation results of the training data set are presented in Table 3.10. The prediction accuracy of the training data set was arrived at by averaging the prediction accuracy obtained from each of the sub-testing data set (i.e., the 10 resamples in Table 3.10). This gives the mean training prediction accuracy of the proposed model. Summary statistics of the ten-fold cross validation is shown in Table 3.11. As one can see, the mean prediction accuracy of the ten-fold cross validation is 0.917, and the minimum and maximum accuracies obtained are 0.840 and 0.980 respectively. In the same way, the mean kappa value obtained in the ten-fold cross validation is 0.829, and the minimum and maximum accuracies obtained are 0.672 and 0.958 respectively. In comparison with similar studies (Rivas et al., 2011; Mistikoglu et al., 2015), these results (e.g., accuracy on new data = 94.31%, Kappa = 0.881, and P-Value: Acc > NIR = 2e-16) indicate that the proposed model is reliable in predicting the nature of injuries that could occur in electrical projects.

Table 3.10. Prediction accuracy of the ten-fold cross validation

S/N	Accuracy	Kappa	Resample
1	0.918	0.834	Fold 1
2	0.980	0.958	Fold 2
3	0.959	0.914	Fold 3
4	0.918	0.828	Fold 4
5	0.898	0.787	Fold 5
6	0.900	0.790	Fold 6
7	0.900	0.800	Fold 7
8	0.940	0.874	Fold 8
9	0.920	0.836	Fold 9
10	0.840	0.672	Fold10

Table 3.11. Summary statistics of the ten-fold cross validation

S/N	Statistic	Accuracy	Kappa
1	Minimum	0.840	0.672
2	First quartile	0.900	0.792
3	Median	0.918	0.831
4	Mean	0.917	0.829
5	Standard deviation	0.038	0.078
6	Third quartile	0.935	0.865
7	Maximum	0.980	0.958

3.6 Discussion

This study aimed to investigate which accident factors have a statistically significant effect on the outcome of accidents (i.e., degree of injury) among electrical contractors. To address the objective, the authors reviewed every report manually to ensure the quality of the final data points, as the quality of data is paramount in any scientific study, without exception. This study also discusses the contributing factors to catastrophic construction accidents among electrical contractors. The results of exploratory, statistical, and machine learning analyzes are discussed here.

3.6.1 Exploratory Analysis

To reveal nature of accidents, summary statistics of 11 variables (e.g., end-use, project type, nature of injury) were reported. As noted by Lee et al. (2012), safety-risk factors in projects are determined by location, type and complexity of projects. Better understanding these influential variables will enable safety managers to strategically allocate their limited safety resources, particularly in small enterprises. As far as electrical contractors are concerned, given situations in which an accident occurred, the results show that the majority of accidents happened in *nonresidential buildings* (e.g., commercial, industrial), *new construction*, and *small projects* (i.e., $50,000 or less). Historic events do not inherently predict future events, but the statistical significance of past accident factors describe conditions in which future accidents may face added risk. Therefore, contractors who are working on these projects should plan for precautionary actions and consider larger contingencies in their budgets.

The main source of injury is *parts and materials* (e.g., electrical parts)—representing 46% of accident sources—followed by *tools, instruments, and equipment* (19%), and *structure and surfaces* (16%). These findings are compatible with a study conducted by Abudayyeh et al. (2003): For non-fatal accidents among electrical contractors from 1992 to 1998 reported in BLS, Abudayyeh et al. found *parts and materials* as the most common source of injury (25%), followed by *structures and surfaces* (i.e., floors, walkways, or ground surfaces – 19%) and all *other* sources (17%). In comparison with Abudayyeh and his colleagues' results, this study shows a large increase in the share of *parts and materials,* which may suggest the need for more training regarding electrical sources. The share of accidents sourced in *tools, vehicles*, and *machinery* has also increased by 10%, 4%, and 3%, respectively, which may indicate the growing application of new tools and machines in construction and may emphasize the need for further, task-specific safety training and planning. One other notable outcome here is that due to the application of the comprehensive content analysis in this study, the *other* category is much smaller here (4%) compared to 17% in Abudayyeh et al. (2003). We posit that this difference can beneficially increase our understanding of accident mechanisms.

The most frequent nature of electrical contracting injuries were *fractures* (31%), *electrocutions* (27%), and *electrical burns* (14%). These are in contrast with the findings of Abudayyeh et al. (2003), as that study reported "sprain and strains" (37%), "all other natures" (23%), and "cut and punctures" (13%) as the top three injury types among electrical workers. Only three cases of "strains/sprains" were reported in OSHA database, which can be attributed to the fact that while these injuries are prevalent among construction workers (Kisner and Fosbroke 1994; Kines et al. 2007; Choi 2015), since they usually do not lead to very serious consequences—such as permanent disability or fatality—they may have not been reported to OSHA inspectors. Indeed, this study's findings relate to more severe injury types that might otherwise be neglected or washed out due to their relatively low frequency. Such a nuance demonstrates the benefit of focusing this study on the accidents within OSHA's catastrophic database.

Considering body parts, the OSHA accident reports have *upper extremities* (25%), *head* (23%), and *body system* (18%) as the main injured body parts. Also, *lower* *extremities* and *trunk* were the two parts with the lowest frequency. Regarding severity, the chance of fatality is higher when the body system or head are injured. Further investigation showed that, most of the incidents in which the whole body was affected were cases of exposure to electricity. It's important to note that the electricity usually enters from upper extremities (e.g., fingers, hands) and most of the time, it's only the magnitude of flow which differentiates between a small injury in upper extremities and a serious (usually fatal) injury in body system. In other words, injuries to upper extremities must be analyzed more carefully, especially among electrical contractors, as they can rapidly escalate to situations in which the whole body can be severely affected.

When considering only non-fatal accidents, Abudayyeh et al. (2003) reported that "contact with objects" (including struck-by and caught in/between), "overexertion," and "falls" are the most common accident types, representing 31%, 22%, and 20% of nonfatal accidents, respectively. Comparatively, in accident reports from OSHA regarding non-fatal cases, *falls* (37%), *exposure to electricity* (36%), and *contact with objects* (19%) caused most of the injuries. Putting aside "transportation incidents" (19%), the main three events for fatal accidents in BLS data were "exposure to harmful substances and environments" (50%), "falls" (21%), and "contact with objects and equipment" (7%). Among fatal cases of OSHA reports, *exposure to electricity* is also the leading event, causing exactly the same 50% of deaths followed by *falls* (28%) and *contact with objects* (19%). The order and magnitude of accident types are very similar in both studies especially in fatal cases. The share of exposure to electricity in fatal cases is much lower for the entire construction industry (18% in BLS data) which can be attributable to a much wider range of work categories in the industry compared to a more limited activities among specialty trades such as electrical contractors. This finding further emphasizes on the necessity of investigating accidents within a specific trade, since focusing one particular type of accident (e.g., exposure to electricity) can reduce the number of fatalities/severe injuries dramatically. For electrical contractors this means more electrical training on main sources and causes of exposures to electricity and ultimately decrease the more severe injuries presented in OSHA's data.

By using these findings from exploratory data analysis, one can start to decipher some of the more common accident scenarios among electrical workers. For instance, *electrocutions* and *burns* to *body system* and *upper extremities* often happened historically in *exposure to electricity* accidents wherein *parts and material* are the source of injury and *small, new, nonresidential buildings* are the location of accident. Though they used different methods, some studies have similarly examined the associations linking accident factors (Chi et al. 2004, 2005, 2012). For instance, Chi et al. (2005) investigated fatal fall accidents to demonstrate how different types of falls are linked to specific causes; the team then suggested several prevention measures based on strong links between a cause and its consequent accident. Chi et al. (2012) also found that the source and cause of injury are significant factors in classifying accident scenarios.

Furthermore, industry would benefit from studying the effect of different factors in fatal scenarios more. According to OSHA reports, and as we describe above, 37% of all catastrophic accidents that occurred to electrical contractors between 2007 to 2013 were fatal. Our investigation suggests when the project type is *demolition*, project budget is *between $5 million to $20 million,* sources of injury is *vehicles or machinery,* causes of injury is *fencing, installing lights and signs*, or *installing plumbing and lighting fixtures, or temporary work*, event type is *exposure to electricity*, injury type is *electrocutions or concussions*, and body part is *body system or head,* there is a higher chance that an accident lead to a fatality (i.e., at least 5% more than the average fatality rate of 37%). Also, human factors—such as *malfunctioned lockout/tagout procedures and inappropriate position for task*—and environmental factors—such as *material-handling equipment/method, overhead moving-/falling-object action, and squeeze-point action*—have contributed to fatalities at larger rates. When planning for injury prevention practices, existence of any of these factors could raise a red flag and consequently, safety managers can design customized interventions to reduce severity and frequency of incidents among electrical contractors. Other than their exploratory values in showing more hazardous situations for electrical contractors, these findings propose that the degree of injury might be affected significantly by some factors that are related to a project's characteristic or a worker's task.

3.6.2 Statistical Analysis

To examine these potential associations, the research team applied chi-square independence test and found that, except for the project end-use, cost, and to a lower degree cause of injury, five accident factors have significant influence on the degree of an injury (Table 3.6 and Table 3.7). Ordered by their Cramer's V values, nature of injury and part of body correlate with the degree of injury most, followed by source of injury, project type, and event type. Considering the effects of a single nature of injury on the degree of injury, this study has found that electrocutions and concussions are associated with more fatal injuries. Regarding injured parts of body, this study found that, among electrical workers, injuries that affect the body system result in fatalities at a greater rate than even injuries to head. Investigating accident scenarios that lead to such body-part specific injuries (e.g., electrocutions that affect the body system or head concussions) should be prioritized in future studies. Similar conclusions, with a lower level of certainty, can be made about other factors such as source of injury and event type. Thus, knowing that a factor (e.g., a specific source of injury) might lead to a fatality more often than another provides empirical evidence for planning decisions that would impact safety of workers on construction sites.

3.6.3 Data Mining Analysis

To illustrate the possibility of forecasting the nature of injury of electrical contractors, a data mining technique (CART) was applied in this study. The algorithm was used to: classify the accident reports into categories of the response variable (nature of injury); gain insight into the relationship between some electrical project features (explanatory or predictor variables); and ascertaining their level of importance in terms of predicting the nature of injury. As earlier mentioned, the model presented in Figure 3.1 displays the relationship between some features of the project in the form of a decision tree. These relationships were defined in the form of decision rules and presented in Table 3.6. These rules/relationships could help safety managers in carrying-out risk assessment on jobsites. For instance, an example of the practical interpretation of a given rule is the decision rule derived with leaf node 7 which suggests that if the source of injury is not parts and materials, then the nature of injury is non-electrical. This could be interpreted to mean that the source of most of the electrical injuries that occur during an electrical project could be attributed to parts and materials. The proposed decision tree model emphasized the importance of parts and materials as a major source of injury by involving it in the first split condition at the root node. Additionally, according to the variable importance list in Table 3.9, parts and materials as a source of injury is highly important and was given a 49 importance score out of a total importance score of 100. In other words, about half of the importance score goes to parts and materials while the remaining half is shared by all other predictor variables. Hence, during electrical workers' safety training, the best ways of handling all parts and materials associated with electrical job sites should be emphasized as this is very essential and could dramatically reduce electrical injuries on site. As seen in Table 3.9, structures, surfaces, tools, instruments and equipment are the other importance sources of injury that could be constantly addressed during site meetings and safety trainings. Another important variable is the environmental factor involving overhead moving- or falling-object action. Hence, it is important to highlight the need to use protective coverings to safeguard site workers from important environmental factors such as *material*-handling equipment or method, and overhead moving- or falling-object action. It is also very essential to protect workers from occupational injuries by making sure they adhere to all safety regulations because this would help prevent accidents from occurring. One can expect an electrical injury for about 40% of the time and a non-electrical injury for about 60% of the time in occupational hazards involving electrical projects. The remaining decision rules are all clearly stated as seen in Table 3.6. These rules give insight into the associations between project information (such as end-use, project type, project cost, source of injury, environmental factor, human factor, and cause of injury) that could help predict and prevent the nature of injury of electrical contractors. The significant amount of safety data being collected on construction sites—e.g., as accident reports—provides a valuable source of information for researchers seeking to better understand the root-causes of accidents. In the process of this ongoing updates of research work on construction safety, SARMAD research group developed a user-friendly mobile application for hazard identification. It is a software that is installed on the cell phones of workers and aimed at providing accurate and handy details regarding potential hazardous incidences. It uses empirical data obtained from safety managers on construction sites. The practical results of this research work can be reviewed by users, and they can enter various inputs () to the program to gain insight into the potential hazards and risks that they may be exposed to on their specific job site and work environment.

3.7 Conclusions

Electrical contractors working in the construction industry are exposed to various hazardous situations leading to high numbers of severe injuries and fatalities (Gholizadeh and Esmaeili 2020 a, b). Electrical contractors have experienced a rise in occupational fatalities in recent years. Identifying statistically significant dependencies between these catastrophic outcomes and a handful of well-defined contributing factors in construction accidents offers a first step in mitigating the risks of construction accidents in this trade. Despite its importance, little has been understood regarding the contributing factors to occupational accident occurrence for small electrical contracting enterprises. To address this knowledge gap, the main objective of the present work is to study the individual effect of different contributing factors (e.g., project characteristics, sources, and causes of injury) on the degree of an injury. Our findings reveal that six factors have significant effects on fatality rates, with nature of injury and injured part of the body having the highest association and project type, source of injury, cause of injury, and event type with moderate impacts. The results of this research work are in line with previous studies and explained the association between electrical project features using accident reports from OSHA's IMIS accidents database from 2007 to 2013 and proposed a model for predicting the nature of occupational injury of electrical contractors. The results of this study apply to the construction industry of the United States of America. The data mining technique known as CART (using decision trees) was employed in this research work to determine if the nature of injury (electrical or non-electrical) of an electrical contractor can be predicted from some project details such as end-use, project type, project cost, source of injury, environmental factor, human factor, and cause of injury. The results of this study, as depicted in the proposed decision tree model gave insight into: (1) the statistics of accident reports affecting electrical contractors; (2) forecasting the nature of injuries due to an accident during an electrical project; and (3) identifying the factors that are most important in forecasting the nature of electrical project injuries. The model proposed by this study revealed that the most important factor for predicting the nature of injury (electrical or non-electrical) is: whether the source of injury is *parts and materials*; followed by whether the source of injury is *tools, instruments, and equipment*. In other words, in predicting (with a 94.31% accuracy) the nature of injury as electrical or non-electrical, whether the source of injury is *parts and materials* and whether the source of injury is *tools, instruments, and equipment* are very important. Seven decision rules were derived from the proposed decision tree model. Safety managers can benefit from these findings to better allocate their limited safety resources and develop personalized interventions (e.g., trainings) to mitigate safety risks. One can conclude that protecting body systems from electric shocks and protecting the head from concussions could be effective ways to reduce occupational fatality rates among electrical workers.

This work discusses several contributing factors to analyze accidents that occur to electrical contractors and provides insight for current and future consideration. In particular, future studies can incorporate other important general variables, such as age and sex of the employee, time of the accident, more specific information on the specific type of accidents (e.g., height for fall accidents, and voltage for exposure-to-electricity cases). The severity of accidents also can be defined more accurately by considering more variables such as monetary cost of injuries or days away from work for non-fatal incidents. Future studies also can include more recent incidents that are available on OSHA's database. Continuing research in this filed will enable safety managers to develop personalize interventions to further reduce severity and frequency of incidents among electrical contractors.

CHAPTER FOUR: DEVELOPING A MULTI-VARIATE LOGISTIC REGRESSION MODEL TO ANALYZE ACCIDENT SCENARIOS: CASE OF ELECTRICAL CONTRACTORS

4.1 Introduction

In 1931, safety theorist Herbert William Heinrich described safety problems as relating to three elements: environment (i.e., safe or unsafe states); decision space (i.e., safe or risky human acts); and the probability that an accident happens given the risky action of humans in an unsafe state (Heinrich, 1941). While Heinrich's concepts have shaped much of the industrial and occupational safety considerations of the last ninety years (Manuele 2002, Penkey and Siddiqui 2015), current fatality and injury trends within various industries demonstrate that these contributing elements have not been sufficiently controlled or considered to keep workers safe. More than 5000 workers have died while working and about 2.8 million workers were injured on jobsites in 2017. Almost one in every five fatalities has occurred in the construction industry (Bureau of Labor Statistics, 2019). These statistics further emphasize that new solutions to managing construction risk appear necessary.

Many studies have leveraged Heinrich's parameters to model occupational accidents and improve worksite safety for workers (Mitropoulos et al. 2005, Bellamy et al. 2007, Manu et al. 2012, Behm and Schneller 2013, Hola and Szóstak 2019, EUROSTAT 2013). For instance, Barkan et al.(1998) integrated signal detection theory into Heinrich's framework to design and execute experiments to study the effects of several factors (e.g., giving positive feedback for good behaviors while imposing punishment for unsafe behaviors) on the probability of choosing a risky task in an unsafe state (i.e., miss rates). Furthermore, to tailor Heinrich's ideas to construction accident investigations, Abdelhamid and Everett (2000) proposed an accident root causes tracing model (ARCTM) which emphasizes unsafe environment/acts as the main root cause of accidents; ARCTM's application is intended to identify accident causes to prevent accident recurrences. At their root, these applications of Heinrich's theory demonstrate the desire to understand the circumstances under which workers—especially construction workers—are more prone to fatalities in occupational accidents, in order to improve safety performance.

However, defining the circumstances of events (i.e., unsafe conditions/acts) can be challenging. Abdelhamid and Everett (2000) suggested that any physical layout, status of materials, tools, and so on, that are in violation of safety standards are a type of unsafe condition. Another study defines an unsafe act as "a violation of an accepted safe procedure which could permit the occurrence of an accident" (Hamid et al. 2008, p. 5). In the dynamic and diverse domain of construction, such specifications are too general to be useful to safety practitioners. Furthermore, these generically defined factors do not exist in a vacuum, so the interactions between environment, behavior, and probability make identifying and proactively managing risk a compounding puzzle for practitioners.

To address this issue, this study builds a multivariate statistical model to identify which elements combine to contribute to more serious accident scenarios (i.e., combinations of project's characteristics, worker's tasks, and accident factors) among those performing a specific trade in the construction industry. To select a trade for analysis, the authors evaluated the specialty trade contractors' classifications within the North American Industry Classification System (NAICS) and found that electrical contractors faced the largest increase in the number of fatalities in recent years (Office of Management and Budget, 2017). In addition, electrical contractors face a wide range of accident types (e.g., such as electrocution, fall, struck-by, and caught in/between) that represent multiple accident scenarios. The variety of accident scenarios enabled the research team to better measure the impact of each individual accident factor. Furthermore, considering the large number of registered firms as electrical contractors and the large number of workers employed by these specialty contractors (Bureau of Labor Statistics, 2019), the research team decided to consider electrical contractors as a prime candidate for the analysis in this study.

In this study, we used historical data about past accidents among electrical contractors to provide a quantitative representation of the unsafe conditions/acts that existed on the construction sites, with the severity of injuries serving as the metric to evaluate the safety performance. Then, using a thorough logistic regression modeling framework, the research team iteratively analyzed the impact of individual accident factors—as described in the accident reports—on the accidents' outcomes; here, the severity of injury among the electrical contractors. As a logistic regression model determines which factors in a pattern have a significant effect on the dependent variable, this approach allowed our team to determine which accident factors contribute significantly to the injury pattern. Given that an accident pattern captures factors describing both unsafe environments and unsafe acts on a construction site, by identifying significant contributory factors affecting accident severity, our model detects high-priority factors that may be preemptively mitigated to prevent or ameliorate accidents. Such knowledge of accident patterns helps safety managers prioritize addressing these conditions in the safety planning and resource-assignment phase of future projects, and therefore supports worker safety while saving safety practitioners time and resources.

The rest of the paper is organized as follows: (1) the background section discusses findings of a literature review regarding the causes and circumstances of occupational accidents among electrical contractors; (2) the methodology section details the logistic regression modeling process, where data analysis and modeling were done using the R language [15]; (3) the results of each modeling process step are demonstrated for the case of electrical contractors, and the hazardous accident patterns observed in this trade are outlined; and lastly, (4) the important findings of the paper are discussed and summarized.

4.2 Background

4.2.1 Analyzing Accident Severity against Contributing Factors

Investigating incident severity through statistical modeling and data-mining techniques has been a popular approach among accident-prevention studies (Chang and Chen 2005; Delen et al. 2006; Huang et al. 2008; Savolainen and Ghosh 2008). A data mining method, association rules, have been used by Liao and Perng (2008), and Cheng et al. (2010) to analyze construction accidents to discover potential associations among accident factors. Liao and Perng (2008) found that in civil engineering projects worker’s age (i.e., between 45–54) and time of service (more than 365 days) are most associated with higher probabilities of fatal injuries. In building construction projects, however, the study found that the worker's salary (i.e., more skilled workers) and day of the week (Mondays and Tuesdays) are more associated with higher fatalities. Cheng et al. (2010) applied association rules to investigate the main factors that are associated with falls in civil engineering and building construction projects. They revealed that "failure to install a work platform or protection when working in a high place, workers' horizontal movements, and failure to use personal protection equipment when at work" are the top three factors contributing to falls in both types of projects (p. 443).

Another popular statistical modeling method to explain relationships between accident variables and the outcome of the incident is logistic regression (Sze and Wong 2007; Tay et al. 2008; Daniels et al. 2010). Many accident analyses studies have utilized variations of logistic regression models. For instance, Al-Ghamdi (2002) applied logistic regression on police reports to examine the effects of several variables on the severity of injuries and revealed that location and cause of accident are the most significant factors. In another study, Yan et al. (2005) analyzed the probability of rear-end accidents using Florida traffic accident data to identify the significance of risk factors such as road environment features (e.g., number of lanes, road surface condition) and driver/vehicle traits (e.g., drivers age, vehicle type) on this type of accident. In a study more relevant to construction safety, Harb et al. (2008) investigated three years of work-zone crash data from Florida using multiple logistic regression models; the authors found roadway geometry, age and gender of the driver, usage of drugs and alcohol, lighting and weather conditions can all be major risk factors of work-zone crashes. Chau et al. (2002, 2004) reviewed 880 cases in the French construction industry to study the effects of the individual characteristics of workers on accident occurrence. The results indicated that sleep disorders and young age significantly contribute to higher injury rates.

Savolainen et al. (2011) performed a comprehensive review of studies related to the application of logistic models in accident analysis; their work shows that among various modeling frameworks, discrete outcome models such as ordered probit and unordered (i.e., nominal) logit models have been applied the most. The choice between these models mainly depends on the nature of the response variable (i.e., degree on injury). When the accidents are classified with either fatal or nonfatal outcomes, binary outcome models become a viable choice for analysis. For instance, in an attempt to quantify the effects of four different street patterns on road accident injury risks—controlling for other parameters such as driver's age, condition, and so on—Riffat and Tay (2009) analyzed 22,704 crashes using a binary outcome logistic regression model; the study showed that some patterns such as lollipops and loops can be marginally safer when compared to the traditional gridiron pattern. Another study by Peek-Asa et al. (2010) investigated the impact of teenager drivers' age on the degree of injury on rural and urban roads. This study found that among teenage drivers the odds of a fatal/severe injury were almost five times higher in non-urban areas than urban environments. The authors concluded that the higher fatality rates in rural areas could be associated with road conditions, uncontrolled intersections, narrower two-lane roads and less visibility. Each of these types of accident-severity analyses attempts to detect accident factors with significant impacts on accident severity, and they can all lead to a better understanding of the latent causes of accidents, as well as to the design of better safety solutions.

While the occupational safety literature is rich in the field of construction, severity-analysis studies based on empirical data are very limited compared to transportation accident studies. The lower number of accidents and fewer data sources, as well as the unique conditions of construction sites—as compared to the relatively ubiquitous conditions in traffic crashes—reasonably explains this gap in the literature. Furthermore, the majority of safety studies/programs in construction have focused on developing methods to reduce the frequency of accidents and not the severity of their outcomes (Moudon et al., 2011). This study addresses these knowledge gaps by analyzing the severity of construction accidents through a multi-variate model. Another limitation, even among crash-injury analyses, is that many details of the modeling process are usually missing from these studies. Thus, discussing the results of these models without describing the details can hurt the final conclusions, as logistic models can be very sensitive to their assumptions. Therefore, the authors provided a detailed description of modeling steps that can be replicated in future safety studies.

4.2.2 Occupational Incidents among Electrical Contractors

Construction is one of the largest industries in the United States. According to the Bureau of Labor Statistics, more than nine million people worked in this industry in 2018 (DataUSA, 2019). The construction industry also consists of a wide range of professions and specialties, and therefore a large number of unique environments around each project. To limit the scope of analysis, the authors decided to focus on only one specialty trade, electrical contractors.

According to the Statistics of U.S. Businesses (SUBS), in 2016 more than 144,000 establishments and firms were registered as electrical contractors, making this subindustry the second largest in business count among all specialty trades (United States Census Bureau 2016). In 2017, this trade also had the third highest number of fatalities among all 18 specialty trades within the industry (Bureau of Labor Statistics, 2019). To better understand accident patterns among electrical contractors, the authors have conducted an in-depth literature review related to occupational incidents in this trade.

Studies investigating the nature of accidents among electrical contractors are minimal (Gholizadeh and Esmaeili, 2016). In one early study, an examination of the mortality patterns of 31,068 U.S. members of the International Brotherhood of Electrical Workers (IBEW) working in the construction industry between 1982 to 1987 showed an elevated proportion of mortality due to causes such as leukemia, brain tumors, melanoma skin cancer, and diseases caused by asbestos (e.g., lung cancer) (Robinson et al. 1999). However, their study reviewed mostly long-term causes of injuries—such as cancer and heart disease—and therefore did not cover day-to-day hazards and occupational injuries on construction sites.

Rossignol and Pineault (1994) investigated fifty-seven electrocution accident reports that resulted in a fatality in Quebec during the period of 1981–1988. The authors defined three phases as pre-electrocution, electrocution, and post-electrocution, and they then set out seven descriptors for the first two phases. The authors used factor analysis to classify the seven pre-defined descriptors. Results of the factor analysis showed more than 90% of cases were covered with only two factors: (1) twenty-six electrocutions occurred during an indoor electrical task performed in direct contact with a source of less than 10,000 volts; and (2) twenty cases happened while performing non-electrical outdoor tasks in indirect contact with a source of more than 10,000 volts. This clear distinction raised questions about the efficiency of safety trainings where around 43% of accidents occurred in non-electrical tasks. The authors suggested a shift in safety strategies from training to elimination or modification of hazard sources. While the outcomes were interesting, the study suffered limitations from a relatively small and old database of cases from outside the United States.

In a study that considered age, company size, experience, tasks performed, source of injury, and accident causes as dependent variables, Chi et al. (2012) performed a chi-square automatic interaction detector (CHAID) analysis of 250 fatal electrocution accidents to classify accidents into seven hazard-pattern scenarios. Among all these predictors, only source of injury and accident causes were indicated to be significant in classifying accident scenarios. Results of the study show that if the source of injury was energized equipment, then almost all electrocutions occurred due to direct contact; for other electrical sources, direct or indirect contact would happen relatively equally. As the authors indicated, the main limitations of this study were inconsistency in reporting accidents and using only two predictors in the analysis, along with the relatively small number of observations.

Considering the high proportion of fatalities due to electrocution in construction even after safety controls have been employed for decades, Zhao et al. (2015 a) examined 486 recommended controls from 132 Fatality Assessment and Control Evaluation (FACE) accident reports to: (1) evaluate the effectiveness of electrical safety recommendations given by NIOSH experts; and (2) assess safety knowledge in the construction industry. Accident reports were coded using the hierarchy of controls: elimination, substitution, engineering, administration, and personal protective equipment. The authors also defined a variable to represent safety knowledge based on the number of recommendations for each accident. They compared these two variables by three important parameters for electrical safety—construction type (residential, commercial, etc.); occupation (electrician, and non-electrician); and electrical condition (low voltage and high voltage)—and found that the number of suggested safety practices was slightly higher for non-electrical workers, and control measures for high-voltage hazards, were statistically less effective than those for low-voltage hazards. The effectiveness of controls was not statistically different by construction type or occupation. Another interesting finding was that behavioral controls in electrical hazard mitigation are overemphasized and more attention should be paid to effective control measures, such as elimination. While the paper made several valuable contributions, it suffered the limitation that FACE reports were collected in multiple decades and were not randomly selected. Therefore, the assumption that the recommendations provided in those reports are common practices in the construction industry would not be correct.

To better understand the chain of decision mistakes that lead to an electrical accident, Zhao et al. (2016) examined 144 FACE accident reports. Considering 12 common decision mistakes (e.g., failure to lockout/tagout) as independent variables and 19 pre-defined factors (e.g., project type, safety training, etc.) as predictors, the authors classified accident reports using Exhaustive CHAID, a type of classification tree. The analysis detected five features of work (i.e., group of activities and sequences) and decisions that could prevent electrocution accidents. One main finding of the study was that the sooner a decision is made, the better it can mitigate the risk of electrical accidents. The results of the study can help safety managers find critical decision points in different situations in order to suggest effective safety controls before the construction phase. Using the same database, Zhao et al. (2015b) studied electrical injuries by considering the interactions among humans, technology, social structure, and environment as a sociotechnical system. By conducting latent class analysis and multiple correspondence analysis, the authors identified three sociotechnical systems in the construction industry: residential-building; heavy and civil-engineering; and non-residential-building construction. The authors provided specific recommendations for each system to reduce the risk of electrocution injuries. However, both of these studies were conducted on a small number of fatal accident reports; for this reason, a severity analysis could not be done on such data.

In another study, Gholizadeh and Esmaeili (2020) studied the effects of different accident types and project end uses on the cost of injury among electrical contractors using robust hypothesis testing methods (i.e., Welch-type procedure, extension of Yuen's method, and percentile bootstrapping). The results of the study confirmed that robust hypothesis testing approaches can be successfully implemented on safety data even when the assumptions of conventional test statistics are violated. The findings showed that various event types and project end-uses can impact the cost of injuries among electrical contractors: caught in/between and exposure to electricity accidents can, on average, lead to higher injury costs than fall to lower levels and struck-by objects/equipment accidents. For a project's end-use, the outcomes indicated that injuries that occurred in nonbuilding projects are significantly costlier than those that occurred in building projects.

The results of the literature review indicated several limitations in previous studies on the occupational health and safety of electrical contractors: (1) the majority of studies have focused only on fatalities and ignored other incident outcomes; (2) most of the studies investigated electrical incidents and not all types of possible hazards (e.g., fall, struck by) among electrical contractors; and (3) there is an absence of robust statistical analysis to support inferences made from incident databases—severity-analysis studies (i.e., using statistical models to explain and predict the outcome of an accident) are limited in the construction safety domain. On the other hand, researchers have been able to enhance our understanding of traffic accidents and their consequences by applying logistic regression models on traffic crash accident data. The authors believe that construction safety managers can also benefit from such analysis by identifying more severe accident scenarios. While the modeling method used in this study has been successfully applied to traffic accident data, its application to construction accident data was not guaranteed before this study. Our findings show that empirical data available publicly through organizations such as the Occupational Safety and Health Administration (OSHA), while they may need improvements, are suited for such analysis, and that statistical models can be built upon construction accident data to find more severe patterns and prioritize managing them. Therefore, this study aims to address these limitations by using a multivariate logistic model to analyze accidents in the OSHA's Integrated Management Information System (IMIS) accident database affecting electrical contractors.

4.3 Materials and Methods

To attain the research objectives, the research team first acquired reliable national data on occupational incidents involving electrical contractors from OSHA's online database of catastrophic accidents. The authors then conducted a thorough content analysis to ensure consistency among variables, to reduce any ambiguity in reported values, and to prepare data for statistical analysis. As the most severe accidents end in a fatality, this study uses fatality rates to describe accident severity. Thus, to investigate and explain the relationship between factors contributing to accidents and the degree of accident injuries, this study executed a multivariate logistic regression model that estimates the fatality rates of different accident scenarios occurring among electrical contractors. The ability to consider several factors in one model and interpret the final coefficients in terms of adjusted odds ratios (i.e., controlled for other factors) makes multivariate logistic modeling a suitable approach for severity analysis. The rest of this section is devoted to explaining each of these steps.

4.3.1 Incident Database and Programming Language

The authors collected 621 accident reports involving electrical contractors between 2007 to 2013 from OSHA's IMIS online database. In total, 689 employees were injured in these incidents while performing their jobs on construction projects. The OSHA IMIS database has been successfully used by previous researchers to study occupational incidents (Esmaeili 2012; Esmaeili et al. 2015a, 2015b). Within this database, a summary of each accident, as reported by OSHA inspectors, appears along with variables that describe the accident (e.g., event type, source and cause of injury), its context (e.g., project end-use, type, and cost), and its consequences (e.g., nature and degree of injuries, and injured part of body). As this study's concern is contributing factors, the analysis only considers variables that manifest before an accident occurs; therefore, variables such as nature of injury (e.g., fracture, burn), part of body (e.g., head, trunk), and event type (e.g., fall, struck by, exposure to electricity) are excluded from the modeling process in this study because these elements are all characteristics of the accident after the accident occurred and therefore do not represent contributory factors to accident severity.

The data analysis and model development steps were done in R environments (R Core Team 2013). R is an open-source language for statistical computing and graphics which provides a broad range of statistical techniques such as classical statistical tests, classification, clustering, time series analysis, and linear and non-linear modeling. R environments can run on Windows, MacOS and Unix platforms.

4.3.2 Odds Ratio

To better understand the associations between each significant covariate and the target variable (i.e., degree of injury) and to interpret the likelihood of a fatal accident in different situations, the unadjusted odds ratios of fatal accidents were calculated at this point. Odds ratios can compare the magnitude of different risk factors over an outcome (Szumilas 2010). For example, consider a comparison of the odds of a fatal injury among all accidents where the source of injury is either a vehicle or a machinery. The odds ratio for the effect of the vehicle category in this example can show the magnitude of its effect on the fatal injuries; an odds ratio of 4 means that the chance (i.e., odds) of a fatal injury is four times more where the source is a vehicle compared to cases where the source is a machine. These ratios can be compared to the adjusted ones once the final model is selected, as such a comparison reveals the effects of controlling factors on odds ratios. Significant changes among unadjusted and adjusted ratios would suggest high correlations between accident factors, and further encourages the use of logistic models to achieve adjusted ratios.

4.3.3 Logistic Regression

While using chi-square tests can show the relationship between a single factor and the degree of injuries, these tests are unable to determine the effects of these variables in the presence of other factors. Logistic regression is a proper method to test the association between potential accident risk factors and a dependent variable (Harb et al. 2008). As mentioned in the background section, many studies have used various types of logistic regression models to investigate potential associations in the field of accident analysis. While more advanced machine learning (ML) methods have emerged and been applied to accident studies in the past decade, logistic regression has some advantages over ML methods to justify its application in this study. First, the results of a regression model are easier to interpret. The model can be represented in one formula using only the independent variables and their coefficients. The coefficients of the model can directly determine important variables, along with the magnitude and direction of association between each independent variable (i.e., risk factor) and the dependent variable. This property is very important in studies where finding the relationship between risk factors and the dependent variable is as significant as the accuracy of the model's predictions. This could be the main reason for the popularity of regression modeling in traffic accident analysis studies. Therefore, while some machine learning methods such as decision trees—and their variants such as random forests and gradient boosting trees—and support vector machines might provide better prediction accuracy, their lack of interpretability can be a disadvantage. Second, once the risk factors are identified, developing a logistic regression model is straightforward and, unlike machine learning methods, does not require tuning various hyperparameters. This quality makes logistic regression the first choice in many predictive studies and a valid baseline for other more complicated classifiers. Third, while methods such as association rules can be useful to find latent patterns in large data sets, these methods are inherently different from classification methods such as logistic regression modeling. As Freitas (2000) has outlined, classification methods are about using the past data to predict the future; prediction is a non-deterministic task, and that's why two different classification methods (e.g., logistic regression and decision trees) could generate different predictions on the same set of values. On the other hand, association rules are deterministic: every algorithm would produce the same set of rules, while some might be faster. One of the objectives behind modeling construction accidents in this paper is to use a reliable and well-defined model to predict fatality rates in common accident scenarios; association rules cannot be used for such predictions.

A logistic regression model can isolate effects and indicate which variables can explain the variability among accidents more accurately. Logistic regression models have been adopted widely in areas ranging from medicine (Higgins et al. 1992; Narayan et al. 2003) to the social sciences (Pierson et al. 1983; Lattimore and Visher 2014). This section details several steps in developing and evaluating a multivariate logistic regression model.

Traditionally, building statistical models starts with selecting variables that can result in a parsimonious model (i.e., having as few variables as possible), explain the data, are stable, and can be generalized to unseen situations (Bursac et al. 2008). To develop such models, this study adopted the purposeful selection of variables procedure proposed by Hosmer et al. (2013). One main advantage of this approach is that it considers both the significance and change-in-estimate criteria when selecting final variables (Dunkler et al. 2014). Each step will be explained here. One should note that to maintain the language of statistical modeling, accident factors are called “covariates” in this section. Three steps were taken to conduct this regression analysis (Figure 4.1):

- Step 1: Select variables and calculate odds ratios

- Step 2: Develop and adjust model
- Step 3: Assess and validate model

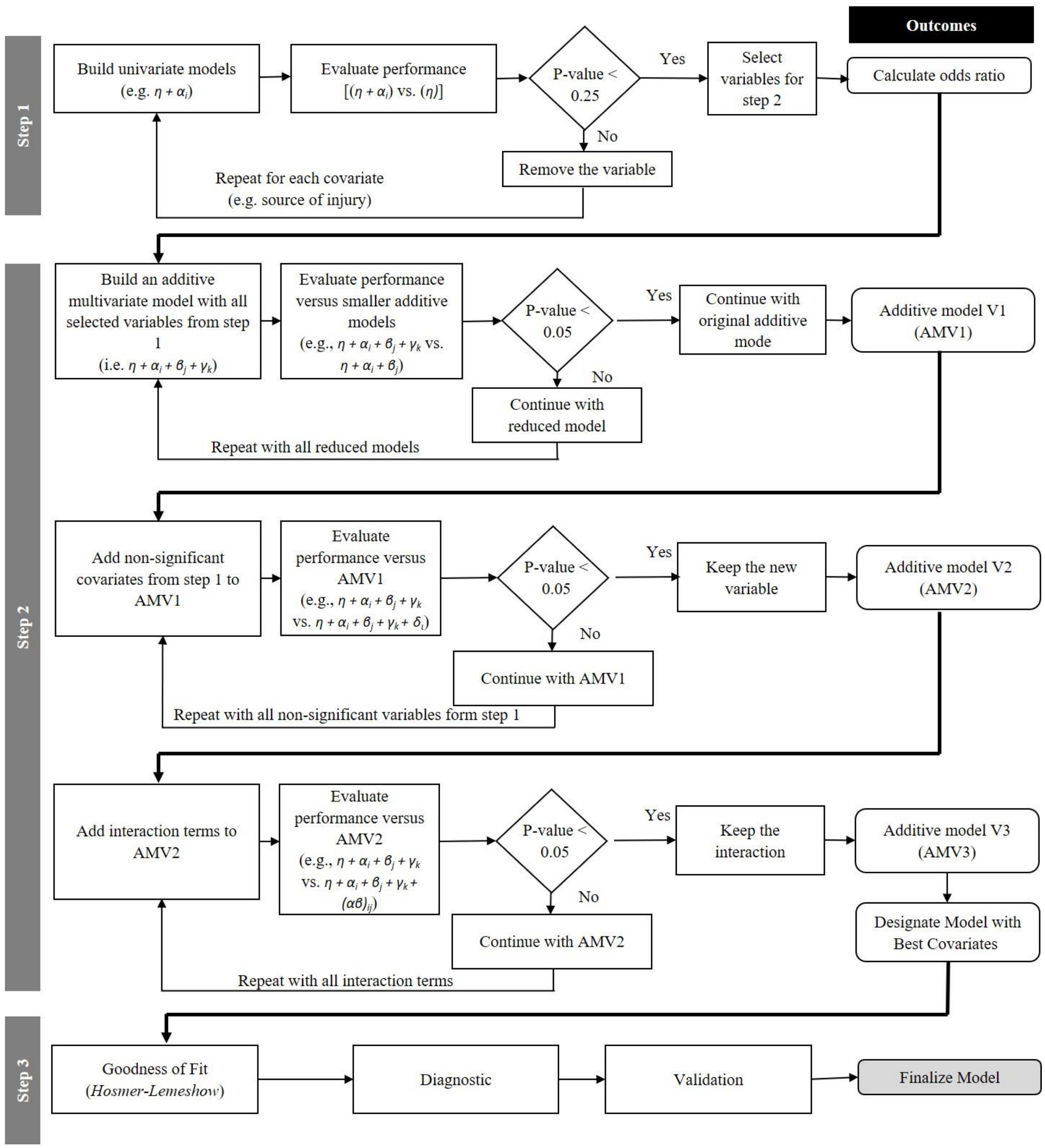


Figure 4.1. Three steps of building and validating a multivariate logistic regression model.

Step 1: Select variables and calculate odds ratios

A purposeful model-building process starts with building univariate (i.e., containing only one covariate such as project end-use or source of injury) models for each covariate and assessing their performance. The performance of each univariate model is calculated as the difference between its deviance and the deviance of a model with only the constant parameter (i.e., no covariate). The glm function from the stats package in R (R Core Team 2013) was used to determine the deviances of models. The significance of difference (i.e., presented by G) is determined through the p-value of a chi-square test (i.e., the pchisq() function in R) with a degree of freedom equal to (d-1), where 'd' is the number of categories of the covariate. Within this framework, a significant result recommends the inclusion of the variable in the final model. As this is the first step, Hosmer et al. (2013) recommended less conservative significance levels (i.e., 0.25 instead of 0.05) to include more variables in the model. In other words, this step allows less significant factors to remain in the model in order to analyze the effects of these apparently less significant factors on more significant ones in future steps. Factors that are neither significant at the 0.05 level nor have an effect on other factors would be excluded from the model eventually.

Table 4.1 presents the log likelihood ratio statistics (i.e., G) for five univariate models. Low values of G indicate that the difference between fatality rates among different categories of a factor are negligible and therefore the variable would not be very predictive in the model. The significance of G also depends on the degree of freedom (i.e., d.f.) which ultimately determines the p-values (i.e., the evidence against a null hypothesis: the smaller the p-value, the stronger the evidence to reject a null hypothesis). One can only check p-values and conclude whether the effect of a variable on fatality rates is significant or not. As the significant level in this step is 0.25, any value less than that is considered as significant at this step. Based on p-values of the chi-square tests, "end use" and "project cost" do not have significant effects on the probability of a fatal accident (even at the significance level of 0.25); in other words, the univariate models with these variables do not differ significantly from a model that has no covariates. The results indicate that the other three covariates are significant and should be considered in the multivariate model.

Table 4.1. Fitting univariate logistic models.

Model	G	Degree of freedom (d.f.)	*p*-Value
End use (EU)	0.4	4	0.983
Project type (PT)	18.7	4	0.001 *
Project cost (PC)	6.36	6	0.384
Source of injury (SoI)	21.58	5	0.001 *
Cause of injury (CoI)	11.80	5	0.038 *

* Statistically significant variables at 0.25.

Step 2: Develop and adjust model

After finding the more important covariates (i.e., "source of injury," "cause of injury," "project type"), one can build an additive (i.e., no interaction) model and test the importance of individual covariates in this multivariate context using traditional significance levels (i.e., 0.05 in this study). The non-significant covariates are temporarily removed from the model—in the case of categorical covariates, all levels should be removed even if only one of the categories is not significant. Next, the deviance of the new/reduced model is compared to the deviance of the original multivariate model (i.e., likelihood ratio test). A large difference means that the removed variables—though independently not significant—have a considerable effect on adjusting the significant variables, and hence should be added back to the model. This process can be repeated several times to make sure that the necessary variables are included in the model. The last step in building a multivariate model includes the interaction effects. As with the single covariates, the interaction terms are added to the model one-by-one, and their effects are measured through the amount of deviance they can reduce. Thus, using this approach, significant interactions also remain in the final model. The modeling step is accomplished using, mainly, the generalized linear modeling [i.e., glm()] function with 'binomial' link in R (R Core Team 2013).

As there are only eighteen possible models of interest based on the different combinations of these variables and their interactions, all of them are shown in Table 4.2. Three subscripts were used to reflect the structure of the data: let πijk be the probability of a fatal accident in the (i, j, k)-th group, where i = 1, 2, 3, 4, 5, 6 indexes "source of injury," j = 1, 2, 3, 4, 5 presents levels of "project type" and k = 1, 2, 3, 4, 5, 6 indicates the categories under "cause of injuries". These variables can produce 180 accident patterns (i.e., 6 × 6 × 5). However, 72 of these patterns have zero cases in the data and therefore should be excluded from analysis, leaving 108 covariate patterns for modeling. As all predictors are categorical, the authors decided to focus on fatality rates among these patterns instead of looking at each accident individually. Table 4.2 shows models in abbreviated notation, formulas for the linear predictor, the deviance, and the degrees of freedom. A unique ID number has also been assigned to each model for future references.

Table 4.2. Deviance for logit models of fatality by "source of injury," "project group," and "Cause of Injury".

ID	Model	logit (π_{ijk})	Deviance	d.f.
1	Null	η	156.87	107
	One Factor			
2	SoI	$\eta + \alpha_i$	135.29	102
3	PT	$\eta + \beta_j$	138.16	103
4	CoI	$\eta + \gamma_k$	145.07	102
	Two Factors			
5	SoI + PT	$\eta + \alpha_i + \beta_j$	113.59	98
6	SoI + CoI	$\eta + \alpha_i + \gamma_k$	123.40	97
7	PT + CoI	$\eta + \beta_j + \gamma_k$	125.49	98
8	SoI × PT	$\eta + \alpha_i + \beta_j + (\alpha\beta)_{ij}$	91.65	80
9	SoI × CoI	$\eta + \alpha_i + \gamma_k + (\alpha\gamma)_{ik}$	98.64	72
10	PT × CoI	$\eta + \beta_j + \gamma_k + (\beta\gamma)_{jk}$	100.98	82
	Three Factors			
11	SoI + PT + CoI	$\eta + \alpha_i + \beta_j + \gamma_k$	101.20	93
12	SoI × PT + CoI	$\eta + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$	78.21	75
13	SoI × CoI + PT	$\eta + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik}$	75.10	68
14	SoI + PT × CoI	$\eta + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk}$	77.79	77
15	SoI × PT + SoI×CoI	$\eta + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik}$	55.17	50
16	SoI × PT + PT × CoI	$\eta + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk}$	54.38	59
17	SoI × CoI + PT × CoI	$\eta + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$	56.75	52
18	SoI × PT + SoI × CoI + PT × CoI	$\eta + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$	37.55	34

Using the deviance and degree-of-freedom from Table 4.2, one should start with the additive model with three covariates (i.e., model 11). Next, one covariate would be excluded from the model to check its effect on the other two covariates. The results of these tests are presented in Table 4.3 (i.e., a1, a2, and a3). The results reveal that the additive model with three variables (i.e., SoI + PT + CoI) represents a significant improvement over all the additive models with two factors (i.e., model 5, 6, and 7). In other words, while some levels of covariates are not significant (results are not shown), they have a significant effect on adjusting the other variables and therefore should remain in the model.

The next step is to investigate the interaction effects, which includes models with one, two, or three interaction terms. Three models (i.e., 12, 13, 14) in Table 4.2 have one interaction term. For example, model 12 includes the main effects of source of injury (SoI), project type (PT), and cause of injury (CoI), and the interaction between SoI and PT. The log likelihood tests in Table 4.3 (i.e., b1, b2, and b3) show that none of these models with one interaction is better than the three-factor additive model (i.e., model 11). One can also consider models involving two interactions between two factors, of which there are three (i.e., models 15, 16, 17). The results show that only one model (i.e., model 16) is marginally (p-value: 0.071) better than the additive model. Model 18, which includes all the interactions between each two variables, could not improve the additive model, and hence cannot be selected as a good model. Considering the p-values in Table 4.3, one can conclude that model 11 (SoI + PT + CoI) is the best model.

Table 4.3. Model comparisons.

Test ID	Test (Model A vs. Model B)	G	d.f.	*p*-Value
a1	11 vs. 5	12.39	5	0.030
a2	11 vs. 6	22.20	4	0.000
a3	11 vs. 7	24.29	5	0.000
b1	12 vs. 11	22.99	18	0.191
b2	13 vs. 11	26.1	25	0.402
b3	14 vs. 11	23.41	16	0.103
c1	15 vs. 11	46.03	43	0.348
c2	16 vs. 11	46.82	34	0.071
c3	17 vs. 11	44.45	41	0.329
d1	18 vs. 11	63.65	59	0.316

The last task in model development is to consider non-significant variables from step 1 in the context of the new multivariate model and check whether they can improve the performance of this model. Table 4.4 shows the results of such comparisons and declares that adding the non-significant variables, one-by-one and together, cannot lead to better results. For instance, adding end use to the model would reduce the model by four degrees of freedom while only reducing the deviance by 0.73, which is not even close to a significant improvement (p-value: 0.949). Therefore, the research team concludes that the additive model with three variables is the best multivariate model among the possible options.

Table 4.4. Additive models with and without non-significant variables.

Model	G	d.f.	*p*-Value
PT + SoI + CoI + EU	0.73	4	0.949
PT + SoI + CoI + PC	4.73	6	0.579
PT + SoI + CoI + EU + PC	5.51	10	0.855

Step 3: Assess and validate model

After building a model by selecting covariates and tuning the model in a purposeful manner, one should investigate the probabilities that are produced by the model against true values in data. To do so, the last step in every statistical modeling process includes model assessment. This section will detail this study's three main attempts to fulfill the requirements needed to assess our final model.

I. Goodness of fit

After building models and comparing deviances to determine covariates and build a parsimonious model, one needs to examine how well the data fits the final model. Lack of fit means that estimated coefficients are biased, odds ratios could be misleading, and future predictions are not accurate. To check fit, one can compare predicted values derived from the model to observed values to confirm that the fitted model is correct (Ma 2018). The Hosmer-Lemeshow (HL) statistic is a popular test for goodness-of-fit and has been used in several clinical studies (Higgins et al. 1992; Narayan et al. 2003; Campbell et al. 2003). The idea is to partition observations based on their estimated probabilities into g (usually 10 to represent deciles of risk) groups featuring approximately the same quantity of observations; here, the first group would represent the lowest probabilities of fatality, the next group would have larger estimated probabilities and so on, until the last group which includes the highest probabilities (Hosmer and Lemeshow 1980; Lemeshow and Hosmer 1982). The HL statistic is then calculated by comparing the sum of probabilities to the number of observed values in each group. A chi-square test on this value with g-2 degrees of freedom can determine whether there is enough evidence to reject the hypothesis that data fits the model. The ‘hoslem.test()’ function from ‘ResourceSelection’ package in R (R Core Team 2013) provides the test statistic and p-value of the HL test.

Table 4.5 shows the results of the Hosmer-Lemeshow tests when dividing the accident patterns into 10 groups. The value of Hosmer-Lemeshow’s goodness of fit computed for the frequencies in Table 4.5 is 5.85 when granted 8 degrees of freedom, with the corresponding *p*-value of 0.664. The large p-value indicates that the null hypothesis that the model fits the data cannot be rejected, demonstrating that the model fits the data quite well. A comparison of observed and expected frequencies in the 20 cells in Table 4.5 also shows close agreements in every decile of risk. For instance, within the highest risk decile (i.e., decile 10), the difference between both observed and expected values are within one point. Figure 4.2 compares the average of observed and predicted values in each decile, which again shows a close agreement in most deciles.

Table 4.5. Observed and estimated frequencies within ten deciles of risk for fatal and non-fatal accidents.

Decile	Cut Point	Non-Fatal		Fatal	
		Observed	Expected	Observed	Expected
1	[0.058, 0.186]	55	56.4	10	8.6
2	(0.186, 0.255]	54	51.3	12	14.8
3	(0.255, 0.295]	48	50.8	23	20.2
4	(0.295, 0.310]	32	32.0	14	14.0
5	(0.310, 0.364]	48	52.3	32	27.7
6	(0.364, 0.390]	35	31.1	15	18.9
7	(0.390, 0.402]	38	33.7	18	22.3
8	(0.402, 0.499]	33	33.0	29	29.0
9	(0.499, 0.561]	26	29.1	35	31.9
10	(0.561, 0.781]	24	23.3	38	38.8

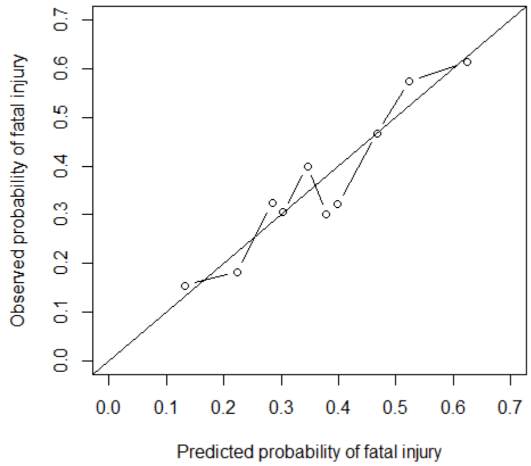


Figure 4.2. Observed and predicted probabilities of fatal injury in ten deciles of risk.

II. Diagnostics

Even a strong fit can be very sensitive to outlying and extreme-leverage points in the data (Lesaffre and Albert 1989). So, while summary statistics such as the HL test can indicate the overall fit of a model to data with a single number, one still needs to check if the model fits over all covariate patterns. Pregibon (1981) introduced a range of diagnostic measures for logistic models with binary outcomes. Two types of measures, based on the leveraged value of covariate patterns, are adopted in this study: (1) those that determine fit in each pattern (i.e., change in value of the Pearson chi-square, $\Delta\chi^2$, and change in the deviance, ΔD); and (2) those that can determine the amount of influence a pattern can have over other patterns (i.e., change in value of the estimated coefficients, $\Delta\beta$).

For each covariate pattern, values of $\Delta\chi^2$ (and ΔD) are calculated as the difference between the Pearson chi-square (and deviance) values of the original model and the model when excluding observations in that pattern. As mentioned by Peng et al. (2002), at the significance level of 0.05, and based on the critical value of the chi-square distribution with one degree of freedom (i.e., 3.84), changes more than four are considered large and demonstrate that the pattern in question contributes significantly to the disagreement between the observed and predicted values. Large values of $\Delta\beta$ also indicate that estimates are not stable. Large values of $\Delta\chi^2$ or ΔD accompanied by large changes in coefficients can signal that a covariate pattern is an outlier and should be investigated in more detail. Figure 4.3 and Table 4.6 show diagnostic measures for covariates in the final model. Regarding the poorest fit, covariate pattern 97 (i.e., SoI: "parts and material," PT: "alteration or rehabilitation," CoI: "other") and 55 (i.e., SoI: "machinery," PT: "other," CoI: "other") induced large changes in Pearson chi-square values. Covariate pattern 97 also created significant changes in deviance values. In terms of effect of a covariate pattern on coefficients of other patterns, pattern 97 again has the largest value, followed by pattern 100 (i.e., SoI: "parts and material," PT: "alteration or rehabilitation," CoI: "installing plumbing and lighting fixtures"). Figure 4.3d combines Figure 4.3a with Figure 4.3c as the size of circles represent the value of $\Delta\beta$. Based on these results, three covariate patterns (i.e., 97, 55, and 100) were selected for further investigation (Table 4.7).

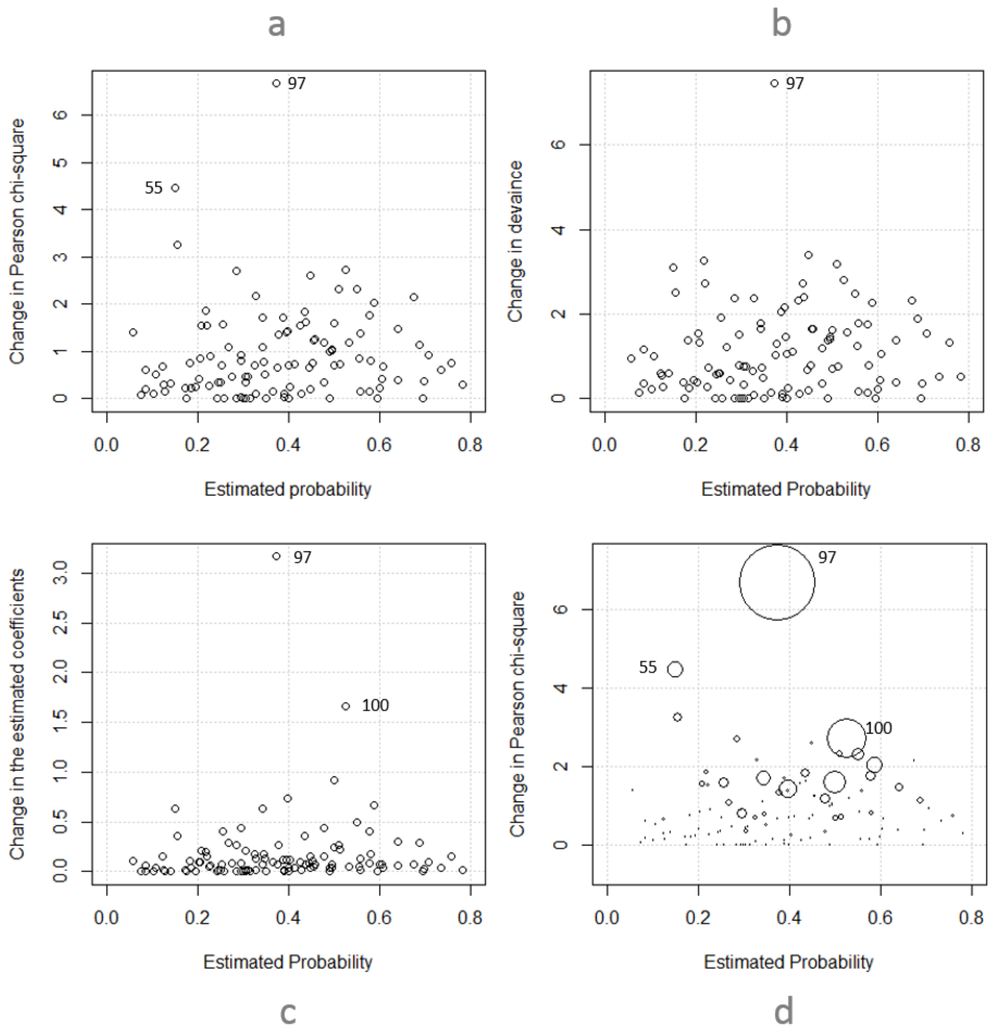


Figure 4.3. Diagnostic measures for the logistic model

(a) Change in Pearson Chi-square when removing individual covariate patterns from data, (b) Change in deviance when removing individual covariate patterns from data, (c) Change in the estimated coefficients when removing individual covariate patterns from data, (d) Combination of a and c: change in Pearson Ci-square (y-axis) and estimated coefficients (size of circles) when removing individual covariate patterns from data.

Table 4.6. Covariate values, observed outcomes (y_{ijk}), number of cases in the covariate pattern (m_{ijk}), observed probability (P_{ijk}), estimated logistic probability ($\hat{\pi}_{ijk}$), and the value of four diagnostic statistics $\Delta\chi^2$ (i.e., change in value of the Pearson chi-square), ΔD (i.e., change in the deviance), h (i.e., leverage), and Δβ (i.e., change in value of the estimated coefficients), for three covariate patterns with at least one large diagnostic value.

Covariate Pattern Number	97	55	100
Source of injury	Parts and materials	Machinery	Parts and materials
Project type	Alteration or rehabilitation	Other	Alteration or rehabilitation
Cause of injury	Other	Other	Installing plumbing, lighting fixtures
y_{ijk}	5	2	14
m_{ijk}	28	4	21
P_{ijk}	0.179	0.500	0.667
$\hat{\pi}_{ijk}$	0.373	0.149	0.524
$\Delta\chi^2$	6.667	4.464	2.737
ΔD	7.452	3.115	2.802
h	0.322	0.125	0.377
Δβ	3.166	0.639	1.658

To further validate this model, one can delete each of the three questionable patterns and look at the model statistics to see if the modified model can still perform well. Columns three, four, and five of Table 4.7 demonstrate deviance, sum of Pearson chi-square residuals, and the Hosmer-Lemeshow goodness-of-fit statistic for three models excluding covariates 97, 55, and 100, respectively. These numbers indicate that the selected additive model (i.e., SoI + PT + CoI) can still perform well and fit the data in each case. The results of three other scenarios in Table 4.7Table 4.7 (i.e., removing two covariates with the poorest fit, removing two covariates with the largest influence, and removing all three covariates) also show that the additive model with three variables can fit the data very well in each scenario. These findings indicate that the additive model performs well with the remaining 105 covariates patterns (representing 91% of all accidents). Based on these results, the three questionable covariates were removed from the rest of analysis.

Table 4.7. Model statistics for the original model and six scenarios.

Model Statistics	All Data Coefficients	97	55	100	Poorest Fit (97, 55)	Largest Influence (97, 100)	All Three
D	101.20	94.02	97.90	98.42	90.81	92.68	89.67
X^2	83.82	77.24	77.24	80.99	74.35	75.62	72.94
C	3.17	1.60	1.01	4.21	3.03	2.92	2.24

III. Validation

As mentioned by Iezzoni (2003), calibration and discrimination are two main methods to assess the performance of a logistic regression model on unseen data. While discrimination measures the ability of a model to distinguish between two classes of the dependent variable, calibration determines the model's capability in producing estimations that are, on average, close to the observed classes (Kramer and Zimmerman 2007). This study is focused on fatality rates among common accident scenarios. Due to the categorical nature of predictors, the fatality rate (i.e., dependent variable) of an accident scenario is calculated as the average of fatalities among all accidents in that scenario. For instance, among the eight accidents that have occurred in alteration/rehabilitation projects with tools and instruments representing the source of injury and interior plumbing/ducting/electrical work being the cause of injury, only one resulted in a fatality. Therefore, the observed/actual rate of fatality in this scenario is around 13%. As the objective is to study these rates and compare the performance of the logistic model to them, the authors have concluded that calibration measures would better serve this purpose.

The same measure (i.e., HL statistic) can be used for calibration/validation purposes. The only difference is that the data would be divided into training and testing sets and the model would be developed on the training set and be tested on the unseen testing set. Large p-values indicate there is no evidence to believe the model does not fit test data. For more information on the test statistics for testing dataset look at Hosmer et al. (2013) (p. 155).

To validate the model on unseen data, a series of training sets were created using 70% of the source data, and a series of testing sets were created using 30% of the source data. To increase the reliability of the results, this study applied a stratified sampling method to generate training/testing data sets. The same measure (i.e., Hosmer-Lemeshow statistic) was calculated to ensure that the model, which was trained only on the training set, can fit testing data well. The chi-square statistic of 6.97 and the *p*-value of 0.540 showed that the proposed model can fit unseen data as well.

4.4 Results

This section starts by presenting the odds ratios of fatality among three significant accident factors and continues by presenting the final model, its coefficients, and the 36 most common accident patterns among electrical contractors that can be explained by the model.

4.4.1 Odds Ratio

Before presenting the final model, one can examine unadjusted odds ratios for the three variables that are included in the model: project type, source of injury, and cause of injury. The significant ratios reveal which categories of a variable caused higher fatality rates. These ratios can also be compared to adjusted ratios (i.e., controlled for other variables) from the logistic model and to identify the effect of controlling factors on each variable. Table 4.8 demonstrates these ratios among project types. In the first row, 1.24 means that fatal accidents in "alteration or rehabilitation" projects are 1.24 times more likely to happen than in "maintenance or repair" projects. This ratio is the same as saying that fatal accidents are 0.81 times (second row) as likely to happen in "maintenance or repair" projects than in "alteration or rehabilitation" ones. As Table 4.8 shows, the odds of fatal accidents in the two largest project types (i.e., "new" and "alteration") are almost identical. Table 4.8 also shows that while the odds of an occupational death are much larger in "demolition" projects than "new," "maintenance," and "alteration" projects, the difference is not significant due to the low frequency of "demolition" projects, which, in turn, produces large confidence intervals. While the definition of the 'Other' category is ambiguous, since no further details were available from accident reports, it was not possible to translate this category among project types.

Table 4.8. Odds ratios of fatal accidents among different project types.

Project Types	Alteration or Rehabilitation	Maintenance or Repair	New Project or New Addition	Demolition	Other
Alteration or rehabilitation †	-	1.24	1.01	0.47	4.06 *
Maintenance or repair †	0.81	-	0.81	0.38	3.28 *
New project or new addition †	0.99	1.23	-	0.47	4.02 *
Demolition †	2.11	2.62	2.13	-	8.58 *
Other †	0.25 *	0.25 *	0.42 *	0.12 *	-

* Statistically significant at 0.05 level; † First term in odds ratio.

Table 4.9 demonstrates the odds ratios among “sources of injuries”. As one can see, when the source was a “tool, instrument, or equipment,” much fewer fatalities happened as compared to tasks that involved “machinery,” “parts and material,” or “vehicles.” On the contrary, “vehicles” could lead to more severe injuries than “parts and materials,” “structures and surfaces,” or “tools, instruments, and equipment.” In fact, having a fatal accident is 2, 3, and 4.4 times more likely to occur when the source is a “vehicle” than when it is a “part and material,” a part of “structure or surface,” or a “tool, instrument, or equipment,” respectively. Among sources of injuries, 56 cases were defined as “other” in OSHA’s database. After reviewing accident reports, all 56 cases were reclassified with meaningful values (vehicles, machinery, etc.). However, there were seven categories (e.g., chemical products, natural gas, containers, etc.) with very low frequencies (less than 2% of sources). To reduce the number of groups in the analysis, the authors combined these categories (accounting for 23 cases all together) into one “other” category. One should note that this “other” category is less ambiguous after content analysis as its components are known.

Table 4.9. Odds ratios of fatal accidents among different sources of injuries.

Source of Injury	Machinery	Parts and Material	Structures and Surfaces	Tools and Instruments	Vehicles	Other
Machinery	-	1.26	1.79	2.57 *	0.58	1.50
Parts and Material	0.79	-	1.42	2.04 *	0.46 *	1.19
Structures and Surfaces	0.56	0.71	-	1.44	0.32 *	0.84
Tools and instruments	0.39 *	0.49 *	0.70	-	0.23 *	0.58
Vehicles	1.73	2.18 *	3.08 *	4.44 *	-	2.59
Other	0.67	0.84	1.19	1.71	0.39	-

* Statistically significant at 0.05 level

The odds ratios for different causes of injuries are shown in Table 4.10.

Table 4.10 Four ratios were found to be significant. When the tasks are “fencing, installing lights, signs, etc.” or “installing plumbing, lighting fixtures,” the likelihood of a fatal accident is 2.88, and 2.01 times the likelihood of a non-fatal accident than when the task is “interior plumbing, ducting, or electrical work” respectively. In other words, “interior plumbing, ducting, or electrical work” is less hazardous than the other two tasks. Also, “fencing, installing lights, signs” and “installing plumbing and lighting fixtures” cause fatality, respectively, 2.39 and 1.67 times more than “Other” causes. One should note that the ‘Other’ category here is consisted of two types of cases: (1) causes with low frequency incidents and (2) reports in which the cause was not reported by OSHA inspectors. The first group represented less than 5% of the data. From 34 reported causes, only five had at least 5% and the rest (e.g., demolition, excavation, cutting concrete pavement) were labeled as ‘Other’. Other than low frequency causes, there were 130 cases in which the cause of injury was not reported by OSHA inspectors. Interestingly, the fatality rate of ‘not reported’ cases was the same as those of the ‘Other’ category in the data: 32% of injuries in both groups were resulted in a fatality. For this reason, the 'not reported' cases were also added to the 'Other' category when calculating the odds ratios and building the regression model.

Table 4.10. Odds ratios of fatal accidents among different causes of injuries.

Cause of Injury	Fencing, Installing Lights, Signs	Installing Equipment	Installing Plumbing	Interior Plumbing, Ducting	Temporary Work	Other
Fencing, installing lights, signs	-	1.74	1.43	2.88 *	1.52	2.39 *
Installing equipment	0.58	-	0.82	1.66	0.88	1.37
Installing plumbing	0.70	1.22	-	2.01 *	1.07	1.67 *
Interior plumbing, ducting	0.35 *	0.60	0.50 *	-	0.53	0.83
Temporary work	0.66	1.41	0.94	1.89	-	1.57
Other	0.42 *	0.73	0.60 *	1.21	0.64	-

* Statistically significant at 0.05 level

4.4.2 Final Logistic Regression Model

After selecting the final multi-variate model, one can discuss its coefficients for different levels of variables. Note that reference groups among variables are set to those with the highest fatality rates: accidents that have occurred in "demolition" projects, those where source of injury was a "vehicle," and those wherein the cause was "fencing, installing lights, signs, etc." From the 14 coefficients, six (excluding the constant) are significant at the level of 0.05, and three are marginally significant at the 0.10 significance level.

New adjusted odds ratios can be calculated easily from Table 4.11. For instance, to compute the adjusted odds ratios between "machinery" and "vehicle," one can first get the estimated logits under the model:

$$\hat{g}(SoI = Machinery; PT^* = Alteration\ or\ rehabilitation; CoI^* = Installing\ plumbing, lighting\ fixtures) = 1.862 - 0.733 - 0.607 - 0.420 = 0.102; \quad (1)$$

$$\hat{g}(SoI = Vehicles; PT = Alteration\ or\ rehabilitation; CoI = Installing\ plumbing, lighting\ fixtures) = 1.862 - 0.607 - 0.420 = 0.835 \quad (2)$$

* Values for PT and CoI could vary as long as same values are used in both cases

The difference between the two estimations is −0.733 and the odds ratio would be the exponential of −0.733, which is 0.48. Comparing this ratio to Table 4.9 reveals that controlling for “project type” and “cause of injury” has decreased the odds of fatality for machinery by around 17% (i.e., 0.58 to 0.48). Note that this odds ratio is more accurate than the unadjusted one in Table 4.9 (i.e., 0.58) as the effects of more accident factors are considered.

Table 4.11. Final model for the electrical contractors (number of remaining accidents = 566).

Accident Factor	Model Parameters	Coefficient	Std. Err.	z	*p*
	Constant	1.862	0.786	2.369	0.018 *
SoI	Machinery	−0.733	0.445	−1.646	0.100 †
	Parts and materials	−0.764	0.336	−2.276	0.023 *
	Structures and surfaces	−1.247	0.381	−3.272	0.001 **
	Tools, instruments, and equipment	−1.613	0.381	−4.235	2.29×10^{-5} ***
	Other	−1.072	0.545	−1.966	0.049 *
PT	Alteration or rehabilitation	−0.607	0.651	−0.933	0.351
	New project or new addition	−0.635	0.629	−1.009	0.313
	Maintenance or repair	−1.082	0.637	−1.699	0.089 †
	Other	−2.456	0.751	−3.268	0.001 **
CoI	Installing equipment (HVAC and other)	0.445	0.435	−1.023	0.306
	Installing plumbing, lighting fixtures	−0.420	0.476	−0.883	0.377
	Interior plumbing, ducting, electrical work	−1.032	0.450	−2.290	0.022 *
	Temporary work (buildings, facilities)	−0.389	0.526	−0.740	0.459

Other	−0.735	0.421	−1.748	0.081 †

Significance levels: 0–0.001: ***; 0.001–0.01: **; 0.01–0.05: *; 0.05–0.1: †. HVAC: heating, ventilation, and air conditioning.

4.4.3 Common Accident Patterns among Electrical Workers

To investigate common covariate patterns and study the performance of the model in predicting the fatality rates among them, the authors decided to look into the observed and predicted fatality rates among patterns with at least six cases (i.e., representing at least one percent of the data). Using this criterion reduced the number of patterns to 36, which represent 411 accidents or 73 percent of the data. Table 4.12 shows these results, which were sorted based on the observed fatality rates in each pattern. Among the 36 patterns, in 11 cases, the predicted value was within 5 percent of the observed fatality rates. These 11 patterns represent 38 percent of the 411 accidents in Table 4.12. Another 10 patterns, representing 28 percent of the accidents, were predicted with 6–10 percent of the observed rates. Among the rest of patterns, four of them did not have close prediction numbers, but yielded ranks predicted correctly as having the highest fatality rates. For instance, the predicted fatality rate for pattern 99 was 62 percent, which is the second highest prediction rate and indicates the high risk-level of this pattern. These numbers indicate the satisfactory ability of the model to detect the level of risk among various accident patterns.

Table 4.12. Observed and predicted probability of fatal accidents in 36 common covariate patterns.

ID	Observed	Estimate	SoI	Cause	Type	*n*
149	0.00	0.07	Structures and surfaces	Other	Other	6
137	0.00	0.23	Structures and surfaces	Other	Maintenance or repair	6
165	0.10	0.22	Tools, instruments, and equipment	Installing plumbing, lighting fixtures	Maintenance or repair	10
160	0.13	0.20	Tools, instruments, and equipment	Interior plumbing, ducting, electrical work	Alteration or rehabilitation	8
106	0.15	0.27	Parts and materials	Interior plumbing, ducting, electrical work	Maintenance or repair	13
179	0.17	0.05	Tools, instruments, and equipment	Other	Other	6
116	0.17	0.14	Parts and materials	Installing equipment (HVAC and other)	Other	6
173	0.17	0.25	Tools, instruments, and equipment	Other	New project or new addition	12
119	0.20	0.11	Parts and materials	Other	Other	10
172	0.27	0.20	Tools, instruments, and equipment	Interior plumbing, ducting, electrical work	New project or new addition	11
158	0.29	0.31	Tools, instruments, and equipment	Installing equipment (HVAC and other)	Alteration or rehabilitation	7
170	0.31	0.30	Tools, instruments, and equipment	Installing equipment (HVAC and other)	New project or new addition	13
113	0.31	0.43	Parts and materials	Other	New project or new addition	32
112	0.33	0.36	Parts and materials	Interior plumbing, ducting, electrical work	New project or new addition	24
108	0.33	0.41	Parts and materials	Temporary work (buildings, facilities)	Maintenance or repair	9
11	0.33	0.63	Vehicle	Other	Alteration or rehabilitation	6
143	0.35	0.32	Structures and surfaces	Other	New project or new addition	26
107	0.36	0.33	Parts and materials	Other	Maintenance or repair	31

171	0.36	0.31	Tools, instruments, and equipment	Installing plumbing, lighting fixtures	New project or new addition	11
104	0.36	0.39	Parts and materials	Installing equipment (HVAC and other)	Maintenance or repair	11
167	0.38	0.17	Tools, instruments, and equipment	Other	Maintenance or repair	8
161	0.38	0.25	Tools, instruments, and equipment	Other	Alteration or rehabilitation	8
142	0.38	0.26	Structures and surfaces	Interior plumbing, ducting, electrical work	New project or new addition	8
159	0.38	0.32	Tools, instruments, and equipment	Installing plumbing, lighting fixtures	Alteration or rehabilitation	8
110	0.41	0.51	Parts and materials	Installing equipment (HVAC and other)	New project or new addition	17
140	0.42	0.39	Structures and surfaces	Installing equipment (HVAC and other)	New project or new addition	12
101	0.47	0.37	Parts and materials	Interior plumbing, ducting, electrical work	Alteration or rehabilitation	17
47	0.50	0.33	Machinery	Other	Maintenance or repair	6
17	0.50	0.51	Vehicle	Other	Maintenance or repair	6
128	0.57	0.39	Structures and surfaces	Installing equipment (HVAC and other)	Alteration or rehabilitation	7
98	0.59	0.51	Parts and materials	Installing equipment (HVAC and other)	Alteration or rehabilitation	17
103	0.67	0.50	Parts and materials	Fencing, installing lights, signs, etc.	Maintenance or repair	9
114	0.67	0.52	Parts and materials	Temporary work (buildings, facilities)	New project or new addition	6
23	0.67	0.62	Vehicle	Other	New project or new addition	9
111	0.78	0.51	Parts and materials	Installing plumbing, lighting fixtures	New project or new addition	9
99	0.83	0.62	Parts and materials	Fencing, installing lights, signs, etc.	Alteration or rehabilitation	6

4.5 Discussion

To explore and quantify the effect of the three significant factors (i.e., “project type,” “source of injury,” and “cause of injury”) on fatality rates, odds ratios were calculated, and their significance was tested at the 0.95 level. Among “sources of injuries,” results indicate that “machinery” and “parts and materials’’ can cause death at significantly higher rates than “tools and instruments”. Moreover, in accidents where the primary “source of injury” was a “vehicle,” odds of a fatal injury would be two times, three times, and four times more than where the sources are “parts and material,” “structures and surfaces,” and “tools and instruments,” respectively. “Vehicles” in this study include highway motorized vehicles, powered off-road and industrial vehicles, and non-powered plant and industrial vehicles. All fatal accidents involving vehicles in this study can be categorized into four scenarios: (1) struck by/run over by passenger vehicles or construction vehicles; (2) electrocuted by touching part of a vehicle (e.g., boom) that was energized mistakenly; (3) fell from truck/forklift/bucket; and (4) pinned between vehicles or part of vehicles. Chi and Lin (2018) have mentioned the imperative of using vehicles that are in compliance with standards such as American National Standard for Powered Industrial Trucks. They also emphasized the need for required training and evaluation of operators based on OSHA standards. For instance, to avoid falls, OSHA 1910.178 declares that operators must avoid unsafe behaviors such as placing arms or legs between the uprights of the mast or outside the running lines of the truck or standing or passing under the elevated portion of any truck. Applying these recommendations could reduce the risk of accident scenarios involving vehicles.

While the odds ratios could show which level(s) of factors are more hazardous (i.e., produce higher fatality rates), they are not enough to explain the relationship between fatality rates and accident scenarios in which more than one factor is present. In other words, considering combinations of attributing factors in a regression context could possibly reveal more about the outcome of accidents (i.e., the probability of a fatal injury in this case) than when only one factor has been studied. As mentioned by Hosmer et al. (2013), the dependent variable is usually associated with more than one predictor and therefore considering only univariate models could be insufficient and misleading. Therefore, the authors took several steps to find the best fitting, most parsimonious, and efficiently interpretable multivariate logistic regression model. The final model is an additive model consisting of three factors: source of injury; type of project; and cause of injury. The analysis shows that this model is better than any of the other 17 possible models and the saturated model (Table 4.2 and Table 4.3), that it can fit the data well (Table 4.5 and Figure 4.2), and that it has a satisfactory performance on unseen data validated by Hosmer-Lemeshow test. To ensure that the model can fit individual patterns, the authors have also performed diagnostic tests, which revealed that three patterns do not fit the model. These patterns have been removed from the data.

Using coefficients of the final model (Table 4.11

Table 4.11), one can study new adjusted odds ratios among factors. These ratios are different from those in Tables 4.2–4.4, as the new ones are controlled for the use of other factors in the model. For instance, controlling for “project type” and “cause of injury,” the odds ratios of “vehicles” to “machinery,” “structure and surfaces,” “tools, instruments, and equipment,” and “other” sources would increase by 20 percent, 13 percent, 13 percent, and 13 percent, respectively, while the odds ratio between “vehicles” and “parts and materials” would be reduced by 1 percent (compared to the ratios that are not adjusted for “project type” and “cause of injury” in Table 4.9: “vehicles’” row). The significance of coefficients strongly confirms the effects of the selected factors on the rate of fatality.

The logistic regression model can also be used to estimate the probability of a fatal injury in a specific accident scenario. Using statistical methods to find common accident scenarios have been tested in previous studies. For instance, Chi et al. (2004) have applied Phi coefficients on fatal accidents to find strong positive associations between industry, sources of injury, and accident type. They reported that fatal falls from structures and construction facilities had the most obvious link to the construction industry. In this paper, the authors have considered both fatal and non-fatal accidents and used the fatality rates as the measure of risk among accident patterns.

From the 105 accident scenarios/patterns used to validate the model, 36 of them happened more frequently. Each of these scenarios occurred at least six times, and together they cover 73 percent of all accidents in this study. To identify and study more hazardous scenarios both in terms of severity and frequency, observed rates of fatality and total number of accidents in each scenario were calculated. Furthermore, the logistic model’s estimated rates of fatality were computed and compared to observed rates to detect which scenarios could be predicted correctly. The two most common patterns share the same source (i.e., “parts and materials”) and cause of injury (i.e., “other”), and only differ in their “project type.” Among the two, most accidents (32 cases) occurred in “new projects or new additions,” followed closely by “maintenance and repair” (31 cases). Considering the fatality rates, though, “new projects” have caused fewer fatalities (31 percent) than “maintenance” projects (36 percent). Without considering the source and cause of injury, the total fatality rates in “new projects” and “maintenance projects” were 40 percent and 35 percent, respectively. The reduction in fatality rates among “new projects” (when putting source and cause into picture) is notable. One can conclude that when the “source of injury” is a “part or material” (e.g., an electrical part) and the cause is “other,” “new projects” are much less hazardous than “maintenance projects.”

In terms of severity, there were nine patterns with at least 50 percent fatality rates. Among them, in five patterns the “source of injury” was “parts and materials”: two “alteration projects,” two “new projects,” and one “maintenance.” Reviewing the fatality rates among these five patterns further confirms the effect of “cause of injury” on fatality rates. Note that with the same source (i.e., “parts and material”) a different cause (i.e., “fencing, installing lights, signs, etc.”) increased the fatality rates from 36 percent to 67 percent in “maintenance projects” (Table 4.12). “Type of project” can also play a significant role in determining the fatality rates. For instance, when “source of injury” was a “part or material” and “cause” is “installing equipment,” the fatality rates could vary from 17 percent in other projects to 59 percent in “alteration or rehabilitation projects.” These jumps in probabilities signal that safety managers may benefit from preemptively addressing such coinciding factors in work situations to plan around the risks and thereby offset them.

Considering the prediction performance, while the range of actual fatality rates among 36 patterns was 0 to 83, the predicted rates represent a smaller range of 5 to 62 percent. This difference is important in interpreting the results as, for instance, an estimated value of 62 percent would refer to the most dangerous patterns whereas the same value might not be as alarming among observed rates. One should note that while this smaller range caused some large variations between the observed and estimated values—especially among larger rates (i.e., greater than 50 percent)—the level of severity was predicted correctly in these cases. As an example, consider the most fatal pattern (83 percent death rate), which occurred in “alteration projects” when the “source” and “cause of injuries” were “part and materials” and “fencing, installing lights, signs, etc.,” respectively. The predicted rate for this pattern (i.e., 62 percent) was the second highest predicted rate, which clearly declares the high risk of this accident pattern. The research team could identify four of such patterns in which the predicted rate—thought not close to actual rates—accurately determined the high-risk level of the pattern. Other than that, 11 and 10 scenarios were estimated within 5 percent and 10 percent of the actual rates, respectively. These 25 patterns represent 69 percent of the patterns and 73 percent of the accidents.

Another practical application of the results is related to risk assessment of projects for electrical contractors. For example, as shown in Table 4.12, in new projects or new addition projects, when workers were “installing plumbing, lighting fixtures” and exposed to “parts and materials”, the risk of fatality was the highest. In another example scenario, if the project was “alteration or rehabilitation” and activity was “fencing, installing lights, signs, etc.”, the risk of fatality was the highest again when workers were exposed to “parts and materials”. In these cases, a project or safety manager can assign more safety resources, conducting job hazard analysis, or toolbox meetings to increase awareness of construction workers regarding potential hazards.

4.6 Conclusions

Electrical contractors working in the construction industry are exposed to various hazardous situations leading to a high number of severe injuries and fatalities (Zhao et al. 2012). This problem necessitates the identification of high-priority accident scenarios both in terms of frequency and severity (i.e., probability of a fatal outcome) to thereby allow for risk mitigation. Accordingly, the objective of the present work was to first study individual effects of different contributing factors (e.g., project characteristics, sources and causes of accident) on degree of an injury and then explain the significant effects through a multivariate logistic model. Rather than estimating the probability of a single accident, the authors utilized three significant factors (i.e., “project type,” “source of injury,” and “cause of injury”) to form 108 accident scenarios (i.e., covariate patterns) and reviewed these scenarios both in terms of frequency and fatality rates. Model assessment techniques verified the fit of the model to the data and its capability to estimate the probability of a fatal injury in future cases. All analysis has been done using R language packages.

Implications of this study suggest that, when controlling for “project type” and “cause of injury,” “vehicles” cause significantly higher fatality rates than “tools, instruments, and equipment,” “structures and surfaces,” “parts and materials,” and to a lower degree “machinery.” Considering common accident scenarios (i.e., 36 scenarios that were repeated at least 6 times in the data), there were nine of patterns with fatality rates of 50 percent or more. The logistic model also assigned the highest probabilities to most of these scenarios, demonstrating the effectiveness of our approach to predictive accident modeling. These patterns should be considered as high-priority safety-risk scenarios by electrical contractors and their safety managers when allocating safety resources and planning for possible interventions. Outside of these practical implications, this study also contributes to the body of knowledge by providing a roadmap for building multivariate regression models for safety studies.

This work has discussed several contributing factors to analyze accidents that occur to electrical contractors. Future studies can incorporate more variables such as “time of the accident” to develop models with better prediction performance. The severity of accidents also can be defined more accurately by considering more variables such as “monetary cost of injuries” or “days away from work” for non-fatal incidents. Furthermore, more advanced, non-parametric machine learning methods can be applied to similar data for the same classification purposes. Addressing these limitations can improve the prediction ability of models in accident-severity classification problems.

CHAPTER FIVE: APPLYING A MULTI-LABEL MACHINE LEARNING APPROACH TO IMPROVE THE PREDICTION OF CONSTRUCTION ACCIDENT OUTCOMES

5.1 Introduction

Construction remains one of the most dangerous industries for its workers in the US (Kang et al. 2017; Baker et al. 2020). In 2019, more than two hundred thousand occupational injuries have occurred in the construction industry and more than one thousand workers lost their lives (Bureau of Labor Statistics 2021). Maintaining safe work conditions to mitigate hazardous incidents depends heavily on effective decision-making by those in key positions (Hayes and Maslen, 2015). However, making good safety decisions under uncertain conditions of construction projects has been a significant challenge for safety managers. Tixier et al. (2016b) pointed out that the human's limitation in inducing knowledge from a large number of historical observations is a main reason for the poor decisions on construction sites. This limitation is particularly important as learning from mistakes is critical in avoiding them to happen again (Hayes and Maslen, 2015).

To this end, unlike humans, machine learning models can be used to accurately predict the outcomes of hazardous situations, improve decision making, and ultimately save lives by learning from massive high-dimensional data (Veropoulos, 2001; Seera and Lim, 2014; Tixier et al., 2016b). Other than the direct safety benefits, correct prediction of accident outcomes can help many other agencies as mentioned by Iranitalab and Khattak (2017): safety planners who are interested in predicting the more hazardous/costly accidents to measure the cost of accidents on communities; the hospitals and first-respondents that need to provide appropriate medical assistant as fast as possible based on the severity of injuries; and insurance companies that need accurate estimations for their costumers' premiums based on, among other factors, the costs of occupational accidents, which depends heavily on accident severity. To benefit from these advantages, this study is focused on predicting the outcomes of construction accidents using machine learning techniques.

Predicting the type of an accident and its outcomes, through statistical modeling and machine learning techniques, has been one of the main topics of safety studies in the last two decades (Sarkar and Maiti, 2020). Most of previous studies have been focused on introducing novel optimization techniques to improve the performance metrics on individual accident outcomes such as nature of injuries or degree of injuries. For instance, Tixier et al. (2016b) have implemented an iterative approach, using out-of-bag error estimation, to optimize three hyperparameters (i.e., number of trees, number of features at each split, and stratified oversampling proportions) for the Random Forests models. Past studies (Mistikoglu et al., 2015; Baker et al., 2020) have validated the application of ML on accident data and expanded our knowledge about the potentials and limitations of these methods. For instance, Mistikoglu et al. (2015) showed that the extracted rules from decision trees can eliminate less important accident attributes and reveal the more critical attributes when predicting the degree of injuries. Baker et al. (2020) demonstrated that Natural Language Processing (NLP) methods can be combined to ML algorithms such as XGBoost and Support Vector Machines (SVM) to facilitate the extraction of accident attributes from historical reports resulting in faster implementations of ML pipelines on accident data.

On the other hand, these studies also enumerated some limitations on the application of ML in safety. For instance, both Mistikoglu et al. (2015) and Baker et al. (2020), recognized the need for collecting more attributes when gathering reports on previous accidents. Furthermore, Baker et al. (2020) also discussed the need for models that can predict the incident occurrence which requires data on non-accident scenarios such as near-misses. Another important aspect, that previous studies have failed to recognize, is that accidents result in multiple outcomes. However, the majority of previous studies have developed ML models to predict a single accident outcome (e.g. degree of injury) at a time which ignores the potential relationships and dependencies among different outcomes such as degree and nature of injury.

One should note that, when describing the outcomes of an occupational injury, inspectors usually identify the type of accident (e.g., fall, struck by), the nature of injury (e.g., burn, fracture), the injured part of body (e.g., head, extremities), or the degree/severity of injury (e.g., fatal, non-fatal). The potential relationship among these outcomes and their predictive power in a machine learning context have not been investigated in safety literature.

To address this knowledge gap, this study has adopted a multi-label approach and developed ML algorithms that can simultaneously consider multiple outcomes of an accident. The study, hence, aims to test the applicability of multi-label methods to improve the performance of ML algorithms with accident data. To achieve this goal, one first needs to study and affirm the correlation among accident outcomes and then verify the impact of these correlations on improving the predictive performance of ML models. Studying these relationships has two main advantages. The first advantage is related to improving the performance of predictive models. In plain terms, if two accident outcomes ‘a’ and ‘b’ are highly correlated, one can assume that knowing the value of ‘a’ can be a good predictor of the value of ‘b’. This can open a new way to improve the performance of ML models in safety studies. The second, and more practical, advantage is that even without a complicated model, one can apply this knowledge in safety planning. For instance, knowing the type and severity of injuries that are related to a specific type of accident (e.g., falls) can help safety management to assign safety resources accordingly.

The rest of the paper is organized in five sections. Section 5.2 summarizes the application of machine learning models in safety studies, provides examples of multi-label classification in other fields, and presents points of departures along with the two main hypotheses of the study. Section 5.3 outlines the methods used in this study to prepare data for machine learning experiments to examine the two main hypotheses. Section 5.4 presents the results of the experiments for both hypotheses and in section 5.5 the authors discuss these results and their implications (theoretical and practical) on construction safety. The conclusions and limitations of the study are presented in section 5.6.

5.2 Literature Review

The authors conducted an in-depth literature review regarding applications of ML in construction safety to determine their strengths and limitations and investigate the feasibility of a multi-label approach in accident predictions through reviewing similar applications of multi-label methods in other fields.

5.2.1 Machine learning in accident analysis studies in construction

Previous studies have shown a wide range of applications for ML in the field of construction safety. Text mining is one important application of ML in construction safety. As historical accident reports are a major source of data in any safety study, text mining methods have been proposed to significantly reduce the amount of time/effort needed for data preparation by developing methods that can automatically classify accident reports (Chen et al., 2015; Chi et al., 2016). As an example, Goh and Ubeynarayana (2017) have applied text mining methods to classify accident reports from OSHA into a limited number of labels (e.g. electrocutions, exposure to extreme temperatures, struck by falling objects) by training six machine learning algorithms. The authors concluded that no ML method could perform best on all labels and each method was suited to predict some labels over others. While all these improvements in applying ML to enhance safety on construction sites are promising, to limit the scope of work, the focus of this study is on the application of ML in predicting the outcome(s) of accidents using historical accident reports.

Machine learning methods have been vastly used in the last two decades to analyze historical accident reports to reveal common and hazardous accident scenarios and predict their outcomes. The two main types of accidents in these studies were vehicle crash incidents and occupational incidents. Many studies have applied methods such as Decision Trees (DT) (Chong et al., 2005; Chang and Wang, 2006; Abellan et al., 2013) and Artificial Neural Networks (ANN) (Mussone et al., 1999; Abdelwahab and Abdel-Aty, 2001; Abdel-Aty and Abdelwahab, 2004; Delen et al. 2006) on crash severity data. While DTs have been particularly popular in this field due to their interpretability and ability to extract accident scenarios/rules from large amounts of data (Montella et al., 2011, Gholizadeh et al. 2021), ANN methods such as multilayer perceptron could provide more accurate predictions (Abdelwahab and Abdel-Aty, 2001) in the cost of "explainability". Other statistical and machine learning methods such as Multinomial Logit (MNL), Random Forests (RF), Support Vector Machines (SVM), and Nearest Neighbors Classification (NNC) have been adopted to predict the severity of crashes in recent years (Iranitalab and Khattak, 2017).

Each method has its own advantages and limitations, and the choice of a particular method depends heavily on the type of data and the objective of the study. For instance, while model parameters (coefficients) in a regression type statistical method such as Multinomial Logit can readily reveal the associations among variables (i.e., each accident attribute and the target variable: severity of injury) (Guo and Geng, 1995, Gholizadeh et al. 2020 b), these models consider strong assumptions about data (linear associations and error distribution) that can lead to inadequate prediction performance (Iranitalab and Khattak, 2017).

On the other hand, machine learning methods utilize different ideas to improve prediction performance. For instance, RF follows the idea of training an ensemble of DTs (instead of one tree) with a random selection of features to split each node to create more generalizable models (Breiman, 2001), while SVM applies the idea of kernel functions to project data from a low-dimensional space to a space of higher dimension (Noble, 2006) to better separate different classes of the target variable (e.g., fatal vs. non-fatal injuries). However, optimizing the hyperparameters and describing the learning mechanisms of these machine learning models are more challenging than linear models. Iranitalab and Khattak (2017) concluded that while machine learning methods, particularly those with less data assumptions such as NNC, outperform MNL when considering the overall prediction performance, no single method can predict all severity levels adequately. This conclusion further reinforces the need for taking a different approach when predicting accident outcomes which is not limited to the characteristics of a specific algorithm but implements a more holistic path by considering the associations among various outcomes of an accident.

Analyzing occupational accidents is another major field in which machine learning methods have been implemented in the last 25 years. Sarkar and Maiti (2020) have pinpointed more than 230 articles on this topic between 1995 and 2019 with most of the studies published after 2010. They found that close to 94% of these studies have analyzed historical data (as opposed to surveys or real-time data), that the researchers from the United States have published the most papers (32%), and that Road and Construction were the sectors with most studies with 37% and 22% of the publications respectively. To this end, several studies have investigated the benefits of ML in predicting different accident outcomes, with more emphasis on the severity of injury as the main outcome (Gholizadeh and Esmaeili, 2016, Gholizadeh et al. 2018, Ayhan and Tokdemir 2019, Gholizadeh et al. 2021). Mistikoglu et al. (2015) have utilized two types of DTs to predict the severity of fall accidents among roofers. While the extracted rules from the DTs in this study can provide valuable information on the relationship between accident attributes such as safety training and the severity of injury, the accuracy of both methods were less than 70% on the test data set. The authors also did not provide the more important performance metrics such as recall and precision. Based on their previous works on automatically creating accident attributes/features from incident reports (Desvignes, 2014; Villanova, 2014; Esmaeili et al., 2015; Tixier et al., 2016a), Tixier et al., (2016b) trained two tree-based ML algorithms to predict four accident outcomes using bagging (i.e., Random Forest) and boosting (i.e., Stochastic Gradient Tree Boosting) ideas. Other than relying on a large dataset of more than 5,000 injury reports and developing a wide set of features, this study also implemented an in-depth procedure to optimize the hyperparameters of each method to improve the performance of the ML models on predicting the four target variables. The results indicated that while these models could predict three target variables well, they were not successful in predicting the most important one (i.e., severity of injury). Another limitation was that the prediction performance of the models was evaluated by a metric called Rank Probability Skill Score instead of standard metrics such as F1-score or the area under ROC curve.

To address these limitations, a follow up study conducted by Baker et al., (2020) that showed improvements in ML metrics for severity of injury. The second paper has benefited from a much larger data set (more than 90,000 samples), two more ML algorithms (i.e., XGBoost, and Linear SVM), and implementing model stacking ideas. The authors also have evaluated the models with standard metrics and analyzed feature importance in more details. While, unlike the first paper, the results show that the ML algorithms can outperform a random model when predicting the degree of injury, the metrics are still not high (e.g., the F1-score for more severe injury class for the best model is 0.37). Also, the evaluation metrics presented in this study are all based on one probability threshold. A more robust metric such as area under ROC curve (ROC_AUC) or a sensitivity analysis on more thresholds could provide a better picture of the models' performance.

This study investigates the third application of ML in construction safety. While previous studies have shown the promising capabilities of ML algorithms in predicting the outcome of injuries, their results clearly indicate that the performance of ML models on safety data is still far from ideal and hence there is a need to explore other methods and ideas. None of previous studies, to the best of our knowledge, have considered the relationships and dependencies among different accident outcomes as a potential to enhance the performance of ML models. To address this limitation, this study aims at first investigating the correlation among accident outcomes and then implement a multi-label approach that can benefit from this correlation to improve the prediction performance of ML models. Section 5.2.2 introduces multi-label classification and a few applications of this method in other fields.

5.2.2 Multi-label classification

A classifier which learns from a set of samples that are each associated with one label l, with K classes, from a set of disjoint labels L, is called a single-label classifier. Single-label classifiers typically address two problems: binary classification where K $=$

2, and multi-class classification if $K > 2$. In contrast, in multi-label classification, each sample is identified by a set of labels $Y \subseteq L$ (Tsoumakas and Katakis, 2007). Figure 5.1 presents different types of classification problem and their presentations in a dataset.

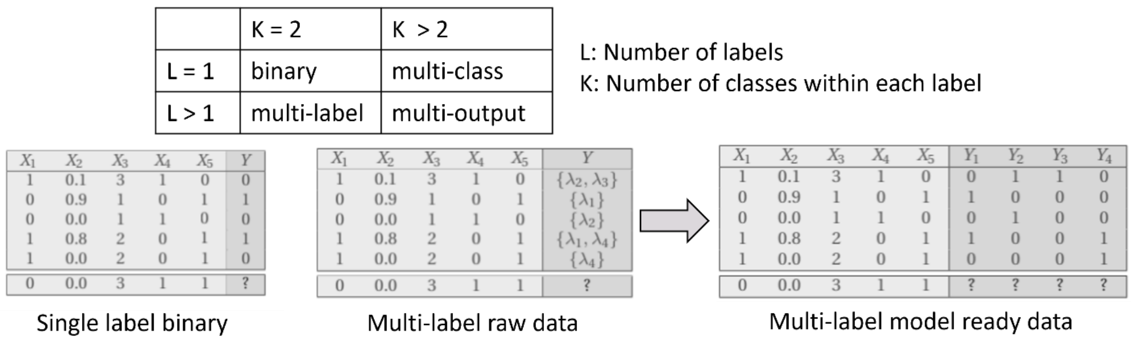


Figure 5.1. Single-label vs. multi-label classification (top) and the presentation of features and target(s) in a dataset

Multi-label classification has many real-world applications, including the automatic labeling of many resources such as texts (Charte et al., 2015), images (Boutell et al., 2004), music (Wieczorkowska et al., 2006), and video (Snoek et al., 2006). As pointed out by Zhang and Zhou (2014), while early researches on multi-label learning mainly applied these methods on text categorization (McCallum, 1999; Schapire and Singer, 2000; Ueda and Saito, 2003), more recent efforts have been focused on a diverse range of ML problems from multimedia contents (Boutell et al., 2004; Qi et al., 2007; Sanden and Zhang, 2011; Trohidis et al., 2008; Wang et al., 2008) to bioinformatics (Zhang and Zhou, 2006), web mining (Tang et al., 2009), rule mining (Veloso et al., 2007), information retrieval (Gopal and Yang, 2010), and tag recommendation (Song et al., 2008).

Methods to solve multi-label problems are usually divided into two groups: algorithm adaptation and problem transformation. As their names suggest, algorithm adaptation methods modify existing machine learning methods to handle multi-label space while problem transformation methods essentially convert the multi-label problem into one or more single-label classification problem(s) (Moyano et al., 2018).

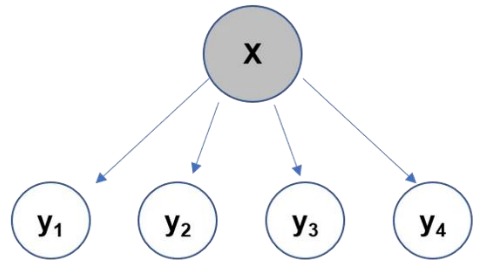
While previous studies have developed numerous algorithms in each group to address various multi-label problems in different fields, comparing the performance of all these methods on safety data is not in the scope of this study. Furthermore, because problem transformation methods are not limited to a specific algorithm and provide a basis for any ML algorithms, it is more reasonable to adopt them without going into the details of a specific algorithm. For these reasons and to limit the number of potential methods to study the dependencies among accident outcomes, this study chooses three problem transformation methods: Binary relevance (BR), Classifier Chains (CC), and Label Power-sets (LP). While all these methods can handle multi-label problems, their most important quality for this study is the degree to which each method considers the associations among labels during training. While BR ignores any potential association and LP only indirectly accounts for the associations when mapping labels into a class, CC directly utilizes these associations during training. This distinction among the three selected methods is essential to measure the impact of multiple accident outcomes on the overall prediction performance of the ML models.

5.2.2.1 Multi-label problem transformation algorithms

Binary Relevance (BR)

BR provides a framework to train multiple binary classifiers (one classifier for each label) using a fixed set of features and fit a model that can score all labels at once. The final model, however, is just a combination of separate binary classifiers and the performance of the model for each label is the same as a binary model for that label. By developing one binary classifier for each label, BR enables all labels to be trained at once and facilitates the computation of multi-label metrics such as exact match and hamming loss which evaluates the performance on all labels and not just one of them.

As the most intuitive method among problem transformation algorithms, BR has been applied as a baseline model in many fields such as emotion detection from text (Huang et al., 2021), emotion detection from music (Li and Ogihara, 2006), cancer pre-diagnosis (Ceylan and Pekel, 2017), and document classification (Yang et al., 2018). Figure 5.2 shows the BR algorithm representation where all labels use one set of features and no dependencies is assumed among labels.



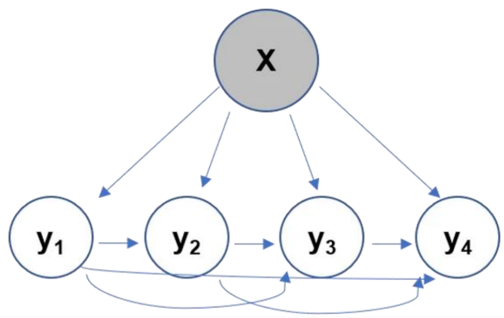
h :	x →	y
h_1:	[0,1,0,1,0,0,1,1,0]	1
h_2:	[0,1,0,1,0,0,1,1,0]	0
h_3:	[0,1,0,1,0,0,1,1,0]	0
h_4:	[0,1,0,1,0,0,1,1,0]	1

Figure 5.2. Binary Relevance

Classifier Chains (CC)

Similar to BR, the CC models train one binary classifier for each label, but the difference lies in the size of feature space in consecutive classifiers. To benefit from the potential correlation between labels, unlike the invariable feature space in BR, the feature space for each binary classifier in CC is extended by all labels of previous classifiers to form a chain. Each label, therefore, has the chance to play an important role in predicting later labels in the chain. Moreover, the order in which labels are organized can affect the outcome significantly. To account for this, when the number of labels in the chain are small, one can consider all order combinations to study the effect of orders and ultimately pick the order with the highest performance metrics.

Introduced by Read et al. (2009), CC has been one of the most popular multi-label methods and have been utilized in many areas since. Briggs et al. (2013), for instance, created an ensemble version of CC models to recognize bird species based on audio recordings of bird sounds. Mohamed et al. (2018) also have adopted CC to predict multiple activities of home residents to better design smart home environments. Figure 5.3 shows the CC algorithm representation where each label is trained on a set of features (X) plus all previous labels in the chain.



$\mathbf{h}$:	$\mathbf{x}' \rightarrow$	$\mathbf{y}$
h_1:	[0,1,0,1,0,0,1,1,0]	1
h_2:	[0,1,0,1,0,0,1,1,0,**1**]	0
h_3:	[0,1,0,1,0,0,1,1,0,**1**,**0**]	0
h_4:	[0,1,0,1,0,0,1,1,0,**1**,**0**,**0**]	1

Figure 5.3. Classifier chains

Label Powersets (LP)

LP considers every unique combination of labels as a separate class (e.g., with four labels 0000 → 0, 0001 → 1, 0011 → 2, etc.) and essentially converts the multi-label case to a multi-class problem. With L labels in the data, the maximum number of classes would be 2^L. As L increases, the number of classes can escalate fast. Some variants of LP have been developed in literature to address this issue by selecting only a sample of L labels randomly (e.g. RAndom k-labELsets or RAKEL [Tsoumakas and Vlahavas, 2007]) or limiting it to only more common labels (e.g. Pruned Sets [PS] and Ensembles of Pruned Sets [EPS] [Read et al., 2008]). For this study, however, *L* is small and one can consider all combinations that exist among labels to create the powersets.

Junior et al. (2017) used LP to classify data streams and presented comparable results to PS, EPS, BR, and CC. Furthermore, to detect hate speech and abusive language in Indonesian Twitter, Ibrohim and Budi (2019) developed several multi label models and revealed that LP models resulted in the best accuracy with fast computational time which shows the ability of LP when the label space is small. Figure 5.4 shows the LP algorithm representation where the set of labels in each data point are converted into one class.

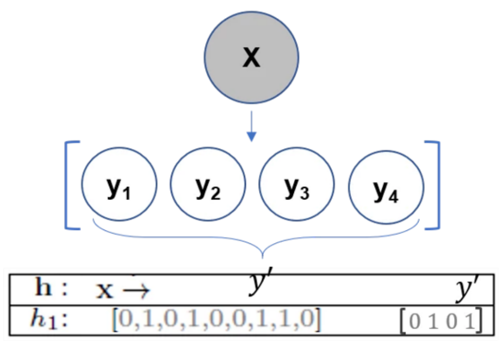


Figure 5.4. Label Powersets

5.2.2.2 Evaluation metrics

To compare the performance of multi-label algorithms, two types of metrics have been calculated in this study: i) metrics to measure the prediction performance of ML models and ii) a dissimilarity metric. Type (i) metrics are essential to determine if one multi-label method is superior to others in correctly predicting labels. Including multiple

Type (i) metrics can also ensure that different aspects of the models' performance are evaluated. Type (ii) metrics are more specific to the objective of this study as they can identify the method that better captures the similarities among labels. These metrics are explained in more details here.

Prediction performance metrics

The following metrics have been considered to evaluate the performance of multi-label classifiers:

- *Exact match (EM)*: For each observation in the test set, exact match is 1 only if all labels are predicted correctly and 0 otherwise. The EM for the whole test set is the average of these individual binary values and therefore is a number between 0 and 1 (Li et al., 2017). This is a conservative metric and higher values indicate better performance.
- *Hamming loss (HL)*: HL is the fraction of labels that are incorrectly predicted and therefore lower values indicate better performance. HL is less conservative than EM as it considers partial score for samples with one or more correct predictions (Li et al., 2017). The HL for the whole test set is the average of these individual binary values and therefore is a number between 0 and 1.
- *Average of area under ROC curve (ROC_AUC)*: The Receiver Operating Characteristic (ROC) curve is presented by a two-dimensional graph wherein the X axis represents values of false positive rate (FPR) and the True positive rates (TPRs) are plotted on the Y axis. Once the prediction

thresholds are selected and sorted, the corresponding TPR and FPR values would be computed for each threshold and their corresponding points would be plotted on the ROC space. The area under this curve (i.e., a value between 0 and 1) can be considered as a metric to evaluate the prediction performance of a classifier (Rakotomamonjy, 2004). The main advantage of using the area under ROC curve (ROC_AUC) as an evaluation metric is its independence from a specific probability threshold (e.g., 0.5) which makes it a more robust measure of performance than those that are calculated based on one threshold such as recall and precision. In multi-label problems, one can calculate the ROC_AUC for each label and then compute the average of these values which is a number between 0 and 1. Higher values represent better performance.

- *Mean Average precision (MAP):* Average Precision (AP) is the weighted average of precisions among thresholds of a precision-recall curve. Each precision value is weighted by the difference between recall of that threshold and recall of the previous threshold (Pecina, 2010).

$$AP = \sum_{n} (Recall_n - Recall_{n-1}) \times Precision_n \quad (1)$$

Similar to ROC_AUC, AP also benefits from considering various thresholds. In the multi-label case, the AP of the positive class is first calculated for each label and then the weighted (i.e., using the number of positives as weight) average of the individual APs is considered as MAP which is a number between 0 and 1. Higher values indicate better performance.

Dissimilarity metric

Other than the prediction performance metrics, one can examine the degree to which the multi-label algorithms could capture the correlation/similarity between labels. For instance, if two target variables are highly correlated, it is expected that the predicted values of them are also highly correlated. A multi-label method that can capture the magnitude of similarities among labels more accurately should have a better prediction performance as well. To calculate dissimilarity between pairs of targets, this study has adopted the Rogers-Tanimoto dissimilarity (RTD) measure (Rogers and Tanimoto, 1960). RTD is defined as:

$$RTD = \frac{2(c_{01} + c_{10})}{c_{11} + c_{00} + 2(c_{01} + c_{10})} \quad (2)$$

Where c_{ij} is the number of corresponding pairs of elements in binary vectors *u* and *v* respectively equal to *i* and *j*. RDT is bounded between 0 and 1 where 0 represents complete agreement ($c_{01} = c_{10} = 0$) and 1 represents total disagreement ($c_{11} = c_{00} = 0$) between the two vectors.

RTD has been used in several machine learning related studies. For instance, Sadri et al. (2006) adopted RTD as the similarity metric to categorized handwritten characters into similar clusters.

5.2.3 Points of Departure and Hypotheses

This study deviates from the current body of literature by proposing a multi-label machine learning approach towards the analysis of occupational construction accidents.

The findings of this study therefore provide new insight on the latent dependencies among accident outcomes and their impact on improving the prediction performance of machine learning models in this area.

To benefit from the relationships among accident outcomes in a predictive modeling context, one needs to address two challenges. The first challenge is to prove that the correlation/impact exists among accident outcomes within a machine learning context. To this end, the authors have hypothesized that:

Null Hypothesis 1: *No significant predictive impact exists among accident outcomes.*

A simple correlation analysis might not be enough to test this hypothesis as the predictive impact of accident outcomes on each other can be more complicated than a linear correlation. Therefore, a set of experiments are proposed in this study to test hypothesis 1 by statistically measuring the predictive impact of accident outcomes on each other.

The second challenge emerges from the fact that accident outcomes are not available before an accident occurs and therefore cannot be used directly as predictor variables of a model in real-world situations. For instance, consider a data analysist who has gathered some valuable information (e.g., the type of the project, the tools that workers are going to use) and plans to use a trained model to estimate the probability of a severe injury for each worker on the job site. As no accident has happened yet, the safety manager cannot use the type of accident or injury as inputs to the model. Therefore, even if hypothesis 1 is rejected, one cannot use accident outcomes as explicit predictors in a machine learning model. This is perhaps one of the main reasons that predictive modeling in safety has been limited to single-label classifiers. To this end, a multi-label approach is proposed that can capture the latent relationships among chains of accident outcomes from historical accident data, without using outcomes as explicit features of the model, and apply this knowledge to better predict accident outcomes in real project scenarios. This approach facilitates testing the second hypothesis of this study that:

Null Hypothesis 2: *Capturing the latent relationships between accident outcomes cannot improve the predictive performance of machine learning models*

5.3 Methodology

In this section we first introduce the data and describe the feature and target variables of the machine learning classifiers. Next, the two main hypotheses of the study and the process to test them are explained.

5.3.1 Data

To gather a reliable data to test the two hypotheses of the study, the authors collected 1,816 accident reports, occurred to specialty contractors, from the Occupational Safety and Health Administration (OSHA)'s Integrated Management Information System (IMIS) database. Specialty contractors in the construction industry play a critical role by performing manual construction tasks. However, such manual work means these workers face hazardous situations more than any other trades of construction. Reports from Bauru of Labor Statistics constantly show that more than 60% of construction fatalities occurs among specialty contractors (BLS 2021). Therefore, finding innovative ways to reduce the number of serious injurie among these trades is a promising option to enhance the overall workplace safety in the construction industry. To focus on specialty contractors that are faced with more hazardous situations, accident reports were gathered from three trades with the highest number of fatalities in recent years: roofing, site-preparation, and electrical contractors. These trades, together, are accounted for 43% of all fatalities among 19 specialty contractor trades between 2012 and 2019. Table 5.1 provides more details about these trades using both data collected from OSHA for this study and overall data from BLS.

Table 5.1. Summary of injury statistics for three specialty contractor trades with the highest number of fatalities

Specialty contractor name	Roofing	Electrical	Site preparation
Specialty contractor NAICS code	238160	238210	238910
Number of accident reports in this study	702	619	495
Average annual number of all non-fatal injuries (BLS: 2012-2019)	7,575	20,975	7,625
Average annual number of all fatal injuries (BLS: 2012-2019)	100	73	79

One should note that OSHA only requires documentation of ‘catastrophic’ accidents such as falls and electrocutions, wherein a work-related accident caused a fatality, in-patient hospitalization, amputation, or loss of an eye. It is also important to note that inclusion in OSHA’s database inherently means an accident occurred. Thus, studying this database enables researchers to assess accidents that occurred historically rather than assessing or predicting rates of accidents, that requires data on non-accident scenarios and near misses which is outside the scope of this study. Within each entry in OSHA’s database appears a summary of accident, reported by OSHA inspectors, and a limited number of variables used to describe the accident (e.g., event type, source, and cause of injury), its context (e.g., project end-use, type, cost, date), and its consequences (e.g., nature and degree of injuries, injured part of body). To process data, this study adopted categories found in the Occupational Injury and Illness Classification Manual (OIICM), developed by the U.S. Department of Labor Bureau of Labor Statistics (Bureau of Labor Statistics, 2012). Gholizadeh and Esmaeili (2020 a) provided more details on this data base and its content analysis approach.

For the purpose of this study, the focus is on the outcomes of construction accidents (i.e., target variable) and variables that can be used in a machine learning model to predict such outcomes (i.e., features or predictors). To this end, four target groups (i.e., event type, nature of injury, injured part of body, and degree of injury) and four feature groups (i.e., project end-use, project type, project cost, and source of injury) were selected from data. Except for the binary degree of injury, target and feature groups are categorical variables with multiple values. To prepare these variables for ML models, one needs to first convert them to binary variables. Table 5.2 presents feature and target categories (column 2). and the binary variables that are derived from them (column 3) along with the frequency and a code, for binary variables, that will be used to refer to them in the next sections.

Table 5.2. Ontology of categorical features and targets in machine learning models ordered by frequency

Variable type	Categorical Variables	Binary Variables	Frequency (proportion %)	Code
Feature		Nonresidential building	927 (51.1)	PE1

	Project End-use	Residential building	525 (28.9)	PE2
		Utility system	156 (8.5)	PE3
		Other heavy and civil engineering	117 (6.4)	PE4
		Highway, street, bridge	91 (5.0)	PE5
	Project Type	New project, new addition	585 (32.2)	PT1
		Maintenance, repair	512 (28.2)	PT2
		Alteration, rehabilitation	416 (22.9)	PT3
		Demolition	168 (9.3)	PT4
		Other project types	128 (7.1)	PT5
	Project Cost	Under $50,000	834 (45.9)	PC1
		$50,000 to $250,000	333 (18.3)	PC2
		$1,000,000 to $5,000,000	172 (9.5)	PC3
		$250,000 to $500,000	154 (8.5)	PC4
		$500,000 to $1,000,000	143 (7.9)	PC5
		$20,000,000 and over	86 (4.7)	PC6
		$5,000,000 to $20,000,000	83 (4.6)	PC7
	Source of Injury	Structures and surfaces	647 (35.6)	SI1
		Parts and materials	380 (20.9)	SI2
		Tools, instruments, equipment	255 (14.0)	SI3
		Other sources	228 (12.6)	SI4
		Machinery	169 (9.3)	SI5
		Vehicles	137 (7.5)	SI6
Target	Event Type	Falls to lower level	877 (48.3)	ET1
		Struck by object or equipment	340 (18.7)	ET2
		Exposure to electricity	301 (16.6)	ET3
		Caught in/between	174 (9.6)	ET4
		Other events	124 (6.8)	ET5
	Nature of Injury[1]	Fracture	648 (35.7)	NI1
		Electrocutions, electric shock	206 (11.3)	NI2
		Concussion	146 (8.0)	NI3
		Burn electrical	84 (4.6)	NI4
		Bruise, contusion, abrasion	81 (4.5)	NI5
		Burn heat	64 (3.5)	NI6
		Cut laceration	64 (3.5)	NI7

Injured Part of Body	Head	470 (25.9)	PB1
	Multiple body parts	336 (18.5)	PB2
	Upper extremities	280 (15.4)	PB3
	Body systems	279 (15.4)	PB4
	Trunk	254 (14.0)	PB5
	Lower extremities	197 (10.9)	PB6
Degree of Injury	Fatal	781 (43.0)	D1

[1] Some smaller categories of 'nature of injury' (e.g., dislocation, foreign body in eye) were not included in the final feature list

Table 5.2 shows that there are 23 potential features and 19 potential targets for machine learning models in this study. Other that the binary features mentioned in Table 5.2, the year of accident has also been added to the features as a continuous predictor.

5.3.2 Experiments to test hypotheses of the study

Two main hypotheses have been defined previously:

Hypothesis 1: *No significant predictive impact exists among accident outcomes.*

Hypothesis 2: *Capturing the latent relationships between accident outcomes cannot improve the predictive performance of machine learning models.*

To test these hypotheses, two sets of experiments were designed and implemented. This section presents these experiments in detail.

5.3.2.1 Testing hypothesis 1

Before explaining the designed steps to test hypothesis 1, the definition of "impact" in the context of machine learning performance in this study needs to be addressed. Impact is defined as the magnitude by which the prediction performance metric of a binary classifier is improved/declined by including additional features (i.e., accident outcomes) to the classifier. Impact is similar to the definition of relevant features introduced by John et al. (1994). A relevant feature influences the target variable (relevancy) in the presence of other features (non-redundancy) (Dash and Liu, 1997). A simple correlation study is not enough to explain the impact that one accident outcome might have on another one. For instance, in the data of this study, the part of body 'upper extremities' has the highest correlation with the event type 'exposure to electricity' (phi coefficient of 0.28). However, when including all the 18 other outcomes in a classifier to predict whether the injured part of body is 'upper extremities', the degree of injury 'fatal' was the feature with the highest predictive power among outcome variables (phi coefficient of 0.21). This manifests the non-linear nature of relationships between accident outcomes in advanced machine learning models, particularly when feature space is large.

To address this limitation of correlation analysis, this paper proposes an empirical method of directly comparing the prediction performance of models with different feature sets while controlling for all other parameters such as sample size, modeling method and model hyperparameters. With this approach, the differences in prediction performance are only the result of the extra features that one model has over the other ones. More specifically, for each target variable, three models are developed: v1, v2, and v3.

The v1 classifiers are baseline models where the features are only those that are available before an accident occurs (i.e., pre-accident variables such as source of injury or project type) and therefore there is no outcome variable in their feature space. As shown in Table 5.2, there are 23 variables in these models. On the other hand, both v2 and v3 contain outcome variables in their feature space, although v3 is a much more conservative approach than v2. The v2 classifiers contain all 23 features of v1 plus all outcome variables except the target of the model (i.e., 18 outcome variables) for a total of 41 features. Other than allowing all outcome variables in their feature space, v2 classifiers also take advantage of the fact that they contain variables from the same group as the target of the model. For instance, if the target of the model is to predict the probability of ‘falls to a lower level’ (from ‘event type’ group), v2 classifiers allow all other event types such as ‘exposure to electricity’, ‘struck by object or equipment’, ‘Caught in/between’, and ‘Other events’, to be part of the features. It is therefore expected that these classifiers show superior performance as they benefit from all possible scenarios of an accident event in this case (it would be easy for the model to predict fall if none of other event types have occurred). To limit this bias of including features from the same category of the target and reducing the size of feature space, the v3 models are proposed where not only no outcome is allowed from the target category but also only one outcome can be added to the feature set from the other three target categories to make a set of 26 features. In the example of target being ‘falls to a lower level’, the three additional targets for v3 should come from the other three groups (i.e., ‘degree of injury’, ‘nature of injury’, and ‘injured part of body’). This way, the model would have no direct information about the event type in the feature space which provides a more impartial comparison to v1.

In this paper, X represents the vector of pre-accident variables with each pre-accident variable denoted by x_i, where $i \in \{1, 2, \ldots, 23\}$, each accident outcome is denoted by y_j, where $j \in \{1, 2, \ldots, 19\}$, and for each outcome one classifier is trained

(h_j). The following steps summarize the proposed approach to check the impact of outcome variables by comparing the performance of three types of classifiers:

Train three different classifiers for each of the 19 targets in Table 5.2:

v1: a classifier only with pre-accident variables (i.e., binary features).

$$\hat{y}_j = h_j(X) \quad (3)$$

v2: a classifier with pre-accident variables (i.e., binary features) and all other accident outcomes (18 level-2 targets).

$$\hat{y}_j = h_j\left(X, y_1, y_2, \dots, y_{j-1}\right) \quad (4)$$

v3: a classifier with pre-accident variables (i.e., level-2 features) and only three outcomes from a different target group.

$$\hat{y}_j = h_j(X, Y), \quad (5)$$

$$where\ Y\ is\ a\ set\ of\ 3\ y\ from\ different\ outcome\ groups\ than\ y_j$$

Compare the average of prediction performance metric (i.e., area under ROC curve) for each classifier.

Identify features with the most contributions in predictions for v2 and v3 to determine if outcome variables are among top predictors. Note that v1 models do not have any outcome variable in their feature space and therefore are excluded from this step.

If the predictive performance of the models with outcome variables is not significantly better than v1, then one can conclude that outcome variables have no significant impact on each other.

As mentioned before, v3 models include one binary outcome variable from each outcome category (i.e., event type, nature of injury, part of body, and degree) except the category that target of the model is from. This one outcome variable should ideally be the most predictive one of its category. One should note that a v2 classifier is already trained for the target and therefore one knows the power of each outcome variable in predicting that target. This ordered list of variables can be used to choose the three outcome variables for v3: from each outcome category, the variable that has the highest contribution in v2 can be selected. Figure 5.5 shows a hypothetical v2 classifier to predict values of y_1 and 15 most important features. Features with the same background color are coming from the same category. The v3 classifier for y_1 would include all X features and one feature from each outcome category that is the most important in its category. One can see that among outcome variables $[y_2, y_3, y_4, y_5]$, that are from the same category, only y_2 is selected for v3.

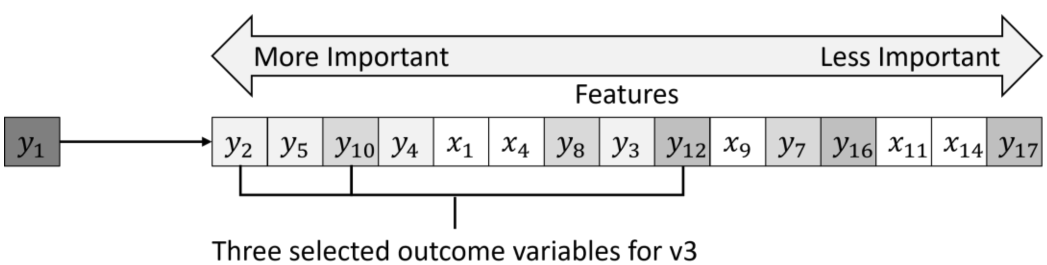


Figure 5.5. An example to show the method of selecting three outcome variables for v3 from features of v2

The importance of each feature is determined by its contribution in generating the value of target predictions. This contribution is calculated through SHAP (SHapley Additive exPlanations) values in this study. In the context of game theory, Shapley value is the average of marginal contributions, of each player, across all possible permutations of a set of players (Winter, 2002). Inspired by this method of measuring the contribution of players in a game, Lundberg and Lee (2017) presented a unified framework for interpreting predictions called SHAP which quantifies the contributions of features in ML models. SHAP has been widely applied in academia and industry to explain the results of machine learning black-box models (Parsa et al., 2020; Rodríguez-Pérez and Bajorath, 2020).

5.3.2.2 Testing hypothesis 2

To test the second hypothesis of the study, the authors first defined all experiments needed to test the hypothesis and then, in each experiment, implemented multi-label algorithms on the data. To achieve this, a complete machine learning pipeline was designed and implemented in Python to automate data preparation, target definition, feature engineering, train/test split, model training, and validating model performance on test data. This section explains these steps in detail.

Creating labels and designing experiments

To test the second hypothesis, we need to consider different ways that these variables can be modeled for the analysis. As presented in Table 5.2, there are 19 target binary variables (e.g., falls to lower level, fracture, head, fatality) from four target groups (i.e., event type, nature of injury, injured part of body, and degree of injury) in this study.

To limit the experiments to more common labels, only those with more than 10% frequency are considered as valid labels which includes: three event types (i.e., fall to a lower level, struck by, and exposure to electricity), two natures of injury (fracture and electrocution), all six parts of body, and degree of injury (a total of 12 labels).

Similar to the v3 models in section 5.3.2.1, the experiments are designed based on the idea that there should be only one label from each of the four target groups in the label-sets. This would result in 36 unique label-sets where each set consists of four binary labels. Table 5.3 shows these experiments with labels representing variable codes from Table 5.2.

Table 5.3. Labels in 36 multi-label experiments

Experiment number	Labels	Experiment number	Labels	Experiment number	Labels
1	ET1-NI1-PB1-D1	**13**	ET2-NI1-PB1-D1	**25**	ET3-NI1-PB1-D1
2	ET1-NI1-PB2-D1	**14**	ET2-NI1-PB2-D1	**26**	ET3-NI1-PB2-D1
3	ET1-NI1-PB3-D1	**15**	ET2-NI1-PB3-D1	**27**	ET3-NI1-PB3-D1
4	ET1-NI1-PB4-D1	**16**	ET2-NI1-PB4-D1	**28**	ET3-NI1-PB4-D1
5	ET1-NI1-PB5-D1	**17**	ET2-NI1-PB5-D1	**29**	ET3-NI1-PB5-D1
6	ET1-NI1-PB6-D1	**18**	ET2-NI1-PB6-D1	**30**	ET3-NI1-PB6-D1
7	ET1-NI2-PB1-D1	**19**	ET2-NI2-PB1-D1	**31**	ET3-NI2-PB1-D1
8	ET1-NI2-PB2-D1	**20**	ET2-NI2-PB2-D1	**32**	ET3-NI2-PB2-D1
9	ET1-NI2-PB3-D1	**21**	ET2-NI2-PB3-D1	**33**	ET3-NI2-PB3-D1
10	ET1-NI2-PB4-D1	**22**	ET2-NI2-PB4-D1	**34**	ET3-NI2-PB4-D1
11	ET1-NI2-PB5-D1	**23**	ET2-NI2-PB5-D1	**35**	ET3-NI2-PB5-D1
12	ET1-NI2-PB6-D1	**24**	ET2-NI2-PB6-D1	**36**	ET3-NI2-PB6-D1

ET: Event Type; NI: Nature of Injury; PB: Part of Body; D: Degree of Injury

As mentioned before, the order of the labels in each set is important in the CC method. With four labels, there are 24 possible permutations to be considered in each CC experiment. In other words, to test hypothesis 2, for each experiment 26 (1 BR + 24 CC + 1 LP) models are trained: a total of 936 models.

Implementing multi-label algorithms

Model development pipeline

To automate the process of training 936 models, a configurable machine learning pipeline was designed and implemented in Python. The pipeline is responsible for all required steps from reading in the raw data, preparing the features (excluding the unnecessary columns, encoding categorical features), creating label-sets and their orders, and randomly splitting data into train and test sets, to training multi-label models, and persisting the performance of them on test data. The details of the pipeline are presented in Appendix 1.

Base classifier

All problem transformation multi-label methods need a base classifier to predict the binary targets. Rivolli et al. (2020) tested several base classifiers in multiple binary strategies and found out that Random Forests was the best base algorithm among non-ensemble strategies (BR, CC, and LP are all non-ensemble methods). In order to control for the effect of this base classifier, the random forest with the hyperparameters presented in Table 5.4 has been implemented in all experiments. One may note that optimizing the base classifier's hyperparameters in each experiment could improve the results of this study. In fact, the parameters presented in Table 5.4 are the product of numerous rounds of optimization on different experiments. However, to control for the impact of hyperparameters on the performance of classifiers and limit the study on the effects of different multi-label algorithms, an invariable set of hyperparameters are used for all experiments.

Table 5.4. Hyperparameters for the base random forest classifier

Base classifier hyperparameters	Value
Number of estimators	2
Maximum depth of trees	3
Split criterion	Gini index
Minimum number of samples required to split	2
Minimum number of samples at a leaf node	1
Number of features to consider in each tree	0.5 * number of all features
Bootstrap samples used when building trees	True
Warm start	False
Class weights	{0:1, 1:2}

Evaluating the performance of multi-label model

As mentioned in section 5.2.2.2, there are four metrics to measure the predictive performance of the models and one metric that measures the degree to which the multi-label algorithms capture the dissimilarity between labels. Rogers-Tanimoto dissimilarity metric can be calculated between two arrays of binary variables. For labels ‘a’ and ‘b’, one can calculate one metric to compute the dissimilarity between actual values (RTD_{actual}) and one RTD to capture the dissimilarity among the predicted values for each multi-label model ($RTD_{predicted-BR}, RTD_{predicted-CC}, RTD_{predicted-LP}$). Next, the difference between RTD of each algorithm and the actual RTD is calculated (i.e., $\Delta RTD_{BR}, \Delta RTD_{CC}, \Delta RTD_{LP}$) and the algorithm with the lowest difference would be considered as the one that captures the dissimilarity the best. In this study, each experiment has four labels and therefore there are six possible pairs to be tested. For each multi-label algorithm, after predicting the labels on the test set, the average of the six $\Delta RTDs$ would be considered as the dissimilarity metric for that algorithm.

5.3.2.3 Statistical tests to compare the results of the experiments

To statistically test the two hypotheses, the Paired Sample t-test was employed. The reason to use the paired t-test instead of the regular t-test is that, for both hypotheses, the subjects/samples are the same in all experiments (Ross and Willson, 2017). To use an example, the experiments here are similar to studies where the effect of a treatment is tested by comparing a metric (e.g., blood sugar) before and after giving a treatment to same subjects. Similarly in this study, the metrics (e.g., ROC_AUC) are calculated for different methods but on the same test data. The paired samples t-statistic is calculated as:

$$t = \frac{\bar{X}_D - \mu_0}{S_D/\sqrt{n}} \qquad (6)$$

Where $\bar{X}_D$ and S_D are the mean and standard deviation of differences among all pairs respectively, n is the number of samples and μ_0 is the true mean of differences among all pairs (i.e., 0 when testing if the difference is significant). The significance level in this study is 0.05.

After determining the significance of the difference between two methods, one also needs to check the magnitude of the difference. Two useful statistics to measure the magnitude of the difference (i.e., the effect size) and the strength of the relationship between the two variables are “Cohen’ d” and “Pearson’s r” respectively.

Cohn’s d is the standardized mean difference between two populations (Cohen, 2013):

$$Cohen's\ d = \frac{\bar{X}_1 - \bar{X}_2}{s} \qquad (7)$$

Where $\bar{X}_1$ and $\bar{X}_2$ are the mean of the two samples and s is the pooled standard deviation. If s_1 and s_2 are the standard deviations and n_1 and n_2 are the sample sizes of sample 1 and 2 respectively, the pooled standard deviation is defined as (Thalheimer and Cook, 2002):

$$s = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2}} \qquad (8)$$

Cohen's d is bounded between -1 and 1 where values close to -1 and 1 indicate larger effect sizes. Pearson's r measures the linear correlation among two sets of data. When comparing two machine learning models, it is expected that the prediction metrics are positively correlated.

5.4 Results

This section presents the results of the two hypotheses of this study through statistical tests. First, section 5.4.1 tests the hypothesis that "no significant predictive impact exists among accident outcomes" using the average values of ROC_AUC among three types of classifiers (i.e., v1, v2, v3). Next, in section 5.4.2, the results of testing hypothesis 2 is presented to determine whether capturing the latent relationships between accident outcomes can improve the predictive performance of machine learning models. To do so, first the results of dissimilarity comparisons among labels are presented to determine which multi-label method can better capture the dissimilarities within the predicted labels. Next, the prediction performances of three multi-label algorithms are compared using the average values of four metrics (i.e., exact match, hamming loss, ROC_AUC, and average precision).

5.4.1 Hypothesis 1

To compare the performance of three classifiers (i.e., v1, v2, v3) presented in section 5.3.2.1, one can calculate the average of the performance metric (e.g., ROC_AUC) for all 19 experiments of each classifier. Table 5.5Table 5.5 provides the summary statistics on the ROC_AUC values for each classifier. As one can see, v2 has the highest ROC_AUC average followed by v3. As mentioned in section 5.3.2.1, both v2 and v3 utilize accident outcomes as model features (i.e., predictor variables). Therefore, the higher prediction performance of v2 and v3 over v1 suggests that the predictions of one accident outcome can be impacted by adding other outcomes into the classifiers' feature space. In other words, adding accident outcomes as predictors to v2 and v3 could improve their prediction performance.

Table 5.5. Summary statistics for ROC_AUC values for the three classifier versions

Classifiers	N	Mean	SD	SE	95% Conf. Interval	
v1	19	0.74	0.14	0.03	0.67	0.80
v2	19	0.94	0.10	0.02	0.89	0.99
v3	19	0.82	0.13	0.03	0.76	0.88

Next, to statistically confirm the prediction performance improvements of v2 and v3 over v1, a paired sample t-test was designed to compare the average of performance metric, and the results are presented in Table 5.6. One may note that, the objective of hypothesis tests, in this section, is to compare the baseline model to each of alternate models separately. For this reason, a variant of t-test was selected instead of ANOVA which can compare the average values among more than two samples. The results clearly show that both v2 and v3 classifiers have significantly higher ROC_AUC values than v1 and hence are superior in performance (Difference < 0 test). The results of ‘v1 - v3’ is particularly interesting as v3 classifiers only use three additional accident outcomes from groups other than the target group. This leads to a more conservative approach than v2 which includes more additional outcomes (18 additional features versus 3) in the feature space and outcomes can be from the same target group. The fact that even v3 models can improve the performance significantly over v1, indicates that accident outcomes have significant impact on predicting each other. In other words, knowing the value of one outcome group can reveal valuable information about the value of another group. The results provide enough evidence to reject the hypothesis that “accident outcomes have no significant impact on each other.”

Table 5.6. Results of paired samples t-tests (v1 – v2 and v1 – v3)

Tests	Differences	df	t statistic	Two side p-value	Diff. < 0 p-value	Diff. > 0 p-value	Cohen’s d	Pearson’s R
v1 - v2	-0.20	18	-6.74	**0**	**0**	1	-1.68	0.85
v1 – v3	-0.08	18	-5.30	**0**	**0**	1	-0.62	0.78

The authors also checked the SHAP values to determine the features with the most contributions to the predictions. As shown in Table 5.7, in all 19 cases, there are one or more outcome variables among the top five SHAP features for both v2 and v3 models. In 13 cases of v3, all three additional outcome variables were among the five most important features (e.g., when predicting fracture, three accident outcomes including “falls to lower level” [ET1], “fatality” [D1], and “lower extremities” [PB6] are among top 5 features). These results further reveal the importance of outcome variables in this experiment.

Table 5.7. Five features with the highest SHAP values in V2 and V3

Target	Feature codes with the most contribution in V2	Feature codes with the most contribution in V3
Falls to the lower level	SI1, **ET2**, SI2, SI3, **ET3**	SI1, SI2, SI3, **NI1**, SI4
Struck by object or equipment	**ET1**, **ET3**, **ET4**, SI1, **NI2**	SI1, **NI2**, SI6, PT4, PE5
Exposure to electricity	SI2, **NI2**, **NI4**, **ET1**, **ET2**	SI2, **NI2**, SI1, **D1**, **PB6**
Caught in/between	**ET1**, **ET2**, **ET3**, SI5, SI6	SI5, SI1, SI6, **PB5**, PE4
Other events	SI4, **NI6**, **ET1**, **ET2**, **PB4**	SI4, **NI6**, **PB4**, **D1**, SI1
Fracture	**ET1**, **D1**, **ET3**, **NI3**, SI2	**ET1**, **D1**, SI2, **PB6**, SI1
Electrocutions, electric shock	**ET3**, SI2, **NI4**, **ET1**, **D1**	**ET3**, SI2, **D1**, SI1, SI4
Concussion	**PB1**, **NI1**, **ET1**, **D1**, SI2	**PB1**, **ET1**, **D1**, SI2, PE2
Burn electrical	**ET3**, **NI2**, SI2, **D1**, **ET1**	**ET3**, **D1**, SI2, **PB3**, PE1
Bruise, contusion, abrasion	**PB1**, **ET3**, PT4, SI3, SI2	**PB1**, **ET3**, PT4, SI3, SI2
Burn heat	**ET5**, SI4, **D1**, SI2, **NI1**	**ET5**, **D1**, SI4, SI2, **PB4**
Cut laceration	**ET2**, SI3, **NI1**, **PB1**, PE3	**ET2**, SI3, **PB1**, **D1**, PE3
Head	**PB2**, **PB4**, **PB3**, **PB5**, **NI3**	**D1**, **NI3**, **ET1**, SI2, SI1
Multiple body parts	**PB1**, **PB4**, **PB5**, **PB3**, **PB6**	**ET1**, SI1, **D1**, **NI3**, PT2
Upper extremities	SI2, **D1**, **PB1**, **PB2**, **PB4**	SI2, **D1**, **ET3**, **NI2**, SI1
Body systems	**D1**, **PB1**, **PB2**, **PB5**, **PB3**	**D1**, **NI2**, SI4, **ET3**, PC7
Trunk	**PB1**, **PB2**, **PB3**, **PB4**, **PB6**	**ET4**, **NI1**, PE1, SI2, **D1**
Lower extremities	**D1**, **NI1**, **PB1**, **PB2**, **PB3**	**D1**, **NI1**, **ET2**, PT3, SI2
Fatal	**NI1**, **PB6**, **PB3**, **PB1**, **NI4**	**NI1**, **PB6**, SI3, SI2, SI6

5.4.2 Hypothesis 2

Three multi-label algorithms were designed to predict 36 different accident scenarios. As mentioned in Table 5.3, each scenario consists of four binary targets (i.e., accident outcomes). To verify hypothesis 2 (i.e., capturing the latent relationships between accident outcomes cannot improve the predictive performance of machine learning models), the authors first utilized a dissimilarity metric to determine the algorithm that can better capture the relationships between accident outcomes. Next, the prediction performance of three algorithms is calculated through four performance metrics. The findings indicate that the algorithm which captures the similarities among labels more accurately can significantly outperform the other two algorithms in prediction performance.

5.4.2.1 Dissimilarity comparison between labels

Table 5.8 provides the summary statistics of the Rogers-Tanimoto dissimilarities for multi-label models in 36 experiments. The results suggest that CC provides predictions that are closest to the actual labels and hence can capture the correlation among target variables better than the other two methods. The t-statistics of the paired samples along with their p-values, shown in Table 5.9, confirm that the difference between CC's label predictions and the actual labels is significantly lower than such difference when the multi-label method is either BR or LP. The comparison between LP and BR also shows that the LP method is less effective than BR in capturing the similarity between label predictions.

Table 5.8. Summary statistics of dissimilarity metric for the three multi-label methods

Multi-Label method	N	Mean	Std	SE	95% Conf. Interval	
BR	36	0.08	0.03	0.01	0.07	0.09
CC	36	0.05	0.03	0.01	0.04	0.06
LP	36	0.22	0.17	0.03	0.16	0.28

Table 5.9. Results of paired samples t-tests (CC – BR, LP – BR, CC – LP) for dissimilarity metric

Test	difference	df	t statistic	Two side p-value	Diff. < 0 p-value	Diff. > 0 p-value	Cohen's d	Pearson's R
CC-BR	-0.03	35	-3.36	0.0019	0.0009	0.9991	-0.85	0.49
LP-BR	0.14	35	5.01	0	1	0	1.33	0.65
CC-LP	-0.16	35	-5.95	0	0	1	-1.61	0.71

5.4.2.2 Prediction performance metrics

This section presents the results of 36 experiments to investigate whether multi-label approaches can capture the significant impact that accident outcomes have on each other. Table 5.10 includes the summary statistics of four multi-label performance metrics of 36 experiments. The average values suggest that the CC method can improve the prediction performance of multi-label classifiers. The significance of these improvements is tested using paired sample t-test and the results are shown in Table 5.11. For all four metrics, CC shows a significant improvement over both BR and LP methods. LP method only shows significant improvement over BR method in one of the metrics (i.e., exact match) while BR outperforms LP regarding ROC_AUC and MAP metrics. There is no significant difference between the Hamming loss values of LP and BR methods.

Table 5.10. Summary statistics of four performance metric values for the three multi-label methods

Metric	Multi-Label method	N	Mean	Std	SE	95% Conf. Interval	
Exact Match	BR	36	0.30	0.08	0.01	0.27	0.33
	CC	36	0.36	0.06	0.01	0.34	0.38
	LP	36	0.33	0.07	0.01	0.31	0.35
Hamming Loss	BR	36	0.26	0.04	0.01	0.25	0.27
	CC	36	0.25	0.04	0.01	0.23	0.26
	LP	36	0.26	0.04	0.01	0.24	0.27
ROC_AUC	BR	36	0.70	0.04	0.01	0.69	0.72
	CC	36	0.72	0.04	0.01	0.71	0.73
	LP	36	0.69	0.06	0.01	0.67	0.71
Avg. Precision	BR	36	0.56	0.08	0.01	0.54	0.59
	CC	36	0.58	0.10	0.02	0.55	0.62
	LP	36	0.56	0.08	0.01	0.53	0.59

Table 5.11. Results of paired samples t-tests (CC – BR, LP – BR, CC – LP) for prediction performance metrics

Metric	test	difference	df	t statistic	Two side p-value	Diff. < 0 p-value	Diff. > 0 p-value	Cohen 's d	Pearson 's R
Exact Match	CC-BR	0.06	35	10.65	**0**	1	**0**	0.90	0.87
	LP-BR	0.03	35	3.76	**0.0006**	0.9997	**0.0003**	0.42	0.54
	CC-LP	0.03	35	4.41	**0.0001**	1	**0**	0.50	0.60
Hamming Loss	CC-BR	-0.01	35	-10.11	**0**	**0**	1	-0.33	0.86
	LP-BR	0.00	35	-0.82	0.42	0.21	0.79	-0.06	0.14
	CC-LP	-0.01	35	-3.24	**0.00**	**0.00**	1.00	-0.28	0.48
ROC_AUC	CC-BR	0.02	35	15.73	**0**	1	**0**	0.40	0.94
	LP-BR	-0.01	35	-3.64	**0.0009**	**0.0004**	0.9996	-0.29	0.52
	CC-LP	0.03	35	7.65	**0**	1	**0**	0.64	0.79
Avg. Precision	CC-BR	0.02	35	5.48	**0**	1	**0**	0.22	0.68
	LP-BR	-0.01	35	-2.24	**0.0319**	**0.016**	0.984	-0.08	0.35
	CC-LP	0.03	35	5.31	**0**	1	**0**	0.28	0.67

5.5 Discussions

An integrated machine learning and statistical testing approach was applied to investigate (1) the predictive correlation/impact among accident outcomes and (2) the effect of such impact on improving the predictive performance of machine learning models. Accordingly, two hypotheses were proposed.

The results showed that there is enough evidence to reject hypothesis I which means that even a model with limited access to accident outcomes in their feature space (i.e., v3 classifiers with only three outcome variables) can reach significantly higher performance levels than a model with only pre-accident attributes as predictors. This finding clearly shows that knowing the nature of the injury, for instance, can help predict the severity of the injury and vice versa. Furthermore, the results of SHAP analysis evidently demonstrated the very powerful contribution of accident outcomes of certain categories in predicting outcomes of other categories. For instance, the three most important variables in predicting concussion were whether the injured part of the body is “head”, whether it was a “fall to lower level” accident, and whether the accident has resulted in “fatality”. Intuitively, it makes sense that knowing the value of these three outcomes can help in predicting concussion. The findings of this study, through statistical tests, showed that this intuition was right. More interestingly, in all 19 experiments that were completed to test this hypothesis, there was at least one outcome variable in the top five predictors for v3 classifiers. In 13 out of 19 experiments, all three accident outcomes were found to be among the top five predictors. This is compelling evidence of a significant correlation/impact between accident outcomes. The v2 and v3 models can also provide practical applications in situations where an accident outcome is missing from data. Using other available outcomes, one can accurately predict the missing outcome.

While accident outcomes demonstrated significant predictive power, these variables cannot be used *directly* as valid features of machine learning models in real situations in the future as they are not available before accident occurrence (Baker et al. 2020). Hence, after showing the relationship between accident outcomes by rejecting the first hypothesis, the authors focused on testing multi-label ML methods that can capture such correlations and their effects on improving prediction performance without *directly* using accident outcomes as predictors. To this end, three multi-label algorithms were selected based on their level of incorporating label correlations during training. While the BR method does not utilize the correlation among labels and the LP method incorporates such correlation inexplicitly, the CC method fully employs the correlations among labels by creating a chain of labels where the *predicted* values of the first label in the chain are being used as a new feature in predicting the values of the second label and so on. After completing 36 experiments with each of the three algorithms, the results first revealed that, on average, the predicted values of CC are closest to the actual labels than the other two methods. This confirms that, in practice, CC preserves the relationships between labels more than the other two methods. This can be mainly due to the fact that each label participates in the training process of predicting the subsequent labels and therefore the potential correlation between labels is reflected, to some extent, in the predicted values. The findings also show that CC has the highest predictive performance in all four performance metrics proposed in this study. Since all the multi-label algorithms in this study utilized the same data, base classifier, and hyperparameters, the authors discuss that the better performance of CC is mainly because of the fact that it captures the correlations among labels more than the other two methods. Therefore, there is enough evidence to reject the second hypothesis which means capturing the latent relationships between accident outcomes can improve the predictive performance of machine learning models.

The findings of this study offer multiple contributions to academia and applications to industry. A major contribution of our study is the validation of the Classifier Chains (CC) multi-label machine learning method, that captures the relationships among accident outcomes, in improving the overall performance of accident outcome prediction models. This result indicates that machine learning methods can benefit from more than one accident outcome during the training to improve their overall prediction performance. This can have multiple implications for safety studies. First, it emphasizes the importance of collecting more variables not only to describe the context and causes of an occupational accident but also to paint a better picture of its outcomes and consequences of it. Most accident outcome studies have been focused on expanding the feature space of their ML models by detecting more pre-accident attributes (Cheng et al. 2020, Zhu et al. 2021). For instance, Tixier et al. (2016 b) built their models based on 79 attributes to predict four accident outcomes. This emphasis on detecting more pre-accident attributes is reasonable when the goal is to train single-label models. However, the results of this study indicate that accident outcomes can also be utilized during the training of ML models using the classifier chain approach. This suggests that more keywords from accident reports can be extracted by NLP methods such as those introduced by Tixier et al. (2016 a) to describe accident outcomes. It also emphasizes the need for more details on the consequences of accidents in the reports of both federal agencies such as OSHA and private companies that would like to benefit from machine learning predictions. As recommended by Baker et al. (2020), construction companies can benefit from machine learning models if they systematically record as much information as possible.

This method has another benefit in the data preparation process. While finding some pre-accident attributes (e.g., project type, cost, end-use) is simpler than detecting hazards as it does not require guessing (Baker et al. 2020), identifying other attributes such as the source/cause of an injury/accident can be challenging and subjective. Very similar accidents can be reported differently when inspected by different individuals. For instance, consider the following report from OSHA:

Report# 202479499 *"… Employee #1 removed a light bulb and cut wires while he was standing on the ladder. Employee #1 grabbed the top of a light fixture with his left* *hand and was trying to remove a bolt with pliers, when he felt an electric shock. This caused Employee #1 to fall from the ladder and he landed on the concrete floor below … sustained hip fracture and scalp laceration in the event…"*

One could list three potential sources for this case: the *electric part*, the *ladder*, or the *floor*. OSHA has considered the electric part as the main source in this case. However, the main event that led to the injury—as mentioned in the report—was a "fall to a lower level". BLS's OIICM guide explains that in these events, "name the equipment or part of the structure (structural element) from or through which the worker fell" as the source and "name the object or substance, if any, that contributed to the worker's fall" as the secondary source (Bureau of Labor Statistics, 2012, p. 110). By this definition, the *ladder* and not the *electric parts* would be the source of injury. Unlike these pre-accident attributes, codifying the outcomes of an accident is usually straightforward as the evidence is apparent and medical records are also available on those outcomes. It's hard to miss the nature of an injury (e.g., fracture, electric shock) or the injured part of the body and almost impossible to report the severity of an injury (i.e., fatality vs. non-fatality) incorrectly. Therefore, using the outcome of injuries in multi-label methods such as CC – which utilizes the correlations among labels - could reduce the noise in machine learning models. Noisy data is one of the main limitations in machine learning practices (Teng 1999, Kalapanidas et al. 2003, Gupta and Gupta 2019). Having a clean data can therefore significantly improve the prediction performance of ML models.

The second implication is on the way machine learning methods have been used in safety studies. Previous studies have introduced various techniques such as automating the creation of binary accident attributes (Tixier et al. 2016 a), automating accident report classification (Goh and Ubeynarayana 2017), Boruta feature selection (Poh et al. 2018), bagging and boosting algorithms with extensive hyper parameter optimizations (Tixier et al. 2016 b, Poh et al. 2018), and model stacking (Baker et al. 2020) to improve the accuracy of machine learning predictions. This study provides a novel approach that has not been tested in safety studies before: incorporating the correlations among accident outcomes in the training process using multi-label algorithms. This is a significant improvement over previous studies as it introduces a new approach to studying accidents and predicting their outcomes that can be generalized to real-world scenarios.

Lastly, the findings of this study can contribute to curtail the quantity and severity of occupational incidents by providing objective knowledge and making accidents more predictable. As discussed by Tixier et al. (2016 b), by identifying accident attributes safety managers can now utilize only one model to accurately predict not only the severity of an injury but also its type and the effected part of body. The safety manager can then use this actionable feedback to replace or remove those attributes before exposure. More specific but hazardous scenarios can also be discussed during safety meetings. As an example, if a task involves certain attributes that can cause a fall to lower levels, the model can simultaneously predict the part of body that is going to be affected most by this accident and its outcome (i.e., fall) to encourage discussions about proper PPE for the predicted body part. Other than the benefits of utilizing the correlations among accident outcomes to improve prediction accuracy, multi-label methods train only one model with multiple predictions instead of training multiple models and trying to combine their results. This consolidation of all outcome predictions into one multi-label model can facilitate the integration of machine learning techniques with safety planning practices.

5.6 Conclusions

The current study validated a multi-label approach to investigate the impact of correlations among accident outcomes on the performance of predictive models and offers potential benefits to both academia and industry. For academia, machine learning experiments that incorporate more than one accident outcome as the model's target will expand existing insight about the latent relationships among multiple accident consequences. The study findings also offer novel intuition into the significance of one accident outcome in predicting the value of other outcomes.

For the industry, this study illustrates the importance of collecting comprehensive data on the outcomes of occupational accidents in the construction industry. The findings will encourage safety managers and accident inspectors to carefully document the consequences of accidents as they can provide a better foundation for machine learning models.

Furthermore, safety practitioners can benefit from the method presented in this study to make safety predictions during the planning phases of the projects. The more accurate prediction of accident outcomes will result in better safety planning and prioritization of safety resources leading to well-informed decisions making which is ultimately the goal of all machine learning efforts in construction safety.

Despite the contribution of this study to construction safety, there are a few limitations that can be addressed in future studies. First, the data for this study were collected from OSHA's online database for around 1,800 accidents among three groups of specialty contractors. Future studies can apply the same methods to larger datasets from private companies to include more accident scenarios. Second, as mentioned in the literature review, the order of labels is important in the CC method, and therefore finding the right order in the chains is one of the main issues in validating the CC method. Due to the low dimension of the label space in this study (each experiment has 4 labels), considering all the permutations of labels (i.e., 24) in each experiment was manageable in this study. However, having more labels would introduce a combinatorial complexity in larger datasets (6 labels would produce 720 different chain orders). Therefore, future studies can incorporate variants of the CC method such as ensemble of chains of random orders (Read et al. 2011), heuristic on conditional label dependence (Zhang and Zhang 2010), and searching the order space given a fixed structure (Read et al. 2015) to address this limitation. Third, interpreting the order of accident outcomes in the best chain was not in the scope of this study. Future research can investigate the best chain order in each accident scenario and provide more insight regarding the relationships between accident outcomes.

In spite of these limitations, this research is proof of the concept that knowing the value of one accident outcome can help predict other outcomes and that multi-label algorithms such as classifier chains can directly benefit from these relationships to improve the performance of models to predict the consequences of accidents in future cases.

CHAPTER SIX: CONCLUSIONS

6.1 Research Summary

The primary outcome of this doctoral dissertation is a validated statistical and machine learning approach for modeling occupational accidents in the construction industry.

Through this systematic method, first the impact of different accident types and project characteristics were quantified by comparing their associated costs in chapter 2. The robust measures of location (trimmed mean) and scale (Winsorized variance) and hypothesis testing methods (the Welch-type procedure, an extension of Yuen's method, and percentile bootstrapping) are suited for comparing the cost information found in safety data since no assumptions about the distribution of samples are required. This is a substantial contribution to the body of safety knowledge since it provides analytical tools for researchers to compare cost of injuries accurately. The validated results revealed that, despite being less frequent than falls and struck by accidents, accident categories like caught in-between can impose a much larger cost to companies and the industry. Safety managers can use this insight to prioritize tasks where caught in-between accidents are more likely to occur. Furthermore, estimations of the expenses that should be paid for injuries in a project can be made more accurately using costs associated to various types of injuries using a probabilistic model. Contractors, especially those that are self-insured, can use this information, and prepare appropriate contingency plans.

In chapter 3, a series of statistical tests were used to assess the level of association between pre-accident attributes (e.g., source and cause of injury, project end-use and cost) and the most significant consequence of accidents (i.e., degree/severity of injury). The results revealed that, except for the project end-use, cost, and to a lower degree cause of injury, five accident elements had a major impact on the severity of an injury. Ordered by their Cramer's V values, nature of injury and part of body have the strongest correlations with injury degree, followed by source of injury, project type, and event type. Based on these statistically validated results, a multi-variate logistic regression model was developed in chapter 4 to predict the severity of injuries using solely pre-accident variables. A stepwise procedure was adopted to find the optimum combination of accident attributes to be included in the model. The proposed model was also validated using a variety of diagnostics metrics to ensure a good fit with the data. This model can be beneficial to safety managers in finding more hazardous accident patterns that are otherwise difficult to identify as they are built based on a variety of accident factors. These results, for instance, can assist safety practitioners in safety planning sessions to predict the type of injuries by knowing the hazards and sources of injuries that are involved in a certain task. This knowledge can then translate into practical policies that will increase workers' awareness before they begin their job on construction sites or establish proper protocols for using particular personal protective equipment related to a specific type of incident. Additionally, safety managers can use these findings to develop tailored training for more dangerous and expensive accident scenarios. The interpretability of statistical models like logistic regression can be useful for both researchers and practitioners, but it has two key drawbacks. First, as demonstrated in chapter 4, one must be extremely careful to examine all the assumptions related to these methods and use extended procedures to ensure the goodness of the model's fit to the data. Second, when there are nonlinear relationships between the predictors and the target variable, these models typically perform poorly when making predictions. Recent studies have used machine learning methods to enhance the prediction ability of models to close this gap.

Lastly, chapter 5 proposed a multi-label approach that can benefit from the relationships between accident outcomes during the model training phase in order to address the shortcomings of linear models in predicting the outcome of injuries and enhance the performance of single-label machine learning models. The fundamental contribution of this research is to verify the correlations among accident outcomes, such as nature and degree of injury, and to validate a modeling strategy that can benefit from such associations to enhance prediction performance. For safety practitioners, this means more accurate insight from historical accidents which translates into enhanced decision making under uncertainty, better resource allocations, optimized safety budgets, and eventually fewer occupational incidents on construction sites.

For organizations like OSHA, the techniques utilized in this study to prepare the data for analysis can be useful. Safety inspectors can improve their reporting exercises by using the systematic content analysis given in this work, and more precise databases can be created for future academics and practitioners.

Overall, the presented methods will lead to a data-driven and objective framework that can be applied to real-word safety practices on construction job sites including safety planning activities.

6.3 Limitations and Areas for Future Research

This study introduces various novel construction safety topics that offer intriguing directions for future study. For instance, future studies can incorporate more accurate cost data from private companies or hospitals into the proposed methods to gain more reliable estimations. Furthermore, considering more predictors—such as safety budgets, experience and education levels of the workers, protective measures, and environmental and human factors—could have provided insightful information about how such factors affect the cost of injuries.

As far as statistical tests on the correlations among accident factors and the severity of injuries and the linear modeling approaches, future studies can incorporate more variables such as "time of the accident" to develop models with better prediction performance. The severity of accidents also can be defined more accurately by considering more variables such as "monetary cost of injuries" or "days away from work" for non-fatal incidents.

The most important area for future research emerges from multi-label classification. First, the data for this study were collected from OSHA's online database for around 1,800 accidents among three groups of specialty contractors. Future studies can apply the same techniques on larger private company datasets to incorporate additional accident scenarios. Second, finding the proper order in the chains is one of the key challenges in verifying the classifier chains approach since the order of the labels is a crucial factor in the performance of these models. Having more labels in the accident data, even though would better characterize the effects of accidents, can quickly increase the combinatorial complexity needed to determine the ideal chain order. Therefore, to overcome this constraint, future studies can use different variants of the classifier chains method such as ensemble of chains of random orders (Read et al. 2011), heuristic on conditional label dependence (Zhang and Zhang 2010), and searching the order space given a fixed structure (Read et al. 2015). Third, interpreting the order of accident outcomes in the best chain was not in the scope of this study. Future research can investigate the best chain order in each accident scenario and provide more insight regarding the relationships between accident outcomes.

APPENDIX

Appendix 1

A configurable pipeline is designed to run all the modeling experiments in this study. The pipeline consists of four stages, one file to define utility functions, two configuration files, and a wrapper to automatically run each experiment.

By running the wrapper file, the pipeline starts with reading the specific configurations of an experiment along with converting the data file, containing 1,816 accident reports in the csv format, into a ‘pandas’ data frame that can be consumed by python functions. The data frame and configurations then go through four stages where in each stage they enter as inputs, go through some data manipulation functions, and being persisted as inputs for the next stage. Details of each stage is described below.

Stage 1: Data cleaning

The data cleaning consists of three simple tasks:

- Convert the column names to lower case
- Drop columns that cannot be used in a specific experiment
- Remove “,” from column values and replace single spaces with “_”

Stage 2: Creating binary features and targets

As shown in Table 5.2, all feature and target variables, except degree of injury, are categorical and need to be encoded as binary variables. This encoding is done through one hot encoding method (i.e., get_dummies function in pandas) in stage 2. While most of the accidents have resulted in one injury, there are 70 cases with multiple injured workers. To capture this, the cumulative values are calculated for binary features. The final step in this stage is removing infrequent nature of injuries such as ‘amputation’, ‘asphyxia’, ‘rupture’, etc.

Stage 3: Splitting

In this stage 40% of data is randomly selected for testing and the rest used to create three folds for training and validation. One should note the test proportion and the number of folds for validation are all configurable.

Stage 4: Modeling

After preparing the data and creating training and test sets in previous stages, the final stage is responsible for fitting the data into multi-label algorithms, optimizing their hyperparameters using the validation sets, and producing final predictions on the test set.

Once the final model for each experiment is trained, the pipeline persists the model as a pickle file along with all the final training and test data frames and performance metrics mentioned in the manuscript in csv format. Another script is written to compute the Shap values using the persisted models and data files for each experiment.

REFERENCES

Abdel-Aty, M. A., and Abdelwahab, H. T. (2004). “Predicting injury severity levels in traffic crashes: a modeling comparison”. *Journal of transportation engineering*, *130*(2), 204-210.

Abdelhamid, T.S., and Everett, J.G. (2000). “Identifying root causes of construction accidents”. *Journal of Construction Engineering and Management*, *126*, 52–60.

Abdelwahab, H. T., and Abdel-Aty, M. A. (2001). “Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections”. *Transportation Research Record*, *1746*(1), 6-13.

Abellán, J., López, G., and De OñA, J. (2013). “Analysis of traffic accident severity using decision rules via decision trees”. *Expert Systems with Applications*, *40*(15), 6047-6054.

Abudayyeh, O., Federicks, T., Palmquist, M., and Torres, H.N. (2003). “Analysis of Occupational Injuries and Fatalities in Electrical Contracting Industry”. *Journal of Construction Engineering and Management*, *129*, 152–158.

Albert, A., Hallowell, M. R., Kleiner, B., Chen, A., and Golparvar-Fard, M. (2014). “Enhancing construction hazard recognition with high-fidelity augmented virtuality”. *Journal of Construction Engineering and Management*, *140*(7), 04014024.

Alexander, R. A., and Govern, D. M. (1994). “A new and simpler approximation for ANOVA under variance heterogeneity”. *Journal of Educational Statistics*, *19*(2), 91-101.

Al-Ghamdi, A.S. (2002). “Using logistic regression to estimate the influence of accident factors on accident severity”. *Accident Analysis & Prevention*, *34*, 729–741.

Algina, J., and Olejnik, S. F. (1984). “Implementing the Welch-James procedure with factorial designs”. *Educational and Psychological Measurement*, *44*(1), 39-48.

Amiri, M., Ardeshira, A., Zarandi MH, F., and 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*, 264–276.

Anders, S., and Huber, W. (2010). “Differential expression analysis for sequence count data.” *Genome biology*, *11*(10), R106.

Argilés-Bosch, J.M., Martí, J., Monllau, T., Garcia-Blandón, J., and Urgell, T. (2014). “Empirical analysis of the incidence of accidents in the workplace on firms’ financial performance.” *Safety Science*. *70*, 123–132.

Ayhan, B. U., and Tokdemir, O. B. (2019). “Predicting the outcome of construction incidents”. *Safety Science*, *113*, 91-104.

Baker, H., Hallowell, M. R., and Tixier, A. J. P. (2020). “AI-based prediction of independent construction safety outcomes from universal attributes”. *Automation in Construction*, *118*, 103146.

Bakker, M., and Wicherts, J.M. (2011). “The (mis)reporting of statistical results in psychology journals”. *Behavior research methods*, 43, 666–678.

Barkan, R., Zohar, D., and Erev, I. (1998). “Accidents and decision making under uncertainty: A comparison of four models”. *Organizational behavior and human decision processes*, *74*, 118–144.

Behm, M., and Schneller, A. (2013). “Application of the Loughborough Construction Accident Causation model: A framework for organizational learning”. *Construction Management and Economics*, *31*, 580–595.

Benjamini, Y., and Hochberg, Y. (1995). “Controlling the false discovery rate: a practical and powerful approach to multiple testing.” *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Berk, R.A. (2008). *Statistical Learning from a Regression Perspective*. Springer: New York, NY, USA.

Berry, D. A. (1987). “Logarithmic transformations in ANOVA.” *Biometrics*, 439-456.

Bellamy, L.J., Ale, B.J.M., Geyer, T.A.W., Goossens, L.H.J., Hale, A.R., Oh, J., Mud, M., Bloemhof, A., Papazoglou, I.A., and Whiston, J.Y. (2007). “Storybuilder-A tool for the analysis of accident reports”. *Reliability Engineering & System Safety*, *92*, 735–744.

Bonafede, M., Espindle, D., and Bower, A. G. (2013). “The direct and indirect costs of long bone fractures in a working age US population.” *Journal of medical economics*, *16*(1), 169-178.

Bondy, J., Lipscomb, H., Guarini, K., and Glazner, J. E. (2005). “Methods for using narrative text from injury reports to identify factors contributing to construction injury.” *American journal of industrial medicine, 48(5)*, 373-380.

Boutell, M. R., Luo, J., Shen, X., and Brown, C. M. (2004). "Learning multi-label scene classification". *Pattern recognition*, *37*(9), 1757-1771.

Box, G. E. (1954). "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification." *The annals of mathematical statistics*, 290-302.

Breiman, L. (2001). "Random forests". *Machine learning*, *45*(1), 5-32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Chapman and Hall: Wadsworth, NY, USA.

Briggs, F., Fern, X. Z., and Irvine, J. (2013). "Multi-label classifier chains for bird sound". *arXiv preprint arXiv:1304.5862*.

Brunner, E., Dette, H., and Munk, A. (1997). "Box-type approximations in nonparametric factorial designs." *Journal of the American Statistical Association*, *92(*440), 1494-1502.

Buckley, C., and Voorhees, E. M. (2017, August). Evaluating evaluation measure stability. In *ACM SIGIR Forum* (Vol. 51, No. 2, pp. 235-242). New York, NY, USA: ACM.

Bureau of Labor Statistics. (2012). Occupational injury and illness classification manual. < https://www.bls.gov/iif/oiics_manual_2010.pdf>, (accessed on March 5, 2019)

Bureau of Labor Statistics. (2019). Census of fatal occupational injuries. < https://www.bls.gov/iif/oshcfoi1.htm> , (accessed on March 12, 2019)

Bureau of Labor Statistics, U.S. Department of Labor, Survey of Occupational Injuries and Illnesses (2021). https://www.bls.gov/data/ (accessed on March 5, 2021).

Bursac, Z., Gauss, C.H., Williams, D.K., and Hosmer, D.W. (2008). "Purposeful selection of variables in logistic regression". *Source code for biology and medicine. 3*, 17.

Campbell, J.C., Webster, D., Koziol-McLain, J., Block, C., Campbell, D., Curry, M.A., Gary, F., Glass, N., McFarlane, J., Sachs, C., Glass, N., McFarlane, J., Sachs, C. and Sharps, P. (2003). "Risk factors for femicide in abusive relationships: Results from a multisite case control study". *American Journal of Public Healt*, *93*, 1089–1097.

Castillo-Rosa, J., Suárez-Cebador, M., Rubio-Romero, J.C., and Aguado, J.A. (2017). "Personal factors and consequences of electrical occupational accidents in the primary, secondary and tertiary sectors". *Safety science, 91*, 286–297.

Cawley, J.C., and Homce, G.T. (2006). Trends in electrical injury, 1992–2002 *IEEE Transactions on Industry Applications*, *4*(44), 962–972.

Ceylan, Z., and Pekel, E. (2017). "Comparison of multi-label classification methods for prediagnosis of cervical cancer". *Graph Models*, *21*, 22.

Chang, L.Y., and Chen, W.C. (2005). "Data mining of tree-based models to analyze freeway accident frequency". *Journal of safety research*, *36*, 365–375.

Chang, L. Y., and Wang, H. W. (2006). "Analysis of traffic injury severity: An application of non-parametric classification tree techniques". *Accident Analysis & Prevention*, *38*(5), 1019-1027.

Charte, F., Rivera, A. J., del Jesus, M. J., and Herrera, F. (2015, September). QUINTA: A question tagging assistant to improve the answering ratio in electronic forums. In *Ieee eurocon 2015-international conference on computer as a tool (eurocon)* (pp. 1-6). IEEE.

Chau, N., Mur, J.M., Benamghar, L., Siegfried, C., Dangelzer, J.L., Francais, M., Jacquin, R., and Sourdot, A. (2002). "Relationships between some individual characteristics and occupational accidents in the construction industry". *Journal of occupational health*, *44*, 131–139.

Chau, N., Mur, J.M.; Benamghar, L., Siegfried, C., Dangelzer, J.L., Français, M., Jacquin, R., and Sourdot, A. (2004). "Relationships between certain individual characteristics and occupational injuries for various jobs in the construction industry: A case-control study". *American journal of industrial medicine, 45*, 84–92.

Chen, G.-X., and Fosbroke, D.E. (1998). "Work-Related Fatal-Injury Risk of Construction Workers by Occupation and Cause of Death". *Human and Ecological Risk Assessment*, *4*, 1371–1390.

Chen, L., Vallmuur, K., and Nayak, R. (2015). "Injury narrative text classification using factorization model". *BMC medical informatics and decision making*, *15*(1), 1-12.

Cheng, M. Y., Kusoemo, D., and Gosno, R. A. (2020). "Text mining-based construction site accident classification using hybrid supervised machine learning". *Automation in Construction*, *118*, 103265.

Cheng, C. W., Leu, S. S., Lin, C. C., and Fan, C. (2010). "Characteristic analysis of occupational accidents at small construction enterprises." *Safety Science*, *48*(6), 698-707.

Cheng, C.-W., Leu, S.-S., Cheng, Y.-M., Wu, T.-C., and Lin, C.-C. (2012). "Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry". *Accident Analysis & Prevention*, 48, 214–222.

Chi, N. W., Lin, K. Y., El-Gohary, N., and Hsieh, S. H. (2016). "Evaluating the strength of text classification categories for supporting construction field inspection". *Automation in Construction*, *64*, 78-88.

Cheng, C., Lin, C., and Leu, S. (2010). "Use of association rules to explore cause–effect relationships in occupational accidents in the Taiwan construction industry". *Safety Science*, *48*, 436–444.

Chi, C.-F., Chang, T.-C., and Hung, K.-H. (2004). "Significant industry—Source of injury—Accident type for occupational fatalities in Taiwan". *International Journal of Industrial Ergonomics*, *34*, 77–91.

Chi, C.F., Chang, T.C., and Ting, H.I. (2005). "Accident patterns and prevention measures for fatal occupational falls in the construction industry". *Applied ergonomics, 36*, 391–400.

Chi, C.F., and Lin, S.Z. (2018). "Classification scheme and prevention measures for caught-in-between occupational fatalities". *Applied ergonomics*, *68*, 338–348.

Chi, C.-F., Lin, Y.-Y., and Ikhwan, M. (2012). "Flow diagram analysis of electrical fatalities in construction industry". *Safety Science*, 50, 1205–1214.

Chi, C.-F., Yang, C.-C., and Chen, Z.-L. (2009). "In-depth accident analysis of electrical fatalities in the construction industry". *International Journal of Industrial Ergonomics*, 39, 635–644.

Choi, S.D. (2015). "Aging Workers and Trade-Related Injuries in the US Construction Industry". *Safety and health at work*, 6, 151–155.

Choi, J., Gu, B., Chin, S., and Lee, J.-S. (2020). "Machine learning predictive model based on national data for fatal accidents of construction workers". *Automation in Construction, 110*, 102974.

Chong, M., Abraham, A., and Paprzycki, M. (2005). "Traffic accident analysis using machine learning paradigms". *Informatica*, *29*(1).

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

CPWR--The Center for Construction Research, (2002). *The Construction Chart Book, The U.S. Construction Industry and its Workers*, 3rd ed. Cpwr-The Center for Construction Research and Training, Silver Spring, MD.

Cramér, H. (1946). *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA

Cribbie, R. A., Fiksenbaum, L., Keselman, H. J., and Wilcox, R. R. (2012). "Effect of non-normality on test statistics for one-way independent groups designs." *British Journal of Mathematical and Statistical Psychology, 65*(1), 56-73.

Cribbie, R. A., Wilcox, R. R., Bewell, C., and Keselman, H. J. (2007). "Tests for treatment group equality when data are nonnormal and heteroscedastic." *Journal of Modern Applied Statistical Methods*, *6*(1), 12.

Daniels, S., Brijs, T., Nuyts, E., and Wets, G. (2010) "Externality of risk and crash severity at roundabouts". *Accident Analysis & Prevention, 42*, 1966–1973.

Dash, M., and Liu, H. (1997). "Feature selection for classification". *Intelligent data analysis*, *1*(1-4), 131-156.

DataUSA. Available online https://datausa.io/profile/naics/construction (accessed on 5 May 2019).

Davis, M. J. (2010). "Contrast coding in multiple regression analysis: Strengths, weaknesses, and utility of popular coding structures." *Journal of Data Science*, 8(1), 61-73.

Delen, D., Sharda, R., and Bessonov, M. (2006). "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks". *Accident Analysis & Prevention*, *38*(3), 434-444.

Dement J. M., and Lipscomb H. (1999). "Workers compensation experience of North Carolina residential construction workers, 1986–1994." *Applied occupational and environmental hygiene, 14*(2), 97–106.

Desvignes, M. (2014). *Requisite empirical risk data for integration of safety with advanced technologies and intelligent systems* (Doctoral dissertation, University of Colorado at Boulder).

Dien, J., and Santuzzi, A. M. (2004). "Application of Repeated Measures ANOVA to High-Density ERP." *Event-related potentials: A methods handbook,* 57-81.

Dunkler, D., Plischke, M., Leffondré, K., and Heinze, G. (2014). "Augmented backward elimination: A pragmatic and purposeful way to develop statistical models". *PLoS ONE, 9*, e113677.

Dunnett, C. W. (1980). "Pairwise multiple comparisons in the unequal variance case." *Journal of the American Statistical Association*, *75*(372), 796-800.

Durbin, D. (1993). “The Cost of Treating Injured Workers: The Changing Workers' Compensation Landscape”. *Benefits Quarterly, 9* (4), 9-21.

Efron, B., and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Elassad, Z.E.A., Mousannif, H., Al Moatassime, H., and Karkouch, A. (2020). “The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review”. *Engineering Applications of Artificial Intelligence, 87,* 103312.

Esmaeili, B. (2012). *Identifying and Quantifying Construction Safety Risks at the Attribute Level*. Ph.D. Thesis, University of Colorado Boulder, Boulder, CO, USA, 2012.

Esmaeili, B., and Hallowell, M.R. (2012). “Diffusion of Safety Innovations in the Construction Industry”. *Journal of Construction Engineering and Management, 138,* 955–963.

Esmaeili, B., Hallowell, M. R., and Rajagopalan, B. (2015a). “Attribute-Based Safety Risk Assessment I: Analysis at the Fundamental Level.” ASCE, *Journal of Construction Engineering and Management*, *141*(8), 04015021.

Esmaeili, B., Hallowell, and M. R., Rajagopalan, B. (2015b). “Attribute-Based Safety Risk Assessment II: Predicting Safety Outcomes Using Generalized Linear Models.” ASCE, *Journal of Construction Engineering and Management*, *141*(8), 04015022.

Gholizadeh, P., and Esmaeili, B. (2020 a). “Cost of occupational incidents for electrical contractors: Comparison using robust-factorial analysis of variance”. *Journal of Construction Engineering and Management*, *146*(7), 04020073.

Gholizadeh, P., and Esmaeili, B. (2020 b). “Developing a multi-variate logistic regression model to analyze accident scenarios: case of electrical contractors”. *International journal of environmental research and public health*, *17*(13), 4852.

Gholizadeh, P., and Esmaeili, B. (2016). Applying classification trees to analyze electrical contractors' accidents. In *Construction Research Congress 2016: Old and New Construction Technologies Converge in Historic San Juan*, *CRC 2016* (pp. 2699-2708). American Society of Civil Engineers (ASCE).

Gholizadeh, P., Esmaeili, B., and Memarian, B. (2018, April). Evaluating the performance of machine learning algorithms on construction accidents: An application of ROC curves. In *Construction Research Congress 2018* (pp. 8-18).

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.

EUROSTAT. (2013). *European Statistics on Accident at Work ESAW*. *Summary Methodology*; Eurostat Methodologies & Working Papers; EUROSTAT: Luxembourg, Luxembourg.

Everett J. G., and Frank, P. B. (1996). “Costs of accidents and injuries to the construction industry.” *Journal of Construction Engineering and Management*, 122(2), 158–164.

Fan, C., Oh, D.S., Wessels, L., Weigelt, B., Nuyten, D.S., Nobel, A.B., Veer, L.J.V., and Perou, C.M. (2006). “Concordance among Gene-Expression–Based Predictors for Breast Cancer”. *New England Journal of Medicine*, *355*, 560–569.

Feng, Y., Zhang, S., and Wu, P., (2015). “Factors influencing workplace accident costs of building projects.” *Safety Science*. 72, 97–104.

Field, A., Miles, J., and Field, Z., (2012). *Discovering statistics using R*., Thousand Oaks, CA: Sage

Fordyce, T. A., Kelsh, M., Lu, E. T., Sahl, J. D., and Yager, J. W. (2007). “Thermal burn and electrical injuries among electric utility workers, 1995–2004.” *Burns*, *33*(2), 209-220.

Fordyce, T.A., Leonhard, M.J., Watson, H.N., Mezei, G., Vergara, X.P., and Krishen, L. (2016). “An analysis of fatal and non-fatal injuries and injury severity factors among electric power industry workers”. *American journal of industrial medicine*, *59*, 948–958.

Fordyce, T.A., Morimoto, L., Coalson, J., Kelsh, M.A., and Mezei, G. (2010). “Neck Injuries Among Electric Utility Workers, 1995–2007”. *ournal of occupational and environmental medicine*, 52, 441–449.

Fox, J., and Weisberg, S. (2011) *An R Companion to Applied Regression*., Thousand Oaks, CA: Sage

Freitas, A.A. (2000). “Understanding the crucial differences between classification and discovery of association rules: A position paper”. *AcM sIGKDD Explorations Newsletter*, *2*, 65–69.

Gholizadeh, P., and Esmaeili, B. (2015). Electrical Contractors’ Safety Risk Management: An Attribute-Based Analysis. In *Proceedings of the ASCE International Workshop on Computing in Civil Engineering*, Austin, TX, USA.

Gholizadeh, P., and Esmaeili, B., (2016). Applying Classification Trees to Analyze Electrical Contractors’ Accidents. In *Construction Research Congress (CRC 2016)*, Puerto Rico.

Gholizadeh, P., Esmaeili, B., and Memarian, B. (2018). Evaluating the Performance of Machine Learning Algorithms on Construction Accidents: An Application of ROC Curves. In *Proceedings of the Construction Research Congress (CRC 2018)*, ASCE, New Orleans, LA, USA.

Gholizadeh, P., and Esmaeili, B. (2020 a) “Cost of Occupational Incidents for Electrical Contractors: Comparison Using Robust-Factorial Analysis of Variance”. *Journal of Construction Engineering and Management*, *146*, 04020073.

Gholizadeh, P., and Esmaeili, B. (2020 b). “Developing a multi-variate logistic regression model to analyze accident scenarios: Case of electrical contractors”. *International journal of environmental research and public health*, *17*, 4852.

Goetsch, D.L., (2013). *Construction Safety and Health*, (second ed.), New Jersey: Pearson Education, Inc.

Goh, Y. M., and Ubeynarayana, C. U. (2017). “Construction accident narrative classification: An evaluation of text mining techniques”. *Accident Analysis & Prevention*, *108,* 122-130.

Gopal, S., and Yang, Y. (2010, July). Multilabel classification with meta-level features. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 315-322).

Guo, J., and Geng, Z. (1995). “Collapsibility of logistic regression coefficients”. *Journal of the Royal Statistical Society. Series B (Methodological)*, 263-267.

Gupta, S., and Gupta, A. (2019). “Dealing with noise problem in machine learning data-sets: A systematic review”. *Procedia Computer Science*, *161*, 466-474.

Haddon Jr, W. (1968). “The changing approach to the epidemiology, prevention, and amelioration of trauma: the transition to approaches etiologically rather than descriptively based.” *American journal of public health and the Nations health, 58(8),* 1431-1438.

Haddon Jr, W. (1972). “A logical framework for categorizing highway safety phenomena and activity.” *Journal of Trauma and Acute Care Surgery, 12(3)*, 193-207.

Hamid, A.R.A., Majid MZ, A., and Singh, B. (2008). “Causes of accidents at construction sites”. *Malaysian journal of civil engineering*, *20*, 242–259.

Harb, R., Radwan, E., Yan, X., Pande, A., and Abdel-Aty, M. (2008). “Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression”. *Journal of Transportation Engineering, 134*, 203–214.

Hasanzadeh, S., Esmaeili, B., and Dodd, M. D. (2017). “Measuring the impacts of safety knowledge on construction workers’ attentional allocation and hazard detection using remote eye-tracking technology”. *Journal of management in engineering*, *33*(5), 04017024.

Hayes, J., and Maslen, S. (2015). “Knowing stories that matter: learning for effective safety decision-making”. *Journal of Risk Research*, *18*(6), 714-726.

Heinrich, N. W. (1931). *Industrial accident prevention*., New York, NY: McGraw-Hill. Book Co.

Heinrich, H.W. (1941). *Industrial Accident Prevention: A Scientific Approach*, 2nd ed.; Wiley: New York, NY, USA.

Herrera, F., Charte, F., Rivera, A. J., and Del Jesus, M. J. (2016). “Multilabel classification. In Multilabel Classification” (pp. 17-31). Springer, Cham.

Hesla, E. (2009). Improvements in safe working practices near overhead lines. In *Conference Record 2009, Proceedings of the IEEE Industrial & Commercial Power Systems Technical Conference*, Calgary, AB, Canada, 3–7 May 2009.

Higgins, T.L., Estafanous, F.G., Loop, F.D., Beck, G.J., Blum, J.M., and Paranandi, L. (1992). “Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients: A clinical severity score”. *JAMA, 267*, 2344–2348.

Hinze, J. (1997). *Construction safety*, Englewood Cliffs, NJ: Prentice-Hall.

Hinze, J., and Appelgate, L. (1991). “Costs of construction injuries.” *Journal of Construction Engineering and Management*, *117*(3), 537–550.

Hinze, J., Devenport, J. N., and Giang, G. (2006). “Analysis of construction worker injuries that do not result in lost time.” *Journal of Construction Engineering and Management*, *132*(3), 321–326.

Hinze, J., and Gambatese, J. (2003). “Factors that influence safety performance of specialty contractors.” *Journal of construction engineering and management, 129(2)*, 159-164.

Health and Safety Executive (1993) *The costs of accidents at work*, London, UK: HMSO.

Hochberg, Y. (1988). “A sharper Bonferroni procedure for multiple tests of significance.” *Biometrika*, *75*(4), 800-802.

Hoła, B., and Szóstak, M. (2019). “Modeling of the Accidentality Phenomenon in the Construction Industry”. *Applied Sciences*, *9*, 1878.

Homce, G.T., Cawley, J.C., and Yenchek, M.R. (2008). *A Performance Evaluation of Two Overhead Power Line Proximity Warning Devices*. Information Circular 9510. DHHS (NIOSH) Publication No. 2009–110; National Institute for Occupational Safety and Health: Atlanta, GA, USA.

Hosmer, D.W., and Lemesbow, S. (1980) “Goodness of fit tests for the multiple logistic regression model”. *Communications in statistics-Theory and Methods*, *9*, 1043–1069.

Hosmer, D.W., Lemeshow, S., and Sturdivant, R.X. *Applied Logistic Regression*; 3rd ed.; John Wiley & Sons: New York, NY, USA, 2013.

Huang, H., Chin, H.C., and Haque, M.M. (2008). “Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis”. *Accident Analysis & Prevention*, *40*, 45–54.

Huang, C., Trabelsi, A., Qin, X., Farruque, N., Mou, L., and Zaiane, O. R. (2021, June). Seq2emo: A sequence to multi-label emotion classification model. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4717-4724).

Huber, P. J., (1981). *Robust statistics*. New York, NY: Wiley

Ibrohim, M. O., and Budi, I. (2019, August). Multi-label hate speech and abusive language detection in Indonesian twitter. In *Proceedings of the Third Workshop on Abusive Language Online* (pp. 46-57).

Iezzoni, L.I. *Risk Adjustment for Measuring Health Care Outcomes*; Health Administration Press: Chicago, IL, USA, 2003; Volume 3.

Iranitalab, A., and Khattak, A. (2017). “Comparison of four statistical and machine learning methods for crash severity prediction”. *Accident Analysis & Prevention, 108*, 27-36.

Islam, S. S., Biswas, R. S., Nambiar, A. M., Syamlal, G., Velilla, A. M., Ducatman, A. M., and Doyle, E. J., (2001). “Incidence and risk of work-related fracture injuries: experience of a state-managed workers’ compensation system.” *Journal of occupational and environmental medicine*, *43*(2), 140–146.

Jallon, R., Imbeau, D., and de Marcellis-Warin, N. (2011a). “Development of an indirect-cost calculation model suitable for workplace use.” *Journal of Safety Research*, *42*(3), 149-164.

Jallon, R., Imbeau, D., and de Marcellis-Warin, N. (2011b). "A process mapping model for calculating indirect costs of workplace accidents." *Journal of safety research*, *42*(5), 333-344.

James, G. S. (1951). "The comparison of several groups of observations when the ratios of the population variances are unknown." *Biometrika*, *38*(3/4), 324-329.

Janicak, C.A. (2008). "Occupational fatalities due to electrocutions in the construction industry". *Journal of safety research*, 39, 617–621.

Jenkins, E.L., Kisner, S.M., Fosbroke, D.E., Layne, L.A., Stout, N.A., Castillo, D.N., Cutlip, C.P., and Cianfrocco, R. (1993). *Fatal Injuries to Workers in the United States, 1980–1989: A Decade of Surveillance: National Profile*, National Institute for Occupational Safety and Health: Cincinnati, OH, USA.

Johansen, S. (1980). "The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression." *Biometrika*, *67*(1), 85-92.

John, G. H., Kohavi, R., and Pfleger, K. (1994). Irrelevant features and the subset selection problem. In *Machine learning proceedings 1994* (pp. 121-129).

Kalapanidas, E., Avouris, N., Craciun, M., and Neagu, D. (2003, November). Machine learning algorithms: a study on noise sensitivity. In *Proc. 1st Balcan Conference in Informatics* (pp. 356-365).

Kane F. (1996). "Maine 200-type programs spread throughout the U.S." *Job Safety Health*, 7(3), 9-15.

Kang, Y., Siddiqui, S., Suk, S. J., Chi, S., and Kim, C. (2017). "Trends of fall accidents in the US construction industry". *Journal of Construction Engineering and Management*, *143*(8), 04017043.

Kassambara, A. (2018) *Machine Learning Essentials: Practical Guide in R*; STHDA: Marseille, France, Available online: http://www.sthda.com/english/articles/35-statistical-machine-learning-essentials/141-cart-model-decision-tree-essentials (accessed on 16 November 2020).

Kenny, D. A., Kashy, D. A., and Bolger, N. (1998). Data analysis in social psychology. *In Handbook of social psychology, 4th ed., Vol. 1 and 2*, (pp. 233-265). Boston, MA: McGraw-Hill

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., and Levin, J. R. (1998). "Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses." *Review of educational research*, *68*(3), 350-386.

Keselman, H. J., Wilcox, R. R., Othman, A. R., and Fradette, K. (2002). "Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and nonnormality." *Journal of Modern Applied Statistical Methods, 1*(2), 38.

Kines, P., Spangenberg, S., and Dyreborg, J. (2007). "Prioritizing occupational injury prevention in the construction industry: Injury severity or absence?" *Journal of Safety Research*, *38*, 53–58.

Kisner, S., and Casini, V. (1998). Epidemiology of electrocution fatalities. In *Worker Deaths by Electrocution: A Summary of NIOSH Surveillance and Investigative Findings*; Department of Health and Human Services (NIOSH): Washington, DC, USA pp. 9–19.

Kisner, S.M., and Fosbroke, D.E. (1994). "Injury Hazards in the Construction Industry". *Journal of Occupational Medicine*, *36*, 137–143.

Kramer, A.A., and Zimmerman, J.E. (2007). "Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited". *Critical care medicine*, *35,* 2052–2056.

Krishnamoorthy, K., Lu, F., and Mathew, T. (2007). "A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models." *Computational Statistics & Data Analysis*, *51*(12), 5731-5742.

Kuhn, M. (2015). "Caret: Classification and Regression Training". *Astrophysics Source Code Library*, ascl-1505

LaBelle, J. E. (2000). "What do accidents truly cost?" *Professional Safety*, *45*(4), 38.

Lattimore, P.K., and Visher, C.A. (2014). "The impact of prison reentry services on short-term outcomes: Evidence from a multisite evaluation". *Evaluation review, 37*, 274–313.

Lee, S., and Ahn, C. H. (2003). "Modified ANOVA for unequal variances." *Communications in Statistics-Simulation and Computation*, *32*(4), 987-1004.

Lee, H.-S., Kim, H., Park, M., Teo, E.A.L., and Lee, K.-P. (2012). "Construction Risk Assessment Using Site Influence Factors". *Journal of computing in civil engineering, 26*, 319–330.

Leigh, J. P., Markowitz, S. B., Fahs, M., and Landrigan, P. J. (2000). *Costs of occupational injuries and illnesses.*, Michigan: University of Michigan Press.

Leigh, J. P., Markowitz, S. B., Fahs, M., Shin, C., and Landrigan, P. J. (1997). “Occupational injury and illness in the United States: estimates of costs, morbidity, and mortality.” *Archives of Internal Medicine, 157*(14), 1557-1568.

Leigh, J. P., Waehrer, G., Miller, T., and Macurdy, S. (2006). “Costs differences across demographic groups and types of occupational injuries and illnesses.” *American journal of industrial medicine, 49*(10), 845–853.

Lemeshow, S., and Hosmer, D.W., Jr. (1982). “A review of goodness of fit statistics for use in the development of logistic regression models”. *American journal of epidemiology, 115*, 92–106.

Leopold, E., and Leonard, S. (1987). “Costs of construction accidents to employers.” *Journal of Occupational Accidents*, *8*(4), 273-294.

Lesaffre, E., and Albert, A. (1989). “Multiple-group logistic regression diagnostics”. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *38*(3), 425–440.

Li, T., and Ogihara, M. (2006). “Toward intelligent music information retrieval”. *IEEE Transactions on Multimedia*, *8*(3), 564-574.

Li, Y., Song, Y., and Luo, J. (2017). Improving pairwise ranking for multi-label image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3617-3625).

Liao, C.-W., and Perng, Y.-H. (2008). “Data mining for occupational injuries in the Taiwan construction industry”. *Safety Science.* **2008**, *46*, 1091–1102.

Liberty Mutual., (2018). Workplace safety index. *Liberty Mutual Research Institute for Safety*.

Lipscomb H. J., Dement J. M., and Behlman R. (2003). “Direct costs and patterns of injuries among residential carpenters, 1995–2000.” *Journal of occupational and environmental medicine*, *45*(8), 875–880.

Lipscomb H. J., Glazner J. E., Bondy J, Guarini K, and Lezotte D. (2006). “Injuries from slips and trips in construction.” *Applied ergonomics*, *37*(3), 267–274.

Lix, L. M., and Keselman, H. J. (1998). “To Trim or Not to Trim: Tests of Location Equality Under Heteroscedasticity and Nonnormality.” *Educational and Psychological Measurement*, *58*(3), 409-429.

Long, J. S., and Ervin, L. H. (2000). “Using heteroscedasticity consistent standard errors in the linear regression model.” *The American Statistician*, *54*(3), 217-224.

Loomis, D., Dufort, V., Kleckner, R.C., and Savitz, D.A. (1999). “Fatal occupational injuries among electric power company workers”. *American journal of industrial medicine*, 35, 302–309.

Luh, W. M., and Guo, J. H. (2001). “Using Johnson's transformation and robust estimators with heteroscedastic test statistics: An examination of the effects of non-normality and heterogeneity in the non-orthogonal two-way ANOVA design.” *British Journal of Mathematical and Statistical Psychology*, *54*(1), 79-94.

Lundberg, S. M., and Lee, S. I. (2017, December). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768-4777).

Ma, W. (2018). *Study of Some Issues of Goodness-of-Fit Tests for Logistic Regression*. Ph.D. Thesis, Florida State Univaersity, Tallahassee, FL, USA.

Manu, P.A., Ankrah, N.A., Proverbs, D.G., and Suresh, S. (2012). “Investigating the multi-causal and complex nature of the accident causal influence of construction project features”. *Accident Analysis & Prevention*, *48*, 126–133.

Manuele, F.A. *Heinrich Revisited: Truisms or Myths*; National Safety Council: Itasca, IL, USA, 2002.

MacKenzie, E. J., Castillo, R. C., Jones, A. S., Bosse, M. J., Kellam, J. F., Pollak, A. N., Webb, L. X., Swiontkowski, M. F., Smith, D. G., Sanders, R. W., Jones, A. L., Starr, A. J., McAndrew, M. P., Patterson, B. M., and Burgess, A. R., (2007). “Health-care costs associated with amputation or reconstruction of a limb-threatening injury.” *JBJS*, *89*(8), 1685-1692.

Manuele, F. A. (2011). “Accident Costs: Rethinking ratios of indirect to direct costs.” *Professional Safety*, *56*(01), 39-47.

Marhavilas, P. and Koulouriotis, D. (2012). “A combined usage of stochastic and quantitative risk assessment methods in the worksites: Application on an electric power provider”. *Reliability Engineering & System Safety*, *97*, 36–46.

Marhavilas, P.K., Koulouriotis, D.E., and Mitrakas, C. (2011). “On the development of a new hybrid risk assessment process using occupational accidents’ data: Application on the Greek Public Electric Power Provider”. *Journal of loss prevention in the process industries*, *24*, 671–687.

McCallum, A. K. (1999). Multi-label text classification with a mixture model trained by EM. In *AAAI 99 workshop on text learning*.

McCann, M., Hunting, K.L., Murawski, J., Chowdhury, R., and Welch, L. (2003). “Causes of electrical deaths and injuries among construction workers”. *American journal of industrial medicine*, *43*, 398–406.

McHugh, M.L. (2012). “Interrater reliability: The kappa statistic *Biochemia medica*, *22*, 276–282.

McHugh, M.L. (2013). “The Chi-square test of independence”. *Biochemia medica*, *23*, 143–149.

Miller, T. R. (1997). “Estimating the costs of injury to US employers.” *Journal of Safety Research*, *28*(1), 1-13.

Miller T. R., and Galbraith M. (1995). “Costs of occupational injury in the U.S.” *Accident Analysis Prevent*. *27*(6), 741–747.

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.

Mitropoulos, P., Abdelhamid, T.S., and Howell, G.A. (2005). “Systems Model of Construction Accident Causation”. *Journal of construction engineering and management*, *131*, 816–825, doi:10.1061/(ASCE)0733-9364(2005)131:7(816).

Mohamed, R., Perumal, T., Sulaiman, M., Mustapha, N., and Zainudin, M. N. (2018). “Multi label classification on multi resident in smart home using classifier chains”. *Advanced Science Letters*, *24*(2), 1316-1319.

Monnery, N. (1998). “The costs of accidents and work-related ill-health to a cheque clearing department of a financial services organisation.” *Safety science*, *31*(1), 59-69.

Montella, A., Aria, M., D'Ambrosio, A., and Mauriello, F. (2011). “Data-mining techniques for exploratory analysis of pedestrian crashes”. *Transportation research record*, *2237*(1), 107-116.

Montgomery, D. C., (2001). *Design and analysis of experiments* (5th ed.). New York, NY: John Wiley.

Moudon, A.V., Lin, L., Jiao, J., Hurvitz, P., and Reeves, P. (2011). “The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington”. *Accident Analysis & Prevention*, *43*, 11–24.

Moyano, J. M., Gibaja, E. L., Cios, K. J., and Ventura, S. (2018). “Review of ensembles of multi-label classifiers: models, experimental study and prospects”. *Information Fusion*, *44*, 33-45.

Mussone, L., Ferrari, A., and Oneta, M. (1999). “An analysis of urban collisions using an artificial intelligence model”. *Accident Analysis & Prevention*, *31*(6), 705-718.

Mustapha, F., and Naoum, S. (1998). “Factors influencing the effectiveness of construction site managers”. *International journal of project management*, *16*, 1–8.

Narayan, K.V., Boyle, J.P., Thompson, T.J., Sorensen, S.W., and Williamson, D.F. (2003). “Lifetime risk for diabetes mellitus in the United States”. *JAMA*, *290*, 1884–1890.

National Safety Council, National Safety Council. Research, and Statistics Dept. (1999). *Injury (accident) Facts*, Itasca, IL: National Safety Council Press.

Neitzel, R.L., Seixas, N.S., and Ren, K.K. (2001). “A review of crane safety in the construction industry”. *Applied occupational and environmental hygiene*, *16*, 1106–1117.

Noble, W. S. (2006). “What is a support vector machine?”. *Nature biotechnology*, *24*(12), 1565-1567.

Office of Management and Budget, (2017), North American Industry Classification System, < https://www.census.gov/eos/www/naics/2017NAICS/2017_NAICS_Manual.pdf>, (March 3, 2019)

Ore, T., and Casini, V. (1996). “Electrical fatalities among US construction workers”. *Journal of occupational and environmental medicine*, 38, 587–592

Osborne, J. W. (2010). “Improving your data transformations: Applying the Box-Cox transformation.” *Practical Assessment, Research & Evaluation*, *15*(12), 2.

OSHA IMIS, (2016), Fatality and Catastrophe Investigation Summaries, https://www.osha.gov/pls/imis/accidentsearch.html, Accessed 2016

Ospina-Mateus, H., Jiménez, L.A.Q., Lopez-Valdes, F.J., Garcia, S.B., Barrero, L.H., Sana, S.S. (2021). Extraction of decision rules using genetic algorithms and simulated annealing for prediction of severity of traffic accidents by motorcyclists. *Journal of ambient intelligence and humanized computing*, 12.

Paez, O., Uahinui, T., Genaidy, A., Karwowski, W., Sun, L., and Daraiseh, N. (2006). “Estimating uninsured costs of work-related accidents, part II: an incidence-based model.” *Theoretical Issues in Ergonomics Science*, *7*(3), 247-260.

Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. K. (2020). “Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis”. *Accident Analysis & Prevention*, *136*, 105405.

Pecina, P. (2010). “Lexical association measures and collocation extraction”. *Language resources and evaluation*, *44*(1), 137-158.

Peek-Asa, C., Britton, C., Young, T., Pawlovich, M., and Falb, S. (2010). “Teenage driver crash incidence and factors influencing crash injury by rurality”. *Journal of Safety Research*, *41*, 487–492.

Peng, C.Y.J., Lee, K.L., and Ingersoll, G.M. (2002). “An introduction to logistic regression analysis and reporting”. *The journal of educational research*, *96*, 3–14.

Penkey, S.P., and Siddiqui, N.A. (2015). “A Review on Accident Pyramid and its Empirical Interpretation in Oil & Gas Industry (Upstream)”. Int. J. Sci. Res. 2015, 5, 1–3.

Perlman, A., Sacks, R., and Barak, R. (2014). “Hazard recognition and risk perception in construction”. *Safety science*, *64*, 22-31.

Perneger, T. V. (1998). “What's wrong with Bonferroni adjustments.” *Bmj*, *316*(7139), 1236-1238.

Pierson, D.E., Bronson, M.B., Dromey, E., Swartz, J.P., Tivnan, T., and Walker, D.K. (1983). “The impact of early education: Measured by classroom observations and teacher ratings of children in kindergarten”. *Evaluation Review*, *7*, 191–216.

Poh, C. Q., Ubeynarayana, C. U., and Goh, Y. M. (2018). “Safety leading indicators for construction sites: A machine learning approach”. *Automation in construction*, *93*, 375-386.

Pregibon, D. (1981). Logistic regression diagnostics. *The annals of statistics, 9*, 705–724.

Qi, G. J., Hua, X. S., Rui, Y., Tang, J., Mei, T., and Zhang, H. J. (2007, September). Correlative multi-label video annotation. In *Proceedings of the 15th ACM international conference on Multimedia* (pp. 17-26).

Quinn, G. P., and Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge, UK: Cambridge University Press.

R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2013; Available online: http://www.R-project.org (accessed on 12 January 2021).

Rakotomamonjy, A. (2004, August). Optimizing Area Under Roc Curve with SVMs. In *ROCAI* (pp. 71-80).

Read, J., Martino, L., Olmos, P. M., and Luengo, D. (2015). “Scalable multi-output label prediction: From classifier chains to classifier trellises”. *Pattern Recognition*, *48*(6), 2096-2109.

Read, J., Pfahringer, B., Holmes, G., and Frank, E. (2011). “Classifier chains for multi-label classification”. *Machine learning*, *85*(3), 333-359.

Read, J., Pfahringer, B., and Holmes, G. (2008, December). Multi-label classification using ensembles of pruned sets. In *2008 eighth IEEE international conference on data mining* (pp. 995-1000). IEEE.

Read, J., Pfahringer, B., Holmes, G., and Frank, E. (2009, September). Classifier chains for multi-label classification. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 254-269). Springer, Berlin, Heidelberg.

Rifaat, S.M., Tay, R. (2009). “Effects of street patterns on injury risks in two-vehicle crashes”. *Transportation Research Record, 2102*, 61–67.

Rikhardsson, P. M., and Impgaard, M. (2004). “Corporate cost of occupational accidents: an activity-based analysis.” *Accident Analysis & Prevention*, *36*(2), 173-182.

Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 1996.

Rivas, T., Paz, M., Martı, J.E., Matias, J.M., Garcia, J.F., and Taboada, J. (2011). “Explaining and predicting work place accidents using data-mining techniques”. *Reliability Engineering & System Safety*, 96, 739–747.

Rivolli, A., Read, J., Soares, C., Pfahringer, B., and de Carvalho, A. C. (2020). “An empirical analysis of binary transformation strategies and base algorithms for multi-label learning”. *Machine Learning*, *109*(8), 1509-1563.

Robinson, C.F., Petersen, M., and Palu, S. (1999). “Mortality patterns among electrical workers employed in the US construction industry, 1982–1987”. *American journal of industrial medicine*, 36, 630–637.

Rodríguez-Pérez, R., and Bajorath, J. (2020). “Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions”. *Journal of computer-aided molecular design*, 34(10), 1013-1026.

Rogers, D. J., and Tanimoto, T. T. (1960). “A Computer Program for Classifying Plants: The computer is programmed to simulate the taxonomic process of comparing each case with every other case”. *Science*, *132*(3434), 1115-1118.

Rom, D. M. (1990). “A sequentially rejective test procedure based on a modified Bonferroni inequality.” *Biometrika*, *77*(3), 663-665.

Ross, A., and Willson, V. L. (2017). “Paired samples T-test”. In *Basic and advanced statistical tests* (pp. 17-19). SensePublishers, Rotterdam.

Sadri, J., Suen, C. Y., and Bui, T. D. (2006, August). A new clustering method for improving plasticity and stability in handwritten character recognition systems. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 2, pp. 1130-1133). IEEE.

Rossignol, M., and Pineault, M. (1994). “Classification of fatal occupational electrocutions”. *Canadian Journal of Public Health*, *85*, 322–325.

Sanden, C., and Zhang, J. Z. (2011, July). Enhancing multi-label music genre classification through ensemble techniques. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 705-714).

Sarkar, S., and Maiti, J. (2020). “Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis”. *Safety science*, *131*, 104900.

Savolainen, P., Ghosh, I. (2008). “Examination of factors affecting driver injury severity in Michigan's single-vehicle–deer crashes”. *Transportation Research Record*, *2078*, 17–25.

Savolainen, P.T., Mannering, F.L., Lord, D., and Quddus, M.A. (2011). “The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives”. *Accident Analysis & Prevention*, *43*, 1666–1676.

Schapire, R. E., and Singer, Y. (2000). “BoosTexter: A boosting-based system for text categorization”. *Machine learning*, *39*(2), 135-168.

Schumacker, R. E. (2014). *Learning statistics using R*. Los Angeles, CA: Sage.

Seera, M., and Lim, C. P. (2014). “A hybrid intelligent system for medical data classification”. *Expert systems with applications*, *41*(5), 2239-2249.

Seixas, N. S., Sanders, J., Sheppard, L., and Yost, M. G. (1998). "Exposure assessment for acute injuries on construction sites: conceptual development and pilot test." *Applied occupational and environmental hygiene, 13(5),* 304-312.

Sharpe, D. (2015). "Your chi-square test is statistically significant: Now what?" *Practical Assessment, Research, and Evaluation*, *20*, 8.

Shirali, G.A., Noroozi, M.V., and Malehi, A.S. (2018). "Predicting the outcome of occupational accidents by CART and CHAID methods at a steel factory in Iran". *Journal of public health research*, 7, 1361.

Silverstein, B., Welp, E., Nelson, N., and Kalat, J. (1998). "Claims incidence of work-related disorders of the upper extremities: Washington state, 1993–1999." *American journal of public health, 88(12)*, 1827–1833.

Sim, J., and Wright, C.C. (2005). "The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements". *Physical therapy*, *85*, 257–268.

Snoek, C. G., Worring, M., Van Gemert, J. C., Geusebroek, J. M., and Smeulders, A. W. (2006, October). The challenge problem for automated detection of 101 semantic concepts in multimedia. In *Proceedings of the 14th ACM international conference on Multimedia* (pp. 421-430).

Song, Y., Zhang, L., and Giles, C. L. (2008, October). A sparse gaussian processes classification framework for fast tag suggestions. In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 93-102).

Storey, J. D., and Tibshirani, R. (2003). "Statistical significance for genomewide studies." *Proceedings of the National Academy of Sciences*, *100*(16), 9440-9445.

Sun, L., Paez, O., Lee, D., Salem, S., and Daraiseh, N. (2006). "Estimating the uninsured costs of work-related accidents, part I: a systematic review." *Theoretical Issues in Ergonomics Science, 7*(3), 227-245.

Sousa, V., Almeida, N.M., and Dias, L.A. (2014). "Risk-based management of occupational safety and health in the construction industry—Part 1: Background knowledge". *Safety Science, 66*, 75–86.

Steinberg, D. (2009). CART: classification and regression trees. In *The top ten algorithms in data mining* (pp. 193-216). Chapman and Hall/CRC.

Sze, N.N., and Wong, S.C. (2007). "Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes". *Accident Analysis & Prevention, 39*, 1267–1278.

Szumilas, M. (2010). “Explaining odds ratios”. *Journal of the Canadian academy of child and adolescent psychiatry*, *19*, 227.

Tang, L., Rajan, S., and Narayanan, V. K. (2009, April). Large scale multi-label classification via metalabeler. In *Proceedings of the 18th international conference on World wide web* (pp. 211-220).

Tang, S. L., Ying, K. C., Chan, W. Y., and Chan, Y. L. (2004). “Impact of social safety investments on social costs of construction accidents.” *Construction Management and Economics*, *22*(9), 937-946.

Tay, R., Rifaat, S.M., and Chin, H.C. (2008). “A logistic model of the effects of roadway, environmental, vehicle, crash and driver characteristics on hit-and-run crashes”. *Accident Analysis & Prevention*, *40*, 1330–1336.

Taylor, A.J., McGwin, G., Valent, F., and Rue, L.W. (2002). “Fatal occupational electrocutions in the United States”. *Injury Prevention*, *8*, 306–312.

Teng, C. M. (1999, June). Correcting Noisy Data. In *ICML* (Vol. 99, pp. 239-248).

Thalheimer, W., and Cook, S. (2002). “How to calculate effect sizes from published research: A simplified methodology”. *Work-Learning Research*, *1*(9).

Therneau, T., and Atkinson, B. (2019). rpart: Recursive Partitioning and Regression Trees. *R Package Version 4*.1-9.

Tixier, A. J. P. (2015). *Leveraging unstructured construction injury reports to predict safety outcomes and model safety risk using natural language processing, machine learning, and probability theory* (Doctoral dissertation, University of Colorado at Boulder).

Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., and Bowman, D. (2016a). “Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports”. *Automation in Construction*, *62*, 45-56.

Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., and Bowman, D. (2016b). “Application of machine learning to construction injury prediction”. *Automation in construction*, *69*, 102-114.

Trohidis, K., Tsoumakas, G., Kalliris, G., and Vlahavas, I. P. (2008, September). Multi-label classification of music into emotions. In *ISMIR* (Vol. 8, pp. 325-330).

Tsoumakas, G., and Katakis, I. (2007). “Multi-label classification: An overview”. *International Journal of Data Warehousing and Mining (IJDWM)*, *3*(3), 1-13.

Tsoumakas, G., and Vlahavas, I. (2007, September). Random k-labelsets: An ensemble method for multilabel classification. In *European conference on machine learning* (pp. 406-417). Springer, Berlin, Heidelberg.

Ueda, N., and Saito, K. (2003). Parametric mixture models for multi-labeled text. In *Advances in neural information processing systems* (pp. 737-744).

United States Census Bureau. (2016). Statistics of US Businesses. < https://www.census.gov/data/datasets/2016/econ/susb/2016-susb.html>. (March 20, 2019)

Veloso, A., Meira, W., Gonçalves, M., and Zaki, M. (2007, September). Multi-label lazy associative classification. In *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 605-612). Springer, Berlin, Heidelberg.

Veropoulos, K. (2001). *Machine learning approaches to medical decision making* (Doctoral dissertation, University of Bristol).

Villanova, M. P. (2014). *Attribute-based risk model for assessing risk to industrial construction tasks* (Doctoral dissertation, University of Colorado at Boulder).

Volberg, V., Fordyce, T., Leonhard, M., Mezei, G., Vergara, X., and Krishen, L. (2017). “Injuries among electric power industry workers, 1995–2013”. *Journal of safety research*, *60*, 9–16.

Waehrer, G. M., Dong, X. S., Miller, T., Men, Y., and Haile, E. (2007a). “Occupational injury costs and alternative employment in construction trades.” *Journal of occupational and environmental medicine*, *49*(11), 1218-1227.

Waehrer, G. M., Dong, X. S., Miller, T., Haile, E., and Men, Y. (2007b). “Costs of occupational injuries in construction in the United States.” *Accident Analysis & Prevention*, *39*(6), 1258-1266.

Walker, H.M., and Lev, J. (1953). *Statistical Inference*; Oxford and IBH Publishing Co.: Calcutta, India.

Wang, M., Zhou, X., and Chua, T. S. (2008, July). Automatic image annotation via local multi-label classification. In *Proceedings of the 2008 international conference on Content-based image and video retrieval* (pp. 17-26).

Weisberg, S. P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R. L., and Ferrante, A. W. (2003). “Obesity is associated with macrophage accumulation in adipose tissue.” *The Journal of clinical investigation*, *112*(12), 1796-1808.

Welch, B. L. (1951). “On the comparison of several mean values: an alternative approach.” *Biometrika*, *38*(3/4), 330-336.

White, H. (1980). “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity.” *Econometrica: Journal of the Econometric Society*, 817-838.

Wieczorkowska, A., Synak, P., and Raś, Z. W. (2006). Multi-label classification of emotions in music. In *Intelligent Information Processing and Web Mining* (pp. 307-315). Springer, Berlin, Heidelberg.

Wilcox, R. R. (1995). “ANOVA: A paradigm for low power and misleading measures of effect size?” *Review of Educational Research*, *65*(1), 51-77.

Wilcox, R. R. (2001). “Pairwise comparisons of trimmed means for two or more groups.” *Psychometrika*, *66*(3), 343-356.

Wilcox, R. R. (2005). “An affine invariant rank-based method for comparing dependent groups.” *British Journal of Mathematical and Statistical Psychology*, *58*(1), 33-42.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing*. Waltham, MA: Academic press

Wilcox, R. R., and Keselman, H. J. (2003). “Modern robust data analysis methods: measures of central tendency.” *Psychological methods*, *8*(3), 254.

Wilcox, R. R., Keselman, H. J., and Kowalchuk, R. K. (1998). “Can tests for treatment group equality be improved? The bootstrap and trimmed means conjecture.” *British Journal of Mathematical and Statistical Psychology*, *51*(1), 123-134.

Winter, E. (2002). “Chapter 53 The shapley value”. *Handbook of game theory with economic applications*, *3*, 2025-2054.

Wu, P. C. (2002). *The central limit theorem and comparing means, trimmed means, one step m-estimators and modified one step m-estimators under non-normality* (Doctoral dissertation, University of Southern California).

Xu, Q., and Xu, K. (2021). “Analysis of the Characteristics of Fatal Accidents in the Construction Industry in China Based on Statistical Data”. *International journal of environmental research and public health*, *18*, 2162.

Yan, X., Radwan, E., and Abdel-Aty, M. (2005). “Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model”. *Accident Analysis & Prevention*, *37*, 983–995.

Yang, P., Sun, X., Li, W., Ma, S., Wu, W., and Wang, H. (2018). “SGM: sequence generation model for multi-label classification”. *arXiv preprint arXiv:1806.04822*.

Yuen, K. K. (1974). “The two-sample trimmed t for unequal population variances.” *Biometrika*, *61*(1), 165-170.

Zhang, M. L., and Zhang, K. (2010, July). Multi-label learning by exploiting label dependency. I*n Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 999-1008).

Zhang, M. L., and Zhou, Z. H. (2013). “A review on multi-label learning algorithms”. *IEEE transactions on knowledge and data engineering, 26*(8), 1819-1837.

Zhang, M. L., and Zhou, Z. H. (2006). “Multilabel neural networks with applications to functional genomics and text categorization”. *IEEE transactions on Knowledge and Data Engineering*, *18*(10), 1338-1351.

Zhao, D., McCoy, A.P., Kleiner, B.M., Du, J., and Smith-Jackson, T.L. (2016). “Decision-making chains in electrical safety for construction workers”. *Journal of Construction Engineering and Management, 142*, 04015055.

Zhao, D., McCoy, A.P., Kleiner, B.M., and Smith-Jackson, T.L. (2015 a). “Control measures of electrical hazards: An analysis of construction industry”. *Safety Science*, *77*, 143–151.

Zhao, D., McCoy, A.P., Kleiner, B.M., Smith-Jackson, T.L., and Liu, G. (2015 b). “Sociotechnical systems of fatal electrical injuries in the construction industry”. *Journal of Construction Engineering and Management*, *142,* 04015056.

Zhao, D., Thabet, W., McCoy, A., and Kleiner, B. (2012). Managing electrocution hazards in the US construction industry using VR simulation and cloud technology. In *eWork and eBusiness in Architecture, Engineering and Construction, Proceedings of the European Conference on Product and Process Modelling, Reykjavik, Iceland, 25–27 July 2012*; CRC Press: London, UK, 2012; pp. 759–764.

Zhu, R., Hu, X., Hou, J., and Li, X. (2021). "Application of machine learning techniques for predicting the consequences of construction accidents in China". *Process Safety and Environmental Protection*, *145*, 293–302.

Zhu, S., Ji, X., Xu, W., and Gong, Y. (2005, August). Multi-labelled classification using maximum entropy method. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 274-281)

Zuppa, D., Issa, R.R.A., and Suermann, P.C. (2009). BIM's Impact on the Success Measures of Construction Projects. In *Computing in Civil Engineering*, 503–512.

BIOGRAPHY

Pouya Gholizadeh graduated from Danesh High School, Karaj, Iran, in 2005. He received his Bachelor of Science from Iran University of Science and Technology in 2010. He received his first Master of Science in Project Management and Construction from Shahid Beheshti University in 2013 and his second Master of Science in Construction Engineering and Management from University of Nebraska-Lincoln in 2017.