Data Analytics Methods for Supply Chain Risk Management with Applications in Transportation and Manufacturing Safety

by

Amir Mehdizadeh

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 11, 2021

Keywords: job rotation, optimization, big data analytics, crash risk prediction, naturalistic driving, near crashes, safety critical events, machine learning

Approved by

Alexander Vinel, Chair, Associate Professor of Industrial & Systems Engineering, Auburn University
Fadel Megahed Associate Professor of Information Systems & Analytics, Miami University
Mark Schall, Associate Professor of Industrial & Systems Engineering, Auburn University
Jia (Peter) Liu, Assistant Professor of Industrial & Systems Engineering, Auburn University
Peter He, Associate Professor of Chemical Engineering, Auburn University

Abstract

Supply chain management aims to understand and explain how organizations should collaborate within a chain to improve the overall competitiveness of the chain and smooth the flow of the money, material, and information between them. Any risks associated with each member will affect the overall performance of the chain. In this dissertation, we consider analytical modeling approaches to two specific components related to risks in supply chains, specifically the risk associated with manufacturing and transportation categories. In the first part (Chapter Two), we consider the problem of employing job rotation schemes to improve worker safety in a manufacturing setting by combining optimization methods with novel modeling techniques developed in the occupational safety community. Recent studies suggest that job rotation schedules may increase the overall risk of injury to workers included in the rotation scheme. We describe an optimization framework evaluating the effectiveness of a job rotation scheme using the fatigue failure model of MSD development and a case study with real injury data. Results suggest that the effect of job rotation is highly-dependent on the composition of the job pool, and the inclusion of jobs with higher risk results in a drastic decrease in the effectiveness of rotation for reducing overall worker risk. The study highlights that in cases when high-risk jobs are present, redesign of those high risk tasks should be the primary focus of intervention efforts rather than job rotation. In the second part (Chapters Three, Four, and Five), the goal is to improve transportation safety in a supply chain. To do so, we first aim to reduce the start-up burden of data collection and descriptive analytics for statistical modeling and route optimization of risks associated with motor vehicles. Then, we focus on improving the safety of truck drivers. The emergence of sensor-based Internet of Things (IoT) monitoring technologies have paved the way for conducting large-scale naturalistic driving studies, where continuous kinematic driver-based data are generated, capturing crash/near-crash safety critical events (SCEs) and their precursors. However, it is unknown whether the SCEs risk can be predicted to inform driver decisions in the medium term (e.g., hours ahead) since the literature has focused on SCE predictions either for a given road segment or for automated breaking applications, i.e., immediately before the event. Here, we examine the SCE data generated from 20+ million miles-driven by 496 commercial truck drivers to address three main questions. First, whether SCEs can be predicted using disparate driving-related data sources. Second, if so, what the relative importance of the different predictors examined is. Third, whether the prediction models can be generalized to new drivers and future time periods. We show that SCEs can be predicted 30 min in advance, using machine learning techniques and dependent variables capturing the driver's characteristics, weather conditions, and day/time categories, where an area under the curve (AUC) up to 76% can be achieved. Moreover, the predictive performance remains relatively stable when tested on new (i.e., not in the training set) drivers and a future two-month time period. Our results can inform dispatching and routing applications, and lead to the development of technological interventions to improve driver safety.

Acknowledgments

I would like to acknowledge and thank my advisor, Dr. Alexandar Vinel, for his enthusiastic and continued support throughout my graduate studies, and for training me to become a researcher. I would also like to express my great appreciation and gratitude to Dr. Fadel Megahed for his support and help during my Ph.D. I would like to thank the rest of my committee members, Dr. Mark Schall, Dr. Peter Liu, and Dr. Peter He for reviewing my work and providing valuable and constructive feedback. Special thanks to my friends, Farnaz, Oguz, Mehran, Ozhan, Mohammad, Armin, Qiong, and Miao for their amazing friendship, companionship, kindness, and help. Finally, I wish to express my deepest love and gratitude to my treasured parents, and my brother for their unconditional love, support, encouragement, and inspiration.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and motivation

1.1 Introduction

Supply chain management aims to understand and explain how organizations should collaborate within a chain to improve the overall competitiveness of the chain and smooth the flow of the money, material, and information between them (Asgari et al. 2016; Stadtler et al. 2015). A supply chain consists of different partners such as customers, distributors, manufacturers, and suppliers. It involves other components within each partner like materials, resources, and activities (Samaranayake 2005). Hence, managing a supply chain requires considering a large number of elements and their interactions to maximize the efficiency of the flow in a chain in terms of money, information, and materials.

Any components in the chain should work appropriately with the other members to deliver a given product to the customers with the promised quality in a reasonable timeline. Any risk associated with each member will affect the overall performance of the chain. It indicates the importance of macro and micro strategies aiming to identify, assess, mitigate, and monitor potential disruptions in supply chains (Aqlan and Lam 2016). Thus, Supply Chain Risk Management (SCRM) plays a vital role in managing uncertainties to minimize the impact of adverse events in supply chains (Ho et al. 2015). It should be noted that since there are a large number of partners, elements, and interactions in each supply chain, it is not straightforward to determine a clear definition for supply chain risk to cover all the associated risks. For example, Zsidisin (2003) defines it as “the probability of an incident associated with inbound supply from individual supplier failures or the supply market occurring, in which its outcomes result in the inability of the purchasing firm to meet customer demand or cause threats to customer life and safety.” Also, Ellis et al. (2010) has a more general definition for supply chain risk: “an individual’s perception of the total potential loss associated with the disruption of supply of a particular purchased item from a particular supplier.”

Ho et al. (2015) classified the supply chain risk factors into seven categories: 1) macro (e.g, natural disaster, war and terrorism) 2) demand (e.g, inaccurate demand forecasts, short products’ life cycle) 3) manufacturing (e.g, employee incidents, insufficient maintenance), 4) supply (e.g., inability to handle volume demand changes, technologically behind competitors), and 5) information (e.g, lack of information transparency between logistics and marketing, E-commerce), 6) transportation (e.g, accidents in transportation, damages in transport), and financial (e.g, exchange rate, currency fluctuations).

Figure 1.1 which used data provided by Ho et al. (2015), shows the distribution of publications based on the risk categories. It is worth noting that most of the publications are focused on the mitigating risk associated with supply and demand sections. There are a limited number of studies focused on risk mitigation methods on issues such as quality risk, lead time uncertainty, capacity inflexibility, machine failures in the manufacturing and transportation sections. Especially, no studies are investigating the effects of safety outcomes such as truck crashes and worker injuries on the disruption of supply chains.

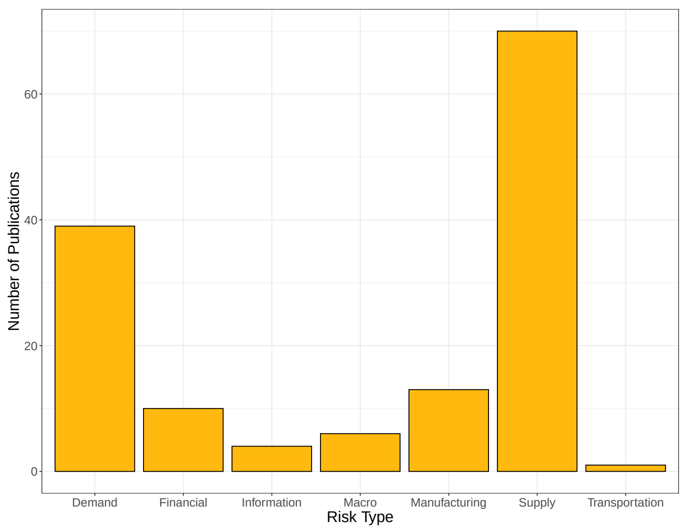


Figure 1.1: Distribution of publications based on the risk categories until 2015

Neal and Griffin (2006) introduced safety climate, which is an antecedent to safety outcomes as “individual perceptions of the policies, procedures, and practices relating to safety in the workplace." It has been shown that safety organizations that value and reward safety affects workers' behavior (Zohar and Luria 2003). Several studies showed that there is a relationship between safety climate and safety performance within organizations. For example, Tharaldsen et al. (2008) suggested that there is a negative correlation between safety climate and incident rate in the offshore industry. Varonen and Mattila (2000) had the same observation in wood processing companies. Clarke (2006) conducted a comprehensive meta-analysis of safety climate research and showed a positive relationship between safety climate and safety performance. Zohar (2000) indicated that in a manufacturing section, workgroup members share a set of perceptions of safety in the workplace. In other words, the safety of each worker can affect the other workers' safety. This point becomes more critical when considering manufacturing safety in a supply chain consisting of several manufacturing companies. Seemingly minor safety-related incidents (e.g., transportation-related incidents such as truck crashes; workplace injuries such as musculoskeletal disorders) can have a direct, cascading effect on the performance of the supply chain.

A recent study by Romero and Stahre (2021) discusses the need for smart resilient manufacturing systems, which is based on human operator resilience and human-machine systems' resilience. They suggest that the manufacturing systems should move toward both self-resilience and system-resilience. Self-resilience is related to biological, physical, cognitive, psychological, occupational health and safety, and workers' productivity. System-resilience refers to interactions between humans and machines by sharing & trading control (Inagaki et al. 2003).

As discussed above, both manufacturing and transportation safety, directly and indirectly, impact the overall performance of each component (e.g., suppliers, manufacturers) of a supply chain. Thus, in this dissertation, we aim to study approaches to use analytical methods to understand and mitigate supply chain risk, specifically as it applies to two particular facets: minimizing the risk associated with manufacturing and transportation categories. In Section 1.2 the contributions of this dissertation are explained in more details.

1.2 Contributions

The contributions of this dissertation are two-folds: In the first part (Chapter Two), we consider the problem of employing job rotation schemes to improve worker safety in a manufacturing setting by combining optimization methods with novel modeling techniques developed in occupational safety community. The work is based on a recently proposed fatigue-failure model for musculoskeletal disorders (MSD) risk evaluation. This part aims to minimize the likelihood of workers getting injured in manufacturing sections so that the overall safety of a supply chain would increase. In the second part (Chapters Three, Four, and Five), the goal is to improve transportation safety in a supply chain. To do so, we first aim to reduce the start-up burden of data collection and descriptive analytics for statistical modeling and route optimization of risk associated with motor vehicles. Then, we focus on improving the safety of truck drivers by predicting the likelihood of a safety-critical event (SCEs) in the next 30 minutes.

Improving safety in manufacturing settings Job rotation is an organizational strategy that can be used, in part, to reduce occupational exposure to physical risk factors associated with work-related musculoskeletal disorders (MSDs). Recent studies, however, suggest that job rotation schedules may increase the overall risk of injury to workers included in the rotation scheme. We describe a novel optimization framework evaluating the effectiveness of a job rotation scheme using the fatigue failure model of MSD development and a case study with real injury data. We examine the efficacy of reducing MSDs through job rotation using numerical simulation of job rotation strategies and utilizing the fatigue failure model of MSD development.

Improving transportation safety The emergence of sensor-based *Internet of Things* (IoT) monitoring technologies have paved the way for conducting large-scale naturalistic driving studies, where continuous kinematic driver-based data is generated, capturing crash/near-crash safety-critical events (SCEs) and their precursors. However, it is unknown whether SCEs can be predicted to inform driver decisions since the literature has focused on SCE predictions for a given road segment or automated breaking applications. Our aim for the third chapter is to examine the SCE data generated from 20 million miles-driven by 497 commercial truck drivers to address three main questions: (a) can SCEs be predicted using disparate driving-related data sources? (b) if so, what is the relative importance of the different predictors examined? and (c) can the prediction models be generalized to new drivers and future time periods? In the fourth Chapter, our goal is to investigate the effect of drivers' type (local, regional, and over-the-road), and different geographical areas on the likelihood of high-risk driving situations.

The remainder of this dissertation is organized as follows. In Chapter Two, we present our research about improving safety in manufacturing settings. This work has been presented as it was published in *Ergonomics* journal. Our recent research about improving the safety of truck drivers is shown in Chapters Three, Four, and Five. Chapter Three and Four involve the parts that are published on *Sensors* and *Accident Analysis & Prevention* journals respectively. In Chapter Five, we expand our work on the previous chapter.

Chapter 2

Job rotation and work-related musculoskeletal disorders: a fatigue-failure perspective

2.1 Introduction

Work-related musculoskeletal disorders (MSDs) are prevalent conditions associated with substantial direct (e.g, medical treatment) and indirect (e.g, lost workdays, lost wages) costs for both workers and organizations (Gallagher and Heberger 2013; Mossa et al. 2016). Exposure to occupational risk factors such as large forces, high rates of repetition, non-neutral postures, and vibration contributes to MSDs, which account for approximately one third of the work-related injuries and illnesses in the United States annually with each MSD injury case requiring an average of 12 days away from work (Bureau of Labor Statistics 2015).

Job rotation is an administrative control that has been used, among other purposes, in an attempt to reduce occupational exposure to physical risk factors associated with MSDs. While potential benefits of job rotation include improved worker cross-training, reduced boredom and monotony associated with simple repetitive tasks (Yoon et al. 2016), and increased motor variability, which may have positive health effects if biomechanical loading is moderated (Srinivasan and Mathiassen 2012; Mathiassen 2006; Sandlund et al. 2017), research examining job rotation to reduce MSDs has been equivocal (Comper et al. 2017; Padula et al. 2017). Trade-offs associated with job rotation may also include increased operational costs related to training workers, particularly those tasks that require high levels of experience or expertise.

Recently, several studies have suggested that job rotation may not have a net positive effect on reducing exposure to physical risk factors, and could even potentially increase the overall risk of a workplace injury in some cases (Leider et al. 2015; Padula et al. 2017; Vinel et al.

2018). To the best of our knowledge, no study has comprehensively evaluated the effectiveness of job rotation by considering multiple risk factors with a job pool consisting of jobs with different ranges of MSD risk.

A significant proportion of the existing literature is based on a linear dependency between risk and exposure. As will be clear from our analysis, if such an assumption is made, a rotation can indeed be expected to be effective at reducing the effect of MSD injuries. At the same time, recent developments in the literature suggest that a nonlinear dependency may be more accurate (Gallagher and Schall Jr 2017; Edwards 2018). In such a circumstance, a rotation scheme would be predicted to increase the overall injury risk, as will be demonstrated in this paper.

The goal of our study is to assess the effect of implementing a job rotation scheme on the overall injury risk of a group of workers by developing a general mathematical model. Our model is constructed as an optimization problem for evaluating rotation effects on multiple affected body segments (e.g., the distal upper extremity and the low back) among a pool of rotating workers based on fatigue-failure theory. Specifically, we use the Lifting Fatigue Failure Tool (LiFFT; Gallagher et al. 2017) and Distal Upper Extremely Tool (DUET; Gallagher et al. 2018) as the underlying exposure assessment models for assessing the probability of an injury.

It must be emphasized that unlike many previous efforts, we are not considering the problem of designing a practical rotation schedule, which requires advanced modeling and algorithmic tools. Our goal is to evaluate the potential of rotation in principle, and hence we are able to relax a number of practical constraints usually considered in the literature, while allowing us to make analytic conclusions on the properties of rotations. Combined with a numerical case study based on realistic data, we can evaluate the expected benefits to MSD risk due to rotation. Specifically, our model confirms some of the claims made in the literature pertaining to the potential for job rotation to have a negative effect on exposure to physical risk factors, especially when high-risk jobs are included in the rotation.

It is worth noting here that MSD risk reduction is not the only reason to implement a rotation. As mentioned above, other benefits, e.g., cross training, reduction of boredom or monotony (and consequently, psychological stress), and increased motor variability are potential advantages to rotating. Our study does not consider these aspects. Instead, our conclusions suggest that MSD risk reduction may not be achievable based on rotation alone (and, in some cases, may lead to an increase in pooled risk to the rotation cohort), and hence all other benefits must be weighed against the effect on MSD risk. Further, we do not consider any of the psychosocial factors in the etiology of MSDs. At the same time, it should also be noted that our findings suggest that those who might use job rotation to attempt to ameliorate psychosocial MSD factors may, in doing so, expose workers to increased physical MSD risk. A future study, including all economic and ergonomic factors involved in rotations is needed to develop comprehensive guidelines to determine when rotation is appropriate in practice.

The remainder of this paper is organized as follows. We present a general overview of the related literature in Section 2.2. Then, we provide background information regarding the Fatigue-Failure Theory (FFT) for evaluating MSDs risk in Section 2.3. In Section 2.4, an optimization model for job rotation scheduling is introduced. We also describe a number of properties of the proposed model and its solution, which then lead to general conclusions regarding job rotation. These conclusions are then illustrated with a case study based on injury data in Section 2.5. Finally, general conclusions and discussions are presented in Section 2.6.

2.2 Literature review

Job rotation is a widely used organizational tool that has received significant attention in the literature. There are a number of benefits associated with rotation. In this paper, we focus specifically on the effect of rotation on MSD risk. A recent survey of relevant literature has been presented by Otto and Battaïa (2017). The authors categorized papers based on the ergonomic risk assessment method applied, how ergonomic risks were considered in the optimization models (as objective function or constraint) and algorithms for solving the optimization models. The reader is referred to this survey for detailed treatment of the related literature. In addition to the articles reviewed in Otto and Battaïa (2017), below we have also included papers found by Web of Science based on the combination of keywords: “job rotation + optimization”.

In our review, in order to provide necessary context, we follow the classification presented in Otto and Battaïa (2017). Specifically, we summarize the major contributions in Tables 2.1 and 2.2. Table 2.1 describes the risk factors that have been used, including general safety criteria, risk assessment methods, the type of risk function (linear or non-linear) applied and the number of risk factors (single or multiple) considered. Table 2.2 focuses on the mathematical models employed (type of optimization problem and the algorithms used) in the studies and summarizes the contributions. Linear or nonlinear risk function refers to the way that subsequent job assignments are aggregated by the ergonomic model used in order to produce the overall risk value for each worker. As will be highlighted by our analysis, whether a linear or nonlinear risk function is used, fundamentally changes the effect of rotation, and hence we explicitly focus on this aspect. Note that some of the studies listed are concerned with risks associated with noise exposure. While noise exposure is, of course, different in nature compared to MSDs, it can be argued that the overall effect is quite analogous. Sound pressure is expressed as Force/Area over which that force is applied. With MSD fatigue failure model, stress on tissues is also represented as Force/Area, and both relationships are exponential in nature, both can be the product of a single highly stressful event or (more often) the result of exposure to repetitive stress (see, for example, Zahnert 2011). For this reason we prefer to include those in the review.

The majority of the studies reviewed focused on methods and algorithms for developing practical rotation schedules. Rotation scheduling can be computationally intensive (in fact NP-hard, see for example, Otto and Scholl 2013), and a significant portion of the literature is devoted to developing efficient algorithms. Most of the approaches are based on an integer programming (IP) formulation where binary variables are used to assign specific workers to specific job stations at specific time intervals, and constraints/objectives are used to enforce a required structure and evaluate selected ergonomic criterion. Since the resulting model is typically not computationally scalable, the authors propose heuristic solution approaches based on either specifically developed approximation ideas or metaheuristics (e.g., genetic algorithms, ant colony, simulated annealing, tabu search). Some of the articles add criteria to the algorithm in addition to ergonomic risk. For example, Diego-Mas et al. (2009) and Mossa et al. (2016) considered worker disabilities and skills, respectively.

Table 2.1: Summary of the reviewed literature: ergonomic evaluation and classification.

Reference	**Safety criteria**	**Evaluation method**	**Type**	**Risk factors**
Carnahan et al. (2000)	MSD	JSI	L	Single
Tharmmaphornphilas et al. (2003)	Occupational noise	DND	L	Single
Tharmmaphornphilas and Norman (2004)	MSD, Occupational noise	JSI, DND	L	Single
Asawarungsaengkul and Nanthavanij (2006)	Occupational noise	DND	L	Single
Bhadury and Radovilsky (2006)	-	-	L	Single
Yaoyuenyong and Nanthavanij (2006)	Occupational noise	DND	L	Single
Seçkiner and Kurt (2007)	Workload cost	General job-specific physical demand parameters	L	Single
Tharmmaphornphilas and Norman (2007)	MSD	JSI	L	Single
Asawarungsaengkul and Nanthavanij (2008a)	Occupational noise	DND	L	Single
Asawarungsaengkul and Nanthavanij (2008b)	Occupational noise	DND	L	Single
Seçkiner and Kurt (2008)	Workload cost	General job-specific physical demand parameters	L	Single
Yaoyuenyong and Nanthavanij (2008)	Occupational noise, energy expenditure	General job-specific physical demand parameters, DND, EnerExp	L	Single
Aryanezhad et al. (2009)	MSD, Occupational noise	JSI, DND	L	Multiple
Diego-Mas et al. (2009)	MSD, Mental and communication capacities	Force loads, awkward and static postures, repetitiveness, capacities of workers	NL	Multiple
Asensio-Cuesta et al. (2012a)	MSD	Force loads	NL	Multiple
Ayough et al. (2012)	-	-	L	Single
Asensio-Cuesta et al. (2012b)	MSD	OCRA, monotony	NL	Multiple
Otto and Scholl (2013)	MSD	EAWS	L	Single
Huang and Pan (2014)	MSD	Discomfort level	NL	Multiple
Mossa et al. (2016)	MSD	OCRA	L	Single
Song et al. (2016)	MSD	NIOSH-eq, force loads, geometry of tasks	NL	Multiple
Yoon et al. (2016)	MSD	REBA	L	Multiple
Digiesi et al. (2017)	MSD	RULA	L	Single
Sana et al. (2018)	MSD	OCRA, RULA, NIOSH-eq	NL	Multiple

Notes. JSI - job severity index, DND - daily noise dosage, EnerExp - energy expenditure, OCRA - occupational repetitive action, EAWS - ergonomic assessment work sheet, NIOSH - National Institute for Occupational Safety and Health, REBA - rapid entire body assessment, RULA - rapid upper limb assessment, L - linear, NL - non-linear

Most of the articles reviewed aimed to minimize the maximum risk among the workers. In other words, the implicit assumption is that improving the worst job assignment leads to the overall improvement of the risk to the worker pool. However, one can intuitively argue that in addition to the importance of reducing the maximum risk, the risk increment faced by the rest of the rotating workers must also be considered. To highlight this point, consider two examples from the reviewed studies. Tharmmaphornphilas et al. (2003) identifies the optimal job rotation schedule for three jobs performed by three workers. According to their results,

Table 2.2: Summary of the reviewed literature: optimization problem classification and major contributions.

Reference	Model	Algorithm	Major Contributions
Carnahan et al. (2000)	IP	Genetic	Cluster analysis provided a general set of rules for designing safe job rotation schedules without the use of a computer.
Tharmmaphornphilas et al. (2003)	IP	-	A case study in a manufacturing setting was presented.
Tharmmaphornphilas and Norman (2004)	IP	-	Rotation interval of approximately 2 hours was determined to be optimal.
Asawarungsaengkul and Nanthavanij (2006)	IP	Heuristic	A hierarchical planning approach proposed as a combination of engineering controls, job rotation, and/or the use of protective devices.
Bhadury and Radovilsky (2006)	Multiobjective IP	Heuristic	A hybrid model considering both the total number of tasks and the total number of the same consecutive tasks assigned to a worker was developed.
Yaoyuenyong and Nanthavanij (2006)	IP	Heuristic	A hybrid procedure utilizing four algorithms was proposed. It outperformed all four algorithms individually.
Seçkiner and Kurt (2007)	IP	Simulated Annealing	The algorithm was successful in identifying the optimal job rotation schedules.
Tharmmaphornphilas and Norman (2007)	IP	Heuristic	Uncertain task demands and non-identical worker profiles were considered.
citetasawarungsaengkul2008heuristic	IP	Genetic	The algorithm first finds the minimum number of workers. Then, based on the minimum total worker-location changeover criteria, workers are assigned to jobs.
Asawarungsaengkul and Nanthavanij (2008b)	IP	Heuristic	A hierarchical planning approach was proposed as a combination of engineering controls and job rotation.
Seçkiner and Kurt (2008)	IP	Ant Colony	The algorithm was successful in identifying the optimal job rotation schedules.
Yaoyuenyong and Nanthavanij (2008)	IP	Heuristic	Heuristic solutions were able to identify the minimum number of workers required.
Aryanezhad et al. (2009)	Multiobjective IP	LP-metric	Results from the multiobjective model dominated the results from the single objective models.
Diego-Mas et al. (2009)	-	Genetic	The algorithm took into account workers' preferences and their abilities as well as their assignments in the previous rotation.
Asensio-Cuesta et al. (2012a)	-	Genetic	The algorithm took into account ergonomic risk, physical skills of the workers and their competences.
Asensio-Cuesta et al. (2012b)	-	Genetic	The algorithm took into account disabilities of workers and their assignments in the previous rotation.
Ayough et al. (2012)	IP	Genetic, Imperialist Competitive Algorithm (ICA)	A multi-period imbalance assignment model which considers both boredom and cost was developed.
Otto and Scholl (2013)	Mixed IP	Heuristic	A tabu search based meta heuristic with promising computational performance
Huang and Pan (2014)	-	Particle swarm optimization (PSO)	An automated job rotation strategy based on the anthropometric measurements of workers was proposed.
Mossa et al. (2016)	Mixed IP	-	Both ergonomic risk and production rate were considered. The effectiveness of the optimal solutions can be significantly increased when flexible workers are employed.
Song et al. (2016)	Multiobjective	Genetic	Risk factors were considered simultaneously. The average ergonomic risk was decreased compared to sequential and single job assignments.
Yoon et al. (2016)	Mixed IP	-	The proposed model prevented sequential assignment to workstations with high workloads.
Digiesi et al. (2017)	Mixed IP	-	The trade-off between production rate and ergonomic risk was investigated. The reduction in ergonomic risk was larger than the reduction in production rate.
Sana et al. (2018)	Multiobjective IP	Genetic	Risk factors were considered simultaneously. The proposed algorithm was competitive with the models with only one risk factor.

the maximum daily noise dose was reduced by 6.0% after the rotation. However, after performing the rotation, the minimum daily noise increased by 8.72%. Therefore, even though the proposed rotation reduced the maximum daily noise dose, it also *increased* the overall daily noise exposure for the rest of workers. Yoon et al. (2016) compared their optimal job rotation scenario which was referred to as job rotation reducing cumulative workload (CWJR) with no job rotation (NJR) and serial job rotation (SJR) scenarios for three workstations (chassis, trip and finishing) using the Rapid Entire Body Assessment (REBA) risk assessment tool (Hignett and McAtamney 2000). Based on their results, CWJR provided the lowest variance of REBA scores. However, in addition to the reduction in variance, the proposed rotation decreased the maximum REBA score by 18.14%, 26.45% and 4.70% for chassis, trim and finishing workstations, respectively, in exchange for an increase in the minimum score by 43.27%, 36.00% and 13.09%, respectively. Therefore, the reduction in the maximum REBA score was always smaller than the increment in the minimum REBA score bringing into question the effectiveness of their proposed rotation. It should be noted that REBA tool guidance is qualitative in nature with higher scores simply implying a greater sense of urgency regarding the investigation of and subsequent improvement of jobs. It is unclear whether or not increases or decreases in tool output can be assumed proportional to scores.

These observations are at the center of our analysis. Specifically, we argue that when developing a job rotation schedule, it is important to consider the overall pool of workers affected, and hence the benefit to the most exposed worker should be measured against the inevitable increase in the risk faced by the rest of the pool. Based on Tables 2.1 and 2.2 we can observe that while the existing studies consider various risk evaluation methods and occupational domains, most share the following properties: a) a linear function for accumulating risk across different tasks; b) a single factor considered in the optimization; c) the objective function aims to minimize the worst assignment. In the current study we propose a model that relaxes these assumptions. We are then able to characterize the trade-off described above and make general conclusions on the effect of rotation on MSD risk.

Note that contrary to many of the reviewed studies we do not design our rotation model as an IP. This stems from the fact that instead of aiming at designing a practical rotation schedule with the corresponding application specific constraints (e.g., rotating once every hour, or limiting the number of unique stations per worker), we focus on evaluating potential benefits of the rotation in principle.

2.3 Material fatigue-failure theory for risk assessment

According to material fatigue failure theory (FFT), all materials fail as a result of an accumulation of damage resulting from loading. The damage resulting from a given load is nonlinearly related to the number of cycles to failure for that load. For example, a structure, be it metal, wood, or biological tissue, can sustain many low force loads (virtually infinite if the forces are low enough), but relatively few loads close to the ultimate strength of the given material. The ultimate strength is the load at which a material will fail with a single loading cycle. The damage resulting from a load configuration can be calculated as a percentage of the ultimate strength of the structure. For example, if a structure, such as a spinal motion segment, requires 20,000 repetitions at a given load to cause failure (structural breakdown/injury), then 2 repetitions at this load would result in 0.01 % damage (2/20,000 = 0.0001= 0.01%). Theoretically, when 100% damage is reached, the structure fails. As the theoretical percentage of damage increases, so does the risk of failure.

It has been recognized for some time that musculoskeletal tissues incur damage in accordance with fatigue failure principles when tested in vitro, including spinal motion segments (Brinckmann et al. 1987; Gallagher et al. 2005), tendons (Schechtman and Bader 1997; Wren et al. 2003), and ligaments (Thornton et al. 2007; Lipps et al. 2013). Further, several lines of evidence have bolstered the notion that fatigue failure may be a causal mechanism of MSDs. For example, epidemiological studies examining the factors of force and repetition have demonstrated a pattern of interaction indicative of a fatigue failure process across a wide variety of MSDs (Gallagher and Heberger 2013). Furthermore, a recent review of fatigue testing in tendons has concluded that the damage observed in biopsies of tendinopathic tendons strongly correlate to the damage that occurs in vitro fatigue testing (Shepherd and Screen 2013). In addition, both the DUET (Gallagher et al. 2018) and LiFFT (Gallagher et al. 2017) risk assessment tools used in the current paper (both based on fatigue failure principles) have demonstrated strong concurrent criterion-related validity against low back and upper extremity MSD outcomes, respectively. An important feature of these risk assessment tools is the ability to sum exposures associated with multiple tasks. Thus, there is substantial evidence supporting fatigue failure as a mechanism in the development of MSDs, and fatigue failure theory has validated methods to assess cumulative risk across multiple tasks. These characteristics support the use of fatigue failure methods to assess the effects of job rotation on MSD risk.

This work is based on a generalized model that follows the form shown in (2.1), where the estimated damage from a given job (y) can be related to the probability of a negative health outcome in populations of industrial workers:

$$f(y) = \frac{ay^p}{1 + ay^p}. \tag{2.1}$$

This equation was used here for both the distal upper extremity and the low back with different coefficients from the fatigue-failure literature, specifically, LiFFT (low back) and DUET (distal upper extremity) risk assessment tools discussed above. The coefficients are given in Table 2.3.

Table 2.3: Equation (2.1) coefficients for LiFFT and DUET tools.

Risk Assessment Method	a	p
LiFFT	$10^{1.723}$	1.024
DUET	$10^{1.573}$	0.747

A key distinction for biological materials is their ability to heal. Hence, a summation of damage exceeding 100% may not result in an injury. Both DUET and LiFFT considered this healing factor when relating damage to injury risk. A job or rotation of jobs that greatly exceeds 8-12 hours per day or includes regular 6 or 7 work weeks may pose much greater risk. A worker's quantity and quality of sleep, rest, and recovery are other potentially important factors. Healing and recovery were not considered here, but on-going work is beginning to address these issues so that they can be incorporated in future work.

2.4 Mathematical optimization model

2.4.1 Problem formulation

Consider a set of N jobs and M workers. Suppose there are K anatomical sites of MSD risk (e.g., injuries to the distal upper extremities, lower back, etc.). The jobs involve highly repetitive work and we model the corresponding risks according to the fatigue-failure theory described in Section 2.3. Specifically, MSD risk for each worker is calculated by determining the cumulative damage on each MSD anatomical site and then applying (2.1). A rotation scheme is then selected in such a way as to minimize a pre-selected measure of the overall worker pool risk.

We denote by X_{ij} the decision variable representing the number of cycles of job j performed by worker i. The total damage that worker i carries from MSD anatomical site k is denoted as CD_{ik}. The corresponding risk based on the fatigue-failure model for worker i from the MSD anatomical site k is denoted as P_{ik}. The overall fatigue-failure risk for worker i is denoted with TP_i, which is calculated according to the standard formula for the probability for a union of random events. The following list summarizes the nomenclature.

List of parameters and variables

Parameters

α number of cycles that each worker is required to perform

a_k, p_k coefficients in risk function (2.1) for the MSD anatomical site k

dpc_{jk} damage on MSD anatomical site k per cycle of job j

K total number of MSD anatomical sites

M total number of jobs

N total number of workers

NC_j number of cycles of job j

Variables

CD_{ik} cumulative damage for worker i on MSD anatomical site k

P_{ik} FFT risk for worker i on MSD anatomical site k

TP_i overall FFT risk for worker i

X_{ij} number of cycles performed by worker i on job j

In order to simplify the analysis, we assume that all workers are of the same skill level, or in other words, there are no restrictions on assigning any job to any worker, and no known preference for any workers to perform any of the jobs. Similarly, we assume that all workers are required to perform the same total number of cycles (α). Note that both of these assumptions can be easily relaxed by introducing additional parameters. Further, we also assume that different MSD anatomical sites are probabilistically independent, which simplifies calculation of the cumulative FFT risk for each worker. While this assumption is rather restrictive, in general it can also be relaxed by introducing correlation coefficients. These assumptions allow us to concentrate on the mathematical properties of the solutions rather than effect of specific parameter values.

The problem of determining the optimal job rotation can be formulated as follows.

$$\min \quad F(TP_1, \ldots, TP_N) \tag{2.2a}$$

s.t.

$$CD_{ik} = \sum_{j=1}^{M} dpc_{jk} X_{ij}, \quad i = 1, \ldots, N, k = 1, \ldots, K, \tag{2.2b}$$

$$P_{ik} = \frac{a_k e^{p_k \ln CD_{ik}}}{1 + a_k e^{p_k \ln CD_{ik}}}, \quad i = 1, \ldots, N, k = 1, \ldots, K, \tag{2.2c}$$

$$TP_i = G(P_{i1}, ..., P_{iK}), \quad i = 1, \ldots, N, \tag{2.2d}$$

$$\sum_{i=1}^{N} X_{ij} = NC_j, \quad j = 1, \ldots, M, \tag{2.2e}$$

$$\sum_{j=1}^{M} X_{ij} = \alpha, \quad i = 1, \ldots, N, \tag{2.2f}$$

$$\mathbf{X} \geq 0. \tag{2.2g}$$

The optimization model above minimizes the objective function $F(TP_1, \ldots, TP_N)$, which evaluates the risk profile of the pool of workers (e.g., average or maximum risk). The constraints ensure that the risk for each worker is calculated according to the fatigue-failure theory, i.e., constraint (2.2b) calculates the total exposure for each worker, and constraint (2.2c) corresponds to equation (2.1) for each anatomical site. Constraint (2.2d) refers to the overall probability of getting injured for worker i from any of the MSD anatomical sites, which is calculated according to the standard formula for a union of independent random events (function

G). For example, if $K = 2$, then $G(p_1, p_2) = p_1 + p_2 - p_1 p_2$. Finally, constraints (2.2e) and (2.2f) ensure that each worker performs the required number of cycles and each job is completed. Note that the problem is only feasible if $\sum_j NC_j = \alpha N$, i.e., the total amount of work demanded is equal to the total worker availability.

2.4.2 Job rotation scenarios and properties

The set of feasible solutions to the problem above describes all possible rotations. Depending on the choice of the objective function different risk profiles will be preferred. For the purpose of the analysis, we define two natural rotation schemes, which we will refer to as *min-max* and *min-average*. In addition we also describe a naive approach (referred to as *complete*), which will be used for comparison in the case study.

Min-max rotation. If $F(TP_1, \ldots, TP_m) = \max\{TP_1, \ldots, TP_N\}$, then the rotation aims at minimizing the worst assignment. Intuitively, this means that if permitted by other constraints, such a rotation should result in assignments with $TP_1 = \ldots = TP_N$, i.e., all workers are equally exposed (see Claim 3). This objective corresponds to the approach used in most studies in the literature, and also naturally coincides with intuitive rotation goal of reducing the risk associated with the worst assignment.

Min-average rotation. If $F(TP_1, \ldots, TP_N) = \frac{TP_1 + \ldots + TP_N}{N}$, then the optimal rotation minimizes the average worker risk without concern for relative equity. It follows from concavity of risk function in equation (2.1) (see Figure 2.1 for the case of two anatomical sites) that average risk will be minimal when the rotation has maximum possible difference between the best and the worst assignments. In other words, the solution can be constructed with a greedy algorithm, which iteratively selects the least risky but still feasible assignments one by one. Most importantly, if the number of jobs is the same as the number of workers, then the *min-average* solution can be obtained by not rotating, i.e., each worker performs a full single job (see Claim 2). For this reason, we will also refer to the model with the min-average function as the *no rotation* case.

Complete rotation. We denote as a *complete* rotation a heuristic idea which distributes equal amounts of each job to each worker, i.e., it assigns $1/N$ of each job to each worker. Intuitively, this results in equal risk distribution between all workers, yet, as we will demonstrate, it is always worse than min-max approach.

Given these definitions, the following results on properties of the rotation schemes can be established. For the purpose of this paper we will omit formal proofs and instead provide intuitive justifications.

Claim 1 *For two independent MSD anatomical sites, the corresponding risk function for calculating the total FFT risk of individual worker, TP_i, is concave.*

Note that for two independent MSD anatomical sites, $TP_i = f(y_{i1}) + f(y_{i2}) - f(y_{i1})f(y_{i2})$, where y_{ik} is the total damage on the MSD anatomical site k for worker i and fatigue-failure function f is given in (2.1). The claim can be verified visually, by inspecting the corresponding graph on Figure 2.1.

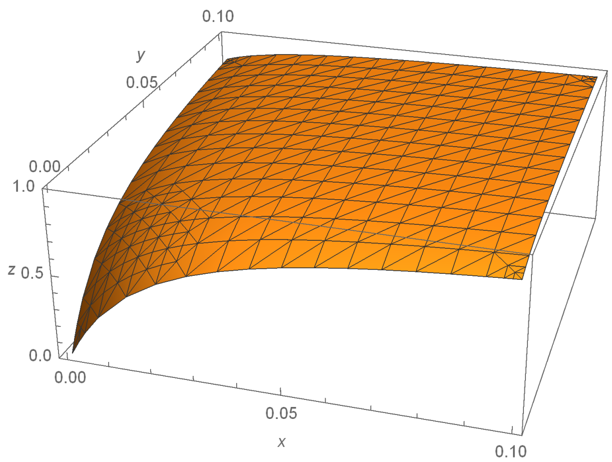


Figure 2.1: Risk predicted by fatigue-failure theory for two independent MSD anatomical sites. x and y axes correspond to the cumulative damage on the two sites, while z-axis measures the risk according to fatigue-failure theory.

Min-average risk criterion is linear, hence, the objective of the corresponding optimization problem is concave, and Jensen's inequality implies the following claim.

Claim 2 *For two independent MSD anatomical sites, if $N = M$, $\alpha = 1$, and $NC_j = 1$ for all j, min-average solution is equivalent to no rotation solution, i.e., $X_{ij} = 0$, for $i \neq j$ and $X_{ii} = 1$, is optimal for min-average problem.*

This claim essentially establishes that if each job requires an equal number of cycles, and each worker will perform equal number of cycles, then the minimum average risk is achieved if no rotation is implemented, i.e., each worker is performing a single job in its entirety. Alternatively, the min-max criterion implies an opposite extreme.

Claim 3 *Optimal min-max solution is such that $TP_i = TP_j$ for all i, j.*

This claim describes a rotation in which all workers are equally exposed to risk. This claim can be justified intuitively. Suppose that there is a risk distribution among the workers such that some of them do not face equal risk. We can then reduce the maximum risk by rearranging the cycles between the worst and the best assignments until the corresponding workers are equally exposed. This argument can be continued until the risk is equally distributed among all workers. Note that if multiple MSD anatomical sites are considered, such assignment is not unique. Hence, in order to obtain an optimal *min-max* rotation, in general, one still needs to solve an optimization problem.

Consequently, *min-max* rotation and *no rotation* can be viewed as two extremes on the spectrum of FFT risk distributions, and any other rotation would result in a probability distribution that is either in some sense "in-between" the *no rotation* and *min-max* rotation or dominated by them. Indeed, *min-max* corresponds to the most conservative approach (all risk is equalized), while the *no rotation* enforces the largest possible inequality.

The results above imply the central claim of our analysis.

Claim 4 *If MSD risk exposure of workers is evaluated through the fatigue-failure theory (or in fact any other approach that is based on a concave risk function, such as* (2.1)*), then it is impossible to design a rotation that does not result in the increase in the average risk exposure of the worker pool.*

Indeed, by design any decrease in the maximum risk exposure, is accompanied by an increase in exposure for some of the rest of the workers, and since the risk function is concave, this trade-off is always negative, i.e., the risk improvement is always smaller than the risk increase. In the next section we illustrate this result with a realistic case study and characterize the scale of this effect depending on the job pool composition.

Note that concavity of the risk function used to aggregate the MSD risk for each worker is the main reason for the negative trade-off above. Indeed, if instead we consider a linear function, then any rotation will not change the average risk in the worker pool, or in other words, all risk increases will be exactly compensated by risk reductions. Consequently, with this assumption one can be expected to be able to take advantage of benefits of rotation (equalization of risk exposure), without paying the corresponding cost (in terms of overall risk increase). This is indeed the assumption made in most of the studies reviewed in Section 2.2. Our conclusions suggest that while very helpful in simplifying the optimization step, a linearity assumption changes the fundamental structure of the problem, and hence should not be made lightly. It is also worth noting that if a convex risk function is used, then a rotation will in fact lead to an improvement to the overall worker pool risk (min-average and min-max solutions coincide). That said, ultimately, the choice of risk function is not up to the decision maker, and rather the risk function should be selected based on the best available ergonomic evidence.

2.5 Case study

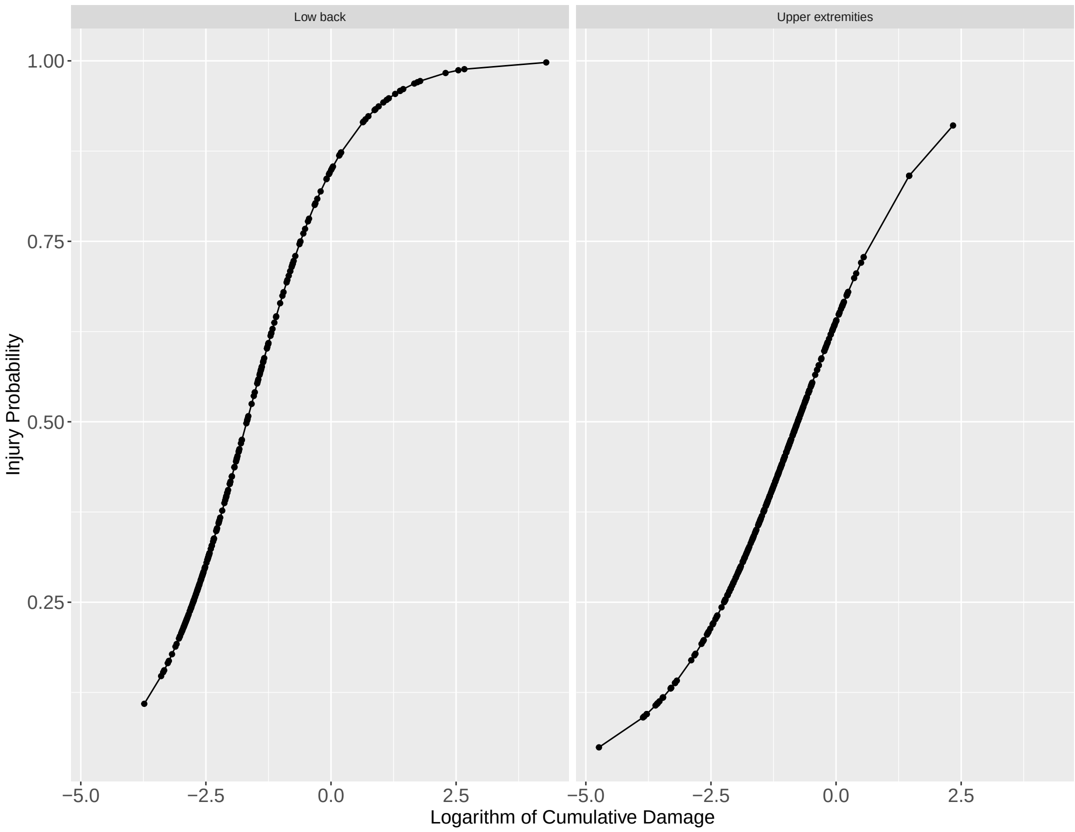
In the case study presented here, each job is characterized in terms of damage per cycle at the low back and distal upper extremity. FFT risks corresponding to these two MSD anatomical sites are determined using the LiFFT and DUET risk assessment tools, respectively, which are described in Section 2.3. Without loss of generality we assume that $\alpha = 1$, (i.e., each job consists of a single cycle, but each worker can be assigned any fraction of that cycle). Our goal is to compare the results of the three described rotation scenarios (*min-max*, *no rotation* and *complete*). *Min-max* rotation is obtained by solving the corresponding optimization problem in AMPL modeling language (Fourer et al. 1990). *No rotation* and *complete* assignments can be obtained analytically.

2.5.1 Data description

We have used two datasets for the case study. The datasets have been collected and described in detail in Marras et al. (1993) and Sesek (2000) respectively, and have been used to validate LiFFT and DUET fatigue-failure tools in Gallagher et al. (2017) and Gallagher et al.

(2018). For our purposes, the datasets consist of 235 and 441 jobs, each represented with total cumulative damage on either low back (LiFFT tool) or upper extremities (DUET tool). The corresponding fatigue-failure probabilities are calculated according to (2.1). The distribution of cumulative damage and probabilities for both datasets is given in Figure 2.2. Both include a wide range of jobs from low- to high-risk.

Figure 2.2: Fatigue-failure probabilities and cumulative damages in the job pools for the two datasets used.



2.5.2 Methodology and experiment setup

We are interested in investigating the effect of rotation depending on job composition. We consider an equal number of jobs and workers. Consequently, a default rotation corresponds to the case of one worker performing a single job. As suggested in the analysis above, this approach (referred as *no rotation* below) is equivalent to the *min-average* rotation scheme, i.e., minimizes the expected number of affected workers. Then, any rotation scheme compared to the *no rotation* case represents a trade-off between improving working conditions for some workers (by giving them portions of less physically demanding tasks) and increasing risk faced by others (by assigning them portions of more physically demanding tasks). Hence, our goal is to evaluate this trade-off. Consequently, we use a combination of numerical characteristics and visual inspection of the risk distribution among workers.

In each experiment we create $M = 10$ jobs, assigned to $N = 10$ identical workers, by sampling from a given collection of jobs. Specifically, since we do not have any real datasets simultaneously measuring fatigue-failure damage for multiple anatomic sites, we create synthetic jobs by combining the datasets described above. For each of $M = 10$ created jobs we sample independently one job from each of the datasets and pair them together. This then results in a collection of 10 jobs that are subjected to rotation schemes. This allows us to study the effect of multiple MSD anatomical sites of risk and changes in job composition, while still relying on realistic data.

To analyze rotation effects, we employ two approaches: a visual inspection of risk distribution and two quantitative measures. Figure 2.3 depicts an example of risk distribution resulting from the three rotation schemes on 10 randomly sampled jobs. We plot the resulting MSD risk calculated according to the fatigue-failure model for each of the 10 workers, where workers are ordered by the risk given by the *no rotation* regime. As analyzed earlier, both the *min-max* and *complete* rotations result in assignments that have the same risk for each worker. This graph lets us visually asses the trade-off due to incorporating a rotation.

We also define two quantitative measures, which will be referred to as Average Increased Probability per Worker (AIPW) and Efficiency Ratio (ER). AIPW provides the average per worker change (increase) in MSD risk due to rotation. Formally, if the risk for worker i before and after the rotation are p_i and r_i respectively, the AIPW is calculated as AIPW $= \sum_i^n \frac{r_i - p_i}{n}$. Similarly, ER measures the total improvement due to rotation as a fraction of the total worsening: ER $= \frac{\sum_i \max\{r_i - p_i, 0\}}{\sum_i \max\{p_i - r_i, 0\}}$. Larger values of ER and lower values of AIPW correspond to greater efficiency in the rotation scheme. Note that, as follows from the analysis in Section 2.4, for any rotation scheme AIPW > 0 and $0 <$ ER < 1. Table 2.4 presents the quantitative description of the random case given in Figure 2.3. Observe that *complete* rotation is objectively worse than *min-max* scheme, though the difference is minimal. Both rotations improve on (decrease) the risk faced by the two most exposed workers, but at the same time, considerably worsen the conditions of the other eight workers. ER at 0.08 means that the observed deterioration is 12.5 times larger than the improvement, and on average each worker's risk is increased by 16.5% (AIPW $= 0.165$).

Figure 2.3: Risk distribution for the three rotation schemes (complete, min-max and no rotation) for an illustrative random case

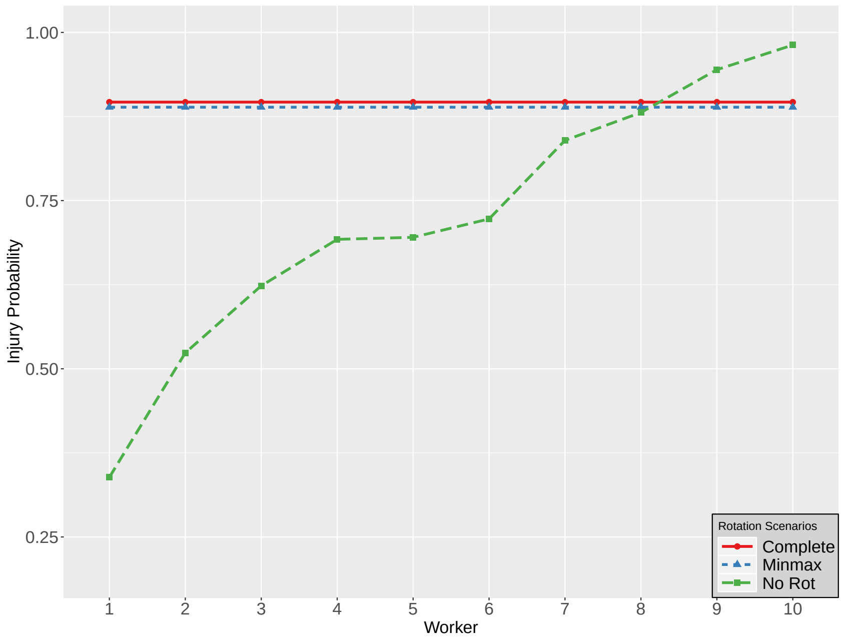


Table 2.4: Summary of rotation effectiveness for the three rotation schemes (complete, min-max and no rotation) for an illustrative random case. Columns *range* and *mean* represent the difference between minimum and maximum risk and average risk among the the workers. AIWP and ER measures are defined in Section 2.5.2.

Scenario	**Range**	**Mean**	**AIPW**	**ER**
No rotation	0.339 - 0.981	0.724	-	-
Min-max	0.889 - 0.889	0.889	0.165	0.083
Complete	0.896 - 0.896	0.896	0.172	0.072

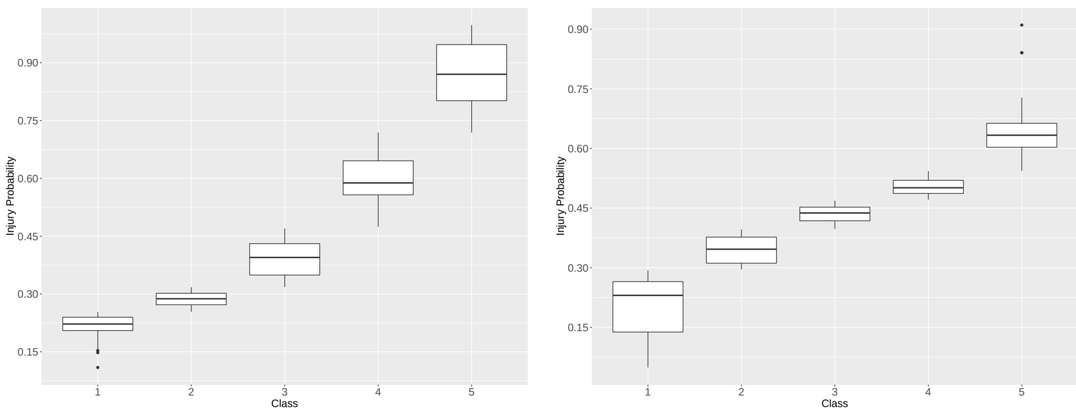
We expect that the effectiveness of rotation will dramatically decrease as we introduce higher-risk jobs into the pool. In order to test it, we implemented the following sampling procedure. We split the jobs from both datasets into five groups using four quantiles (20%, 40%, 60% and 80%), which are denoted as classes C1 through C5, where class C1 corresponds to the 20% least risky jobs, class C2 is the 40% least risky, etc. Tables 2.5 and 2.6 and Figure 2.4 provide descriptions of job composition in each class. We then sample separately from each class, which guarantees that we obtain job pools with wide variety of risk levels. We repeat sampling 100 times for each class. Note that this specific sampling procedure is adopted as a way to create cases with a diverse set of exposure levels, which could have been achieved in alternative ways. The selected procedure enables us to analyze model performance for various profiles of MSD risk.

Table 2.5: Risk class composition for low back dataset

Class	**Risk**	**# of Jobs**
C1	(0, 0.25]	47
C2	(0, 0.32]	94
C3	(0, 0.47]	141
C4	(0, 0.72]	188
C5	(0, 1]	235

Table 2.6: Risk class composition for distal upper extremity dataset

Class	**Risk**	**# of Jobs**
C1	(0, 0.29]	89
C2	(0, 0.40]	180
C3	(0, 0.47]	265
C4	(0, 0.54]	353
C5	(0, 1]	441



(a) low back dataset (b) distal upper extremity dataset

Figure 2.4: Job risk distribution in the risk classes for the two datasets.

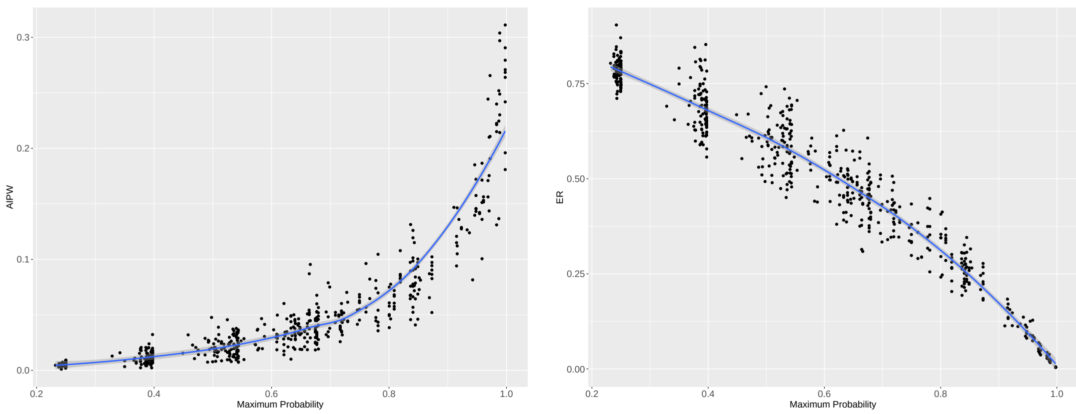
2.5.3 Case study results

Figure 2.5 plots AIWP and ER measures as functions of the maximum fatigue-failure probability in the job pool. In other words, it depicts how potential effectiveness of rotation changes as we include more and more risky jobs into the pool. Results for representative samples from each class are also provided in Table 2.7 and Figure 2.6. As discussed earlier, *complete* rotation

Table 2.7: Effectiveness measures (AIWP and ER) and rotation scheme characteristics for the representative instances from each of the classes.

Class	Scenario	Range	Mean	AIPW	ER
	No rotation	0.239 - 0.452	0.396	-	-
C1	Min-max	0.404 - 0.404	0.404	0.008	0.707
	Complete	0.404 - 0.404	0.404	0.008	0.701
	No rotation	0.290 - 0.583	0.483	-	-
C2	Min-max	0.496 - 0.496	0.496	0.012	0.668
	Complete	0.496 - 0.496	0.496	0.013	0.659
	No rotation	0.308 - 0.711	0.560	-	-
C3	Min-max	0.585 - 0.585	0.585	0.025	0.579
	Complete	0.587 - 0.587	0.587	0.026	0.563
	No rotation	0.321 - 0.839	0.633	-	-
C4	Min-max	0.682 - 0.682	0.682	0.049	0.437
	Complete	0.687 - 0.687	0.687	0.054	0.402
	No rotation	0.339 - 0.981	0.724	-	-
C5	Min-max	0.889 - 0.889	0.889	0.165	0.083
	Complete	0.896 - 0.896	0.896	0.172	0.072

is always outperformed by the *min-max* scheme. At the same time, the difference is mostly insignificant.

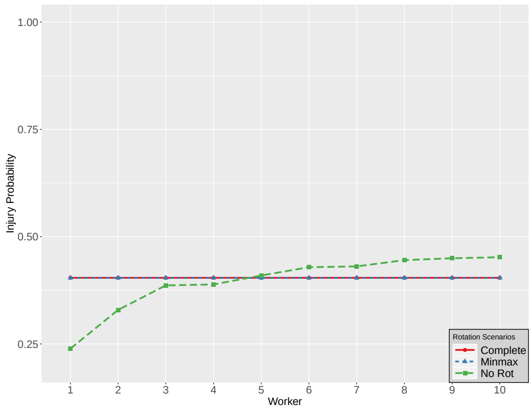


(a) AIPW vs maximum risk in the pool

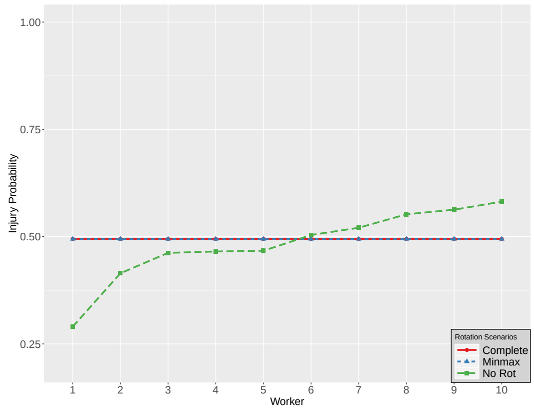
(b) ER vs maximum risk in the pool

Figure 2.5: Effectiveness of job rotation depending on the job pool composition.

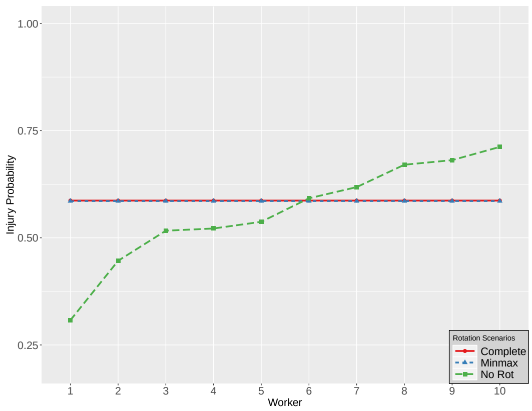
As shown in Figure 2.5, both performance measures (AIPW and ER) indicate that the effectiveness of job rotation is reduced by the inclusion of higher risk jobs in the pool. For example, as indicated by the figure, if there is even one job with the risk probability of 90% in the job pool, the average risk for each worker in a *min-max* rotation increases by around 15% (AIPW $= 0.15$) and the observed improvement of the worst assignments is around 5.5 times



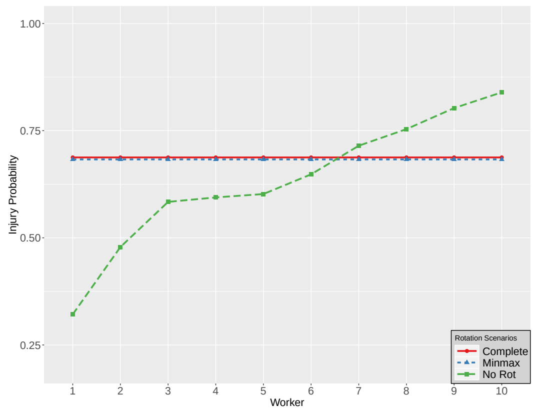
(a) C1 job class



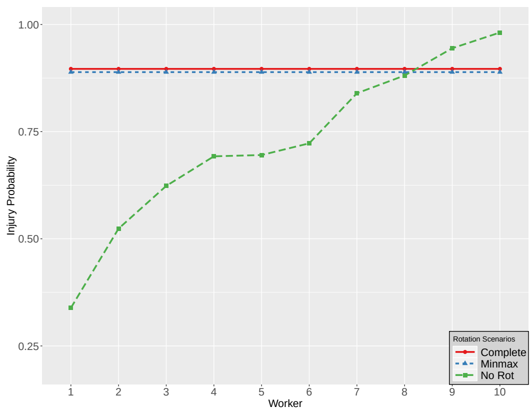
(b) C2 job class



(c) C3 job class



(d) C4 job class



(e) C5 job class

Figure 2.6: Comparison of illustrative cases for the five risk classes. Each case presents the three rotation schemes for a representative instance sampled from the corresponding job class.

lower than the deterioration observed for easier jobs (ER $= 0.18$). On the other hand, if only low-risk jobs are present, while improvement due to rotation is still less in absolute value compared to the increased risk for easier jobs, the trade-off inefficiency is considerably less drastic. Compare the risk profiles given in Figures 2.6(a) and 2.6(e). In the former (sampled from low-risk class C1), the rotation improves the risk for seven workers by increasing exposure faced by three workers. The average risk is essentially unchanged (AIPW $= 0.007$) with relative difference between improvement and deterioration at 0.7 (worsening of conditions is 1.4 times larger than improvement). On the other hand, in the representative sample for class C5, eight workers see their risk increase in order to improve the conditions for only two workers, with average increase in MSD risk of 16%. Even more concerning, the relative difference is only 0.08.

2.6 Conclusions

In this study, the effectiveness of using job rotation as a strategy for reducing exposure to physical risks associated with MSDs was evaluated following the precepts of fatigue-failure theory. Based on the evidence presented, we have drawn the following conclusions:

1. In general, any rotation leads to a trade-off between increasing the risk associated with low-risk jobs for the sake of reducing the risk of high-risk jobs.

2. *Min-max* and *min-average* rotations can be viewed as two extremes on the spectrum of risk distributions. Any rotation would result in a probability distribution that is either "between" the *min-average* and *min-max* solutions or dominated by them.

3. Due to concavity of the risk-function, any gain in efficiency is guaranteed to be smaller in absolute value when compared to the losses if measured according to fatigue-failure theory. The difference varies from very large, if even a single high-risk job is present in the pool, to small if all jobs are fairly low risk.

4. The effectiveness of job rotation is highly dependent upon the composition of the job pool.

5. It is unlikely that any rotation alone can be expected to substantially help in alleviating the effect of high risk jobs on MSD injuries. On the other hand, in the case of a job pool which contains only low risk jobs, a rotation may be helpful in achieving risk equity between workers. In the case of medium-risk pools, a general conclusion cannot be made and the optimal strategy depends on the associated costs and specific decision preferences.

6. The current study considers the effect of rotation on exposure to physical risk factors, specifically highlighting their potential to increase MSD risk. Any judgement on the effectiveness of a particular rotation scheme in a practical setting must be evaluated considering the full spectrum of potential benefits and costs, e.g., psychosocial factors, worker training and productivity.

It must be emphasized that our conclusions are consequences of a mathematical model, and hence are subject to our assumptions. At the same time, results of this study have important implications for occupational health and safety practitioners. Most importantly, job rotation *alone* does not appear to be an effective means of redistributing injury risk between low-risk and high-risk jobs when considering the fatigue-failure perspective. The increase in risk for those in formerly low-risk situations as a result of a rotation scheme can exceed the decrease in risk for a worker in a high-risk situation. Job rotation will thus be counterproductive in terms of overall injury risk in such scenarios. When high-risk jobs are present, the best recommendation remains the use of ergonomic principles to redesign such jobs to reduce injury risk. However, if all jobs in the rotation are relatively low risk, according to our model, workers may derive benefits from rotation such as decreased boredom, increased skill development, and increased motor variability without significant increases to injury risk.

The proposed procedure for evaluating the effectiveness of job rotation has some limitations. The model itself relies on a number of assumptions. First, we assume that MSD risk is accurately described with fatigue-failure theory, which, while supported by the literature (Gallagher and Schall Jr 2017; Edwards 2018), remains an approximation of a real physical system.

Musculoskeletal tissues have the capacity to recover, heal, and adapt that is not currently captured by the fatigue failure tools used in this optimization. Future work examining the healing and recovery of musculoskeletal tissues and its relationship with the fatigue failure mechanism is warranted.

Another limitation is our assumption that anatomical sites are independent and that there are no constraints on rotation structure (e.g., each worker can perform any amount of each job). By assuming a ”typical” worker and not including personal characteristics (e.g., gender, body mass index, level of experience/training) in the proposed model, the specificity to any particular occupational group is limited. Individual differences and abilities, particularly as they relate to subsequent susceptibility or resistance to MSD injuries, should be considered in future work as the workforce is both aging and increasingly obese.

Finally, we considered MSDs as the sole safety criterion and the severity of injury was not considered. Of course, other criteria (e.g., occupational noise, exposure to psychosocial risk factors, production rate and the resulting return on investment) come into consideration when deciding whether to adopt job rotation. These effects should certainly be considered in any practical evaluation of job rotation plans. We believe, however, that our findings have relevance to those who may use job rotation for other risks and/or purposes, since, as evidenced by our findings, any rotation scheme has the potential to inadvertently lead to increased physical MSD risk, regardless of potential positive effects of other factors.

Supplementary materials

The data and the corresponding code (AMPL and R) used in the case study are available through the following GitHub Repository: `https://github.com/mehdizadeamir/Job-rotation.git`.

Chapter 3

Review of data analytic applications in road traffic safety. descriptive and predictive modeling

3.1 Introduction

Despite the significant technological advances in motor vehicle sensing technologies (e.g., lane departure detection and collision mitigation sensing systems), road crashes have remained a pressing global health issue. The World Health Organization estimated that road injuries are the 8^{th} leading cause of death worldwide, resulting in 1.4 million deaths annually World Health Organization (2018a). Perhaps more importantly, the incidence of such crashes and their severity are on the rise. By 2030, traffic-related deaths are predicted to become the 7^{th} leading cause of death worldwide (World Health Organization 2018a). The increase in annual deaths is seen in low- and high-income countries alike. For example, in the U.S., an estimated 37,133 people died in road crashes in 2017 (National Highway Traffic Safety Administration, NHTSA 2018), which constituted a 7.5% increase from the average annual deaths recorded in 2012-2016 (Insurance Institute for Highway Safety 2018). In addition to the massive loss of life, motor vehicle (which is used to capture passenger cars, motorcycles, buses and trucks) crashes cause significant economic losses. According to the World Health Organization (2018b), “road traffic crashes cost most countries 3% of their gross domestic product.” In the U.S., it is estimated that the total value of societal harm from motor vehicle crashes exceeds $830 billion annually (Blincoe et al. 2015), which is equivalent to $\approx 4.4\%$ of the country’s gross domestic product (World Bank, 2018).

Consequently, there are multiple diverse streams of research dedicated to curbing such driving-related risks. This review focuses on data analytics approaches, which revolve around the idea of using data to characterize and predict traffic risk in order to prescribe better (safer) routes, driver assignments, rest breaks, etc. With the advances in information technology it is possible to collect ever increasing amounts of relevant data, such as comprehensive incident databases, real-time driving data feeds, or relevant factor characteristics (e.g., detailed historical and forecasted weather and traffic reports). Further, there has been a tremendous improvement in the variety and capabilities of data analytics tools and methods that can be applied to all steps of modeling (data collection, processing, prediction, or prescription). The goal of this study then, is to pull together and categorize the existing literature on different aspects of research relevant to enabling data-driven analytics approaches to traffic safety.

The study was inspired by an observation that there exists an apparent disconnect between two essential facets of pertinent research efforts: statistical modeling of crash risk on one hand and prescriptive modeling for decision making on the other. For example, it is very common in operations research (OR) literature to assume that the crash probability is time-invariant (Erkut et al. 2007; Androutsopoulos and Zografos 2012), and is, in fact, in the range of 10^{-8} to 10^{-6} per mile (Abkowitz and Cheng 1988). This contradicts the findings from the predictive stream of research, with multiple efforts studying the effect of real-time crash risk factors (traffic and weather conditions) on the likelihood of a crash. According to the reviews in Theofilatos and Yannis (2014) and Roshandel et al. (2015) different traffic and weather conditions would result in different crash risk profiles, bringing into question the effectiveness of the methods often used by OR community for considering risk in decision-making process.

In order to further examine this apparent gap we have conducted a more formal bibliographic study. Based on the keywords and search strategy described in the Supplementary Materials Section, we identified 856 relevant documents (i.e., published articles, proceeding papers, and book chapters). To categorize these documents for this review, a text/bibliometric analysis was performed using the *bibliometrix* **R** package (Aria and Cuccurullo 2017), with the goals of: (a) examining the co-occurrences of keywords within documents since this shows a link between the topics captured by these keywords; and (b) constructing a conceptual structure map of the literature based on a more streamlined keywords list ("Keyword Plus", see Garfield and Sher (1993) for a detailed introduction). The results are shown in Figures 3.1 and 3.2, respectively.

In the keyword co-occurence network, induced by the documents found, a pair of keywords is connected by a link, if they appear in the same document (the links are weighted according to the number of co-occurrences). This network is then clustered with K-means clustering algorithm (all parameters selected automatically by *bibliometrix* package). The clusters and most important links (corresponding to more than four co-occurrences) are depicted on Figure 3.1 with the black and red links depicting within-cluster and between-cluster connections respectively. The conceptual structure map (Figure 3.2) aims at identifying the common emerging concepts in the expanded "Keyword Plus" network. Here, dimensionality reduction technique (multidimensional scaling) is applied to the concept co-occurrence network in order to project it to two dimensions, and the result then clustered with a K-means clustering algorithm. More details on the precise implementation can be found in (Aria and Cuccurullo 2017).

Based on Figures 3.1 and 3.2, two important observations can be made. First, the literature can indeed be grouped into two main groups: (a) an explanatory/predictive modeling stream, where the keywords emphasize the collected data (loop detector data), predictors (traffic, weather, time and/or infrastructure), models used (regression, spatial-analysis, Poisson-gamma and negative binomial), and model outcomes (rates, crash frequencies, and crash prediction); and (b) a prescriptive modeling stream, where the focus is on developing algorithms to manage risk, particularly for hazardous materials (hazmat) trucking, through the selection of paths and routes. Second, the cluster agreement between the keyword co-occurrence network and the concept map generated using the Web of Science's Keywords Plus field implies that there is a clear division between the two research streams, despite the fact that the outputs from the first stream should be inputs for the optimization models used for prescriptive decision-making. Based on the second insight and a separate thorough examination of the relevant operations research (OR) literature we can then conclude that the prescriptive literature largely ignores the recent results on factors influencing crash risk.

Against this backdrop, the primary purpose of this review is to help bridge the gap between the different research streams that relate to the modeling and minimization of crash risk. Our goal is to bring the research into better focus and to encourage future work that crosses

Figure 3.1: A keyword co-occurrence network of the literature, depicting the 60 most used keywords. The nodes correspond to the keywords, with node size reflecting relative frequency. The links are limited to keywords that co-occurred at least five times (black and red lines correspond to between and within clusters, respectively). The network plot divides the literature into two clusters: prescriptive modeling (left), and explanatory/predictive modeling (right).

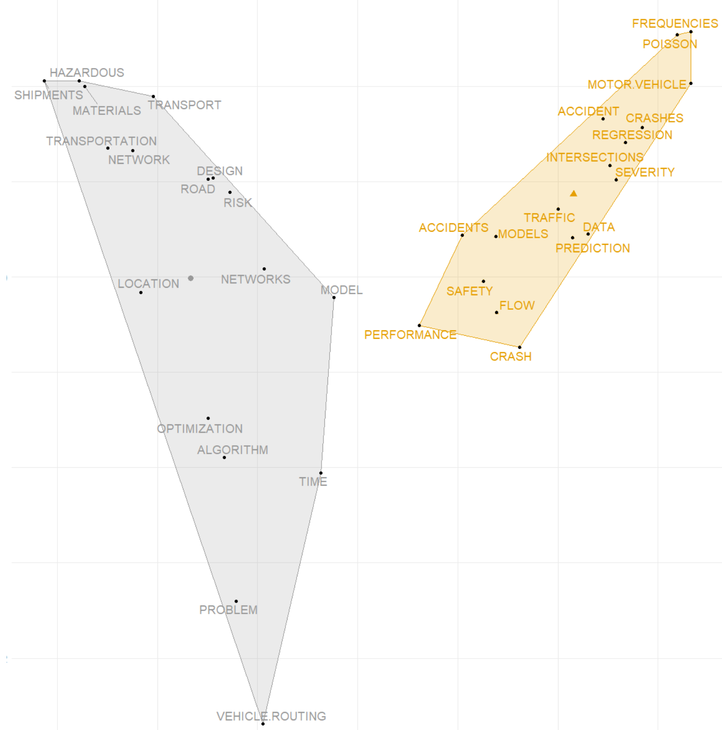


Figure 3.2: A data-driven conceptual structure map based on “Keywords Plus” (keywords tagged by the ISI or SCOPUS database scientific experts) and the application of multiple correspondance analysis and k-means clustering. The nodes are limited to keywords that have occurred ≥ 5 times, and the gray circle and orange triangle depict the corresponding cluster center. Similar to Fig 3.1, the concept map also divides the literature into the same two clusters.

the siloed divisions within the literature. To achieve this goal, we divide this review into two parts. Part 1 covers the sensing, data acquisition, data exploration, and explanatory/predictive modeling, i.e., focuses on the first research stream. Part 2 reviews the prescriptive modeling component (i.e., second stream), provides a simple case study for how both streams can be integrated, and presents ideas for future research. Note that the research presented in Part 2 primarily targets hazardous materials (hazmat) trucking operations, where optimization models are used to minimize crash risk through path/route selection and/or rest-break scheduling, while meeting delivery requirements. On the other hand, in Part 1, the research relates to both commuters and commercial drivers since the unit of analysis is a "road segment".

This paper is structured to follow the standard data analytics framework: data collection $\longrightarrow$ data exploration $\longrightarrow$ predictive modeling. The final part—prescriptive modeling—is discussed in Part 2 of this effort. We would like to emphasize that in addition to the need for connecting siloed research streams identified above, there also may exist a relatively high "start-up cost" for initiating new efforts in this area. Specifically, as we survey in the remainder of this paper, there exist multitudes of disparate datasets, data processing approaches and statistical methods that all may be relevant. Hence, the goal of this review is to attempt to reduce this burden by categorizing the existing efforts. The remainder of the first part of this review is structured utilizing a data analytic framework (data collection $\longrightarrow$ data exploration $\longrightarrow$ predictive modeling). We present an overview of the sensors and data collection mechanisms used in these studies in Section 3.2. In Section 3.3, we provide a taxonomy and review of the commonly utilized data exploration and summarization techniques. Then, we synthesize the explanatory/predictive modeling techniques used for crash risk modeling in Section 3.4. We offer our concluding remarks in Section 3.5, and provide links for our code and analysis in the Supplementary Materials Section.

3.2 Data acquisition protocols: An overview of the types of collected data and their associated sensing systems

In this section, we provide an overview of the data acquisition strategies typically used in motor vehicle safety studies as well as a brief to introduction to the corresponding sensing systems.

The ability to extract such data is an indispensable component in any crash risk prediction study, yet it is typically under-described. Thus, we view this section as an important practical contribution of our review since a potential reason for the gap between the predictive and prescriptive analytic research streams can be attributed to the "large start-up burden", associated with the lack of sufficient/targeted documentation for collecting quality data. While we primarily focus on U.S.-based systems, the protocols described here can be extended to many transportation locales. To facilitate and encourage the collection of data pertaining to important factor sets (per the reviews of Theofilatos and Yannis (2014) and Roshandel et al. (2015)) in future prescriptive studies, we provide **R** code that can be used to scrape data for many important crash risk predictors (see the link in our Supplementary Materials Section). '

It must be emphasized that both data sources needed and data acquisition methods used to access these sources depend on the design of the study in question. Specifically, since this review is focused on the literature dedicated to models for quantifying crash risks, the corresponding studies can generally be divided into two main study designs: (a) retrospective case-control studies in which police crash reports are used, and (b) prospective naturalistic driving studies (NDS), in which a pre-specified set of drivers is followed for a certain period of time. As one can expect, the choice of study design affects the data collection mechanism (as well as the statistical methodologies used for analysis, which are discussed in Section 3.4). For the sake of completeness, we provide some background on each of these two design strategies in the following subsection.

3.2.1 Background: Study designs

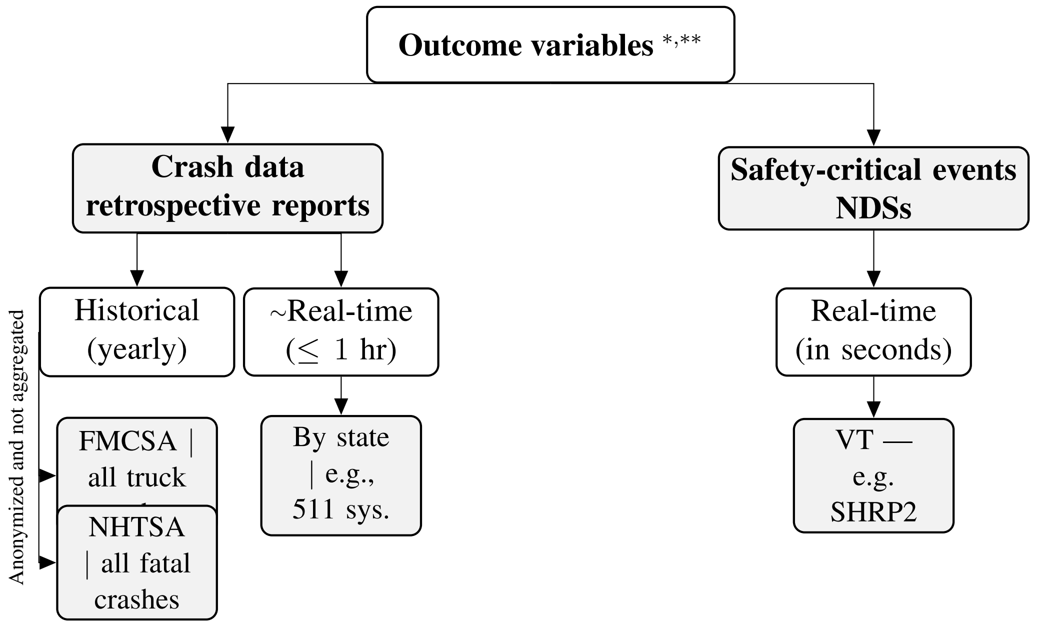
Most research on motor vehicle safety has assumed that the sampling unit is a spatiotemporal snapshot of a highway, i.e., researchers typically study a given section of a highway for a pre-specified time period. Note that it is not sufficient to study the conditions under which crashes tend to occur; one must also study the conditions under which crashes do not occur, and compare the two. The problem is analogous to that faced by epidemiologists when investigating the cause(s) of a disease, where they examine the prior behavior of individuals with and without a disease and attempt to identify differences in their prior behavior. The most common design that epidemiologists use is the case-control design. A number of individuals with the disease are first identified, representing the cases. The demographic and behavioral characteristics (e.g., age, sex, race, smoking status, body mass index, etc.) for the cases are then determined/computed. A control group, as similar as possible to the case group, is then identified. In a matched pair case-control study, each case is matched with one or more control subjects.

In motor vehicle highway safety applications, these retrospective case-control studies are typically conducted using police crash reports. In the U.S., crash reports include information pertaining to number of vehicles, involvement of pedestrians, number of injuries/fatalities, road type, crash location, date-time, intersection type, presence of a nearby work zone, weather conditions, and road surface conditions (Thiese et al. 2017; Newnam et al. 2019). While a lot of information can be captured in these reports, case-control studies are inherently limited for two main reasons. First, the information captured in the crash reports combines: (a) factual information, e.g., type of road and number of vehicles involved in the crash; (b) information that is estimated by the police officer, e.g., classifying weather into one of pre-defined categories; and (c) information captured from witnesses which is subject to recall and/or information bias, e.g., it is often hard to gauge the veracity of information extracted from drivers involved in the crash. Second, the inference from case-control studies can be limited when the denominator (e.g., non-crashes or healthy individuals) is unknown to the researchers (Dingus et al. 2011). In highway safety research, traffic flows can be captured using cameras and on-the-road sensors; however, such information is not typically available for every road segment (e.g., in rural local roads and/or for all highway exits). Thus, this is a prevalent issue in existing case-control highway safety studies.

To alleviate the limitations in case-control studies, there has been an increasing number of prospective naturalistic driving studies (NDSs) in the past decade. Contrary to the case-control studies, the information is captured via one or more sensors that are mounted in the vehicle in an effort to collect (Guo 2019): (a) high-resolution real-time driving data under real-world circumstances; (b) location/GPS, speed, and multiple views of the driver/road; and (c) naturalistic/individualized driving behaviors that can help explain differences if a crash is observed during the study period. Compared to traditional case-control studies, NDSs resemble prospective cohort studies, where a pre-specified set of drivers is followed for a certain period of time. The sampling units here are the drivers instead of road segments, and all the events or non-events of the sample drivers are collected. Therefore, it is possible to compare the rates of events in NDSs. In addition, the data are automatically collected using sensors, which minimizes the impact of police/witnesses' judgement in imputing the data and/or estimating values for certain predictors.

3.2.2 Outcome variables used in crash risk modeling

In retrospective case-control studies, the most frequently used outcome variable is crash counts. In the U.S., historical crash data are hosted by different Department of Transportation (DoT) divisions depending on: (a) the types of vehicles involved, i.e., commercial vehicles or personal commuter vehicles; and (b) whether the crash resulted in any fatalities. When these models are utilized/deployed for predictive purposes, real-time traffic data can often be used as model inputs. In the U.S., such data can be obtained from state specific reporting systems. For example, the 511 reporting system highlighted in Figure 3.3, is the predominately used sensing system in the U.S. since it is used by more than 45 states (Federal Highway Administration 2016). On the other hand, in prospective NDSs, the use of safety-critical events (SCEs) as a proxy outcome variable is more common since: (a) NDSs do not focus on crash-prone highways, (b) SCEs have a much higher incidence rate than crashes, and (c) they are assumed to be positively correlated with the incidence of crashes (Guo et al. 2010; Dingus et al. 2011). SCEs are defined as events that avoid crashes by last-second evasive maneuver(s) (Dingus et al. 2011). The most commonly studied SCE is "hard brakes", which can be detected using accelerometers/inertial measurement units mounted in the vehicle or through a driver's smart phone. The identification of a "hard break" is threshold dependent; for example, several studies equate a "hard break" to a deceleration higher than 3.0 m/s^2 (Jansen and Simone Wesseling 2018; Mollicone et al. 2019). Several detailed reviews have been published on surrogate indicators using in the field of traffic safety.Zheng et al. (2014); Johnsson et al. (2018); Mahmud et al. (2017). It is important to note that, while SCE has been extensively used as the outcome variable in NDSs, its validity and causal relationship with crashes have not yet been conclusively confirmed (Knipling 2015, 2017). We provide a visual summary of the hierarchical nature of the described outcome variables in Figure 3.3.



* Acronyms: FMCSA = Federal Motor Carrier Safety Administration, NHTSA = National Highway Traffic Safety Administration, VT = Virginia Tech.

** Code: To simplify the data collection process, we present the **R** code needed to scrape and clean these different data sources at: `https://caimiao0714.github.io/TrafficSafetyReviewRmarkdown/`.

Figure 3.3: A hierarchical view of outcome variables in crash risk modeling studies. The first level captures the data type, the second level shows the frequency, and the third level highlights examples and sources.

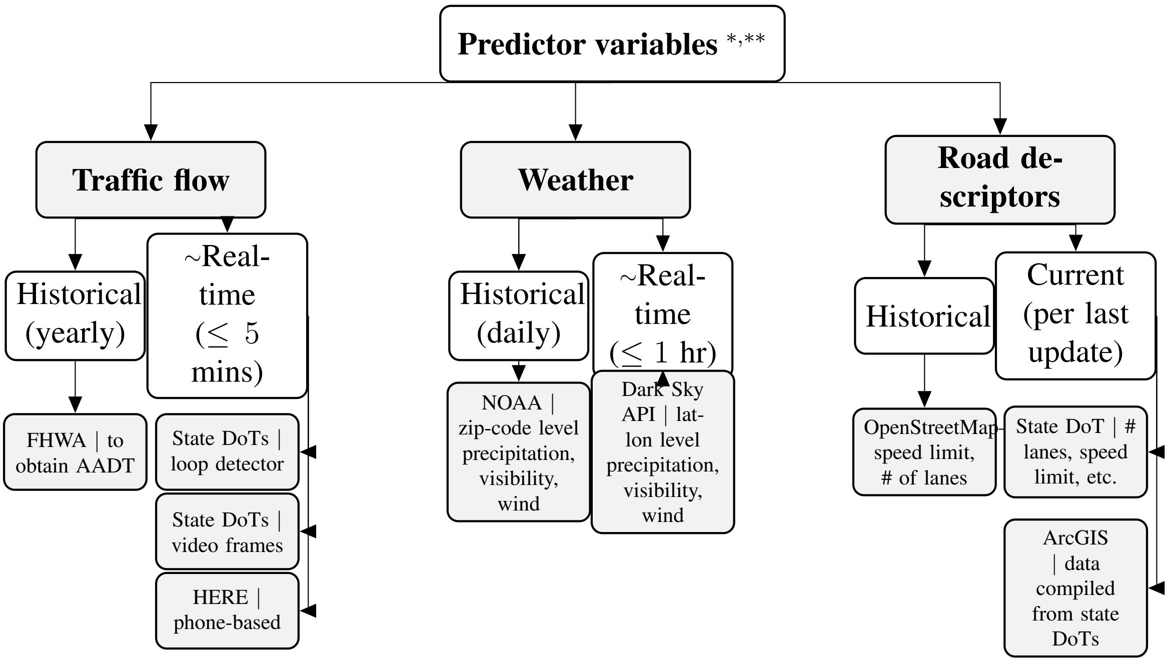
3.2.3 Predictor variables used in crash risk modeling

Factors that have been shown in the literature to contribute to motor vehicle crash risk are discussed in detail in Section 3.4. Here we concentrate on strategies and sensing technologies used to obtain relevant data.

From a data acquisition viewpoint, the sensors can be divided into (Guerrero-Ibáñez et al. 2018): (a) intra-vehicular sensing platforms, where conditions extracted from the vehicle are captured, and (b) urban sensing platforms, where the sensors are integrated in the road infrastructure. Intra-vehicular sensors can capture driver behavior, vehicle speed, traffic environment, etc. (Abdelhamid et al. 2014), and are widely used in NDS studies. On the other hand, urban sensing platforms are more commonly utilized in case-control studies.We can categorize such platforms into the following three categories: (a) traffic sensing systems (e.g., traffic cameras, inductive loop detectors, infrafred sensors), which can be used to estimate traffic flow, speed, occupancy, and volume (Guerrero-Ibáñez et al. 2018); (b) weather sensing systems, which can be used to compute/estimate important factors for both explanatory/predictive (e.g., visibility, rain/ snow accumulation, and potential for icy conditions) and prescriptive modeling (e.g., wind direction and speed which are important considerations in hazardous material routing since they are used in predicting the severity of a possible crash through estimating the radius of dispersion of toxic materials); and (c) geometric road descriptors (e.g., number of lanes, speed limit information, longitudinal grade, road shoulder width, and whether the road segment of interest contains a straight, merge, and/or diverge sections), which are typically tagged in geographic information systems (GIS) and can be accessed using popular application programming interfaces (APIs) such as *OpenStreetMaps* (Wikipedia contributors 2019; Eugster and Schlesinger 2013). A visual summary of predictor variables extracted from urban sensing systems is provided in Figure 3.4

3.3 Descriptive analytic tools used for understanding crash data

In this section, we review the exploratory data analysis (EDA) techniques used to examine transportation datasets prior to the explanatory/predictive modeling stage. EDA is an especially important pre-processing steps when dealing with large datasets, where predictive modeling and optimization can be computationally intensive. In Figure 3.5, we depict the two major goals of EDA as well as the methodologies used to achieve these goals. Note that these methods may not be mutually exclusive and can be used to complement each other.



* Acronyms: AADT = Annual Average Daily Traffic, FHWA = U.S. Federal Highway Administration, DoT = U.S. Department of Transportation, and NOAA = U.S. National Oceanic & Atmospheric Administration.

** Code: To simplify the data collection process, we present the **R** code needed to scrape and clean these different data sources at: `https://caimiao0714.github.io/TrafficSafetyReviewRmarkdown/`.

Figure 3.4: A hierarchy of predictor variables used in modeling crash risk. The first level captures the data type, the second level shows the frequency, and the third level highlights examples and sources.

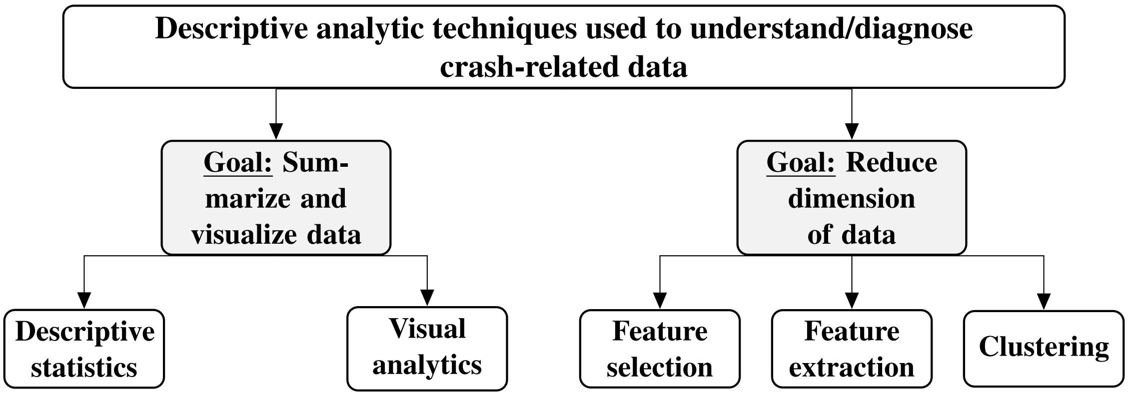


Figure 3.5: Exploratory data analysis (EDA) goals and their associated techniques/methodological frameworks.

3.3.1 Data summarization and visualization

Data summarization include both univariate (e.g., central tendency, dispersion, etc.) and multivariate tools (e.g., correlation). We assume that both predictive and prescriptive modeling researchers are well-versed with these methods, and thus we will not discuss them here (see Washington et al. Washington et al. (2010) for a detailed introduction). As a complement to data summarization, data visualization is a succinct approach to understanding trends, patterns, and anomalies in data. In a survey paper on the application of visualization techniques for traffic datasets, Chen et al. Chen et al. (2015) categorized visualization approaches based on four data types: (a) temporal data, (b) spatial data, (c) spatiotemporal data, and (d) multivariate data. This framework can be extended to more comprehensive crash modeling studies where traffic, weather and other predictor sets are combined. Table 3.1 presents an overview of the appropriate/recommended visualization techniques for each data type, with example references from the literature. In the following subsections, we discuss each of these groups in further detail.

Visualization of time-oriented data

Line graphs are the most frequently used visualization technique for time-oriented data, where the x-axis represents time and y-axis demonstrates transportation-related variable. There are numerous applications of line graphs in traffic/crash visualizations, for example, visualizations of tips per trip and fare per miles-driven among New York City taxi drivers (Ferreira et al. 2013), carbon monoxide pollution over the course of the day in London (Croxford et al. 1996), traffic volumes in Beijing, China (Han et al. 2006) and Porto, Portugal (Alam et al. 2017), or the effects of road surface conditions and time of day on traffic volumes (Nookala 2006). Since line graphs can become visually overwhelming as the number of variables increases. Other time-series based graphs can be considered in this case, such as *ThemeRiver stacked chart* (Havre et al. 2000), which uses a flowing river metaphor to capture changes in several variables of interest over time. This chart was used by Guo et al. (2011) for understanding traffic volume patterns.

Table 3.1: Categorizing visualization techniques for transportation data, adapted from Chen et al. (2015).

Variable type (Main group)	Subgroup	Visualization techniques	Examples
Time-series data	*Linear time*	Line and stacked graphs	Han et al. (2006); Nookala (2006); Guo et al. (2011); Ferreira et al. (2013); Tsai et al. (2015); Alam et al. (2017)
	Periodic time	Radial layout and cluster-and-calendar based visualization	Pu et al. (2013); Tsai et al. (2015)
Spatial	*Point-based*	Symbol maps	NHTSA (2018)
	Line-based	Line maps, edge bundling, and kernel density estimation charts (KDE)	Xie and Yan (2008); Lovelace et al. (2019)
	Region-based	Radial metaphor charts, choropleth, proportional symbol maps, and heat maps	Kraak (1999); Erdogan (2009); Wongsuphasawat et al. (2009); Liu et al. (2013); Zeng et al. (2013)
Spatiotemporal	-	Space-Time-Cube (STC), animated maps, GeoTime, and stacking-based STC	Kraak (2003); Kapler and Wright (2005); Romero (2015); Galka (2016); Tominski et al. (2012)
Multiple properties	-	Parallel coordinates plot, trellis plot, and multidimensional scaling	Pack et al. (2009); Wongsuphasawat et al. (2009); Cottrill and Thakuriah (2010); Pack (2010); Chu et al. (2014); van Huysduynen et al. (2015); Liu et al. (2017); Das et al. (2018)

When the data are inherently periodic or cyclic, three charts can be applied (Chen et al. 2015): radial layout, cluster- and calendar-based (where line graphs are used for showing cluster averages over time, and calendar-based charts are used to show cluster membership per day) Van Wijk and Van Selow (1999), and statistically derived charts. Pu et al. (2013) used the radial layout chart to depict traffic volumes in different days and times. Tsai et al. (2015) showed how the cluster- and calendar-based charts can be effective in understanding traffic flows in the state of Alabama. In their case study, they showed that the data exhibited eight distinct clusters of daily traffic volumes (at hourly intervals within each day). Two of the clusters were somewhat unexpected, where one captured game-day traffic for college football, and the other captured travel patterns around different holidays (including Fourth of July, Thanksgiving, and Christmas). Statistically derived plots (based on time-series analysis techniques) can be used to quantify the periodic/seasonal nature of the data. From a time-series analysis perspective, the data can be decomposed into: (a) seasonal, (b) trend, and/or (c) cyclical components within a season. These components can be visualized, along with the autocorrelation function (ACF) and the partial autocorrelation function (PACF) for the differenced series to provide an understanding of what type of time-series models to use. The reader is referred to Washington et al. (2010) for a detailed coverage of time-series modeling applied to transportation data analyses.

Visualization of spatial and spatiotemporal data

Crash datasets provide rich spatial information including the location of vehicles, construction sites, road closures, and crashes. Visualizing them spatially gives insight(s) on the geographical patterns and clusters, which may improve the decisions made when setting up the dataset for predictive/ prescriptive modeling. Chen et al. (2015) presented three visualization options (point-based, line-based, and region-based visualizations), which should be selected based on the dataset's aggregation level.

In point-based visualizations, each symbol on a map represents the position of an object at a given point in time. An example of such a visualization is the motor vehicle fatality symbol map, which is used by NHTSA to depict fatalities (NHTSA 2018). We provide a screenshot of their dashboard in Figure 3.6, showing the location of vehicle occupants killed in speed-related crashes on Saturdays in December, 2016.

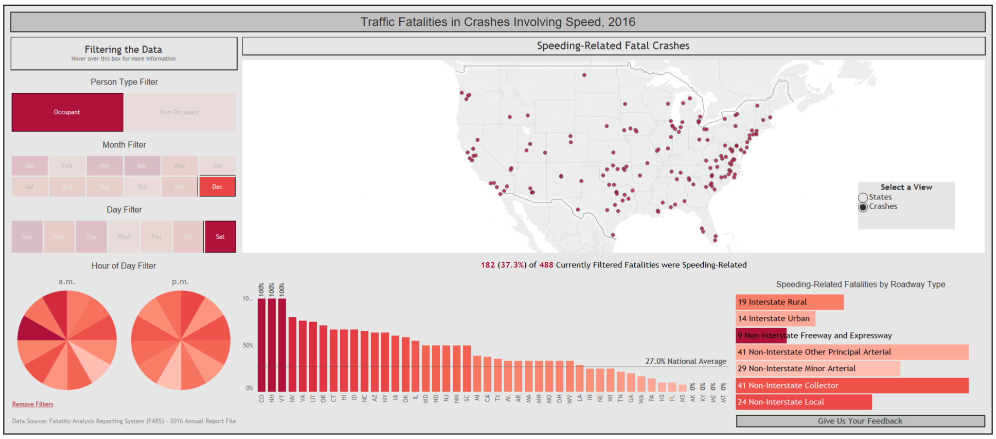


Figure 3.6: Symbol map showing the location of vehicle occupants killed in speed-related crashes in the US in December, 2016. The dashboard is available at (NHTSA 2018).

Popularized by the ubiquity of modern navigation applications, a line map visualizes travel routes and traffic flow. An example can be found at Lovelace et al. (2019), who presented the trip patterns in Bristol, England. They used the "line width" to encode the number of trips and "color" to encode active travel percents. Given the widespread use of navigation applications, we do not discuss other examples in this review.

Region-based spatial visualizations include three popular visualization techniques. The first is the "proportional symbols map" (Kraak 1999), where the size of a point/symbol in a map is proportional to the number of observations in that location. This can be seen as an extension to the point-based visualization, where the point-position on the map is now used to encode count. The second technique is based on "choropleth maps" (Erdogan 2009; Wongsuphasawat et al. 2009; Liu et al. 2013), where areas/regions in maps are shaded, colored, or patterned relative to the value of the metric of interest. These maps are common when comparing crash/fatality rates between larger geographic regions (e.g., counties, states, or countries). The third, and least commonly used visualization is the "radial metaphor“. One existing application was provided by Zeng et al. (2013), who used a "radial metaphor" chart to visualize interchanging traffic patterns among different regions of a city.

For spatiotemporal visualizations, there are two overarching strategies that can be used. The first strategy is intended for web-based visualizations, where a time effect is added to the map by animation or transition effects. Examples can be found in Romero (2015) and Galka (2016). On the other hand, the second strategy is intended for print and utilizes dedicated visualization methodologies. Space-time-cube (STC) visualizations, are the most commonly utilized approach, where the x and y axes are used to capture spatial information, while the temporal information is shown on the z axis (Kraak 2003). Applications of such technique include: (a) traffic analysis where the changes in a traffic-related variable of multiple vehicles across time and space is shown by stacking-based STC (Tominski et al. 2012); and (b) crash analysis where crashes are displayed and tracked based on their spatiotemporal information by an enhanced version of standard STC (Kapler and Wright 2005; Gudes et al. 2017). Despite their perceived utility for showing spatiotemporal patterns in a 2-dimensional screen/paper, we do not recommend this approach since the actual values cannot be easily shown and comparisons depend on one's ability to estimate the patterns over space and time. Instead, we would recommend the use of either panel visualizations (i.e., trellis/ small multiples), or a tabulated representation of the results to show the time component.

Visualization of high-dimensional datasets

For high-dimensional data, visualization requires more data cleaning and curation. On the lower end of the spectrum, Parallel coordinates plots (PCP) and trellis (small multiples of bar charts or scatter plots) are commonly used fast plotting tools and require less data preprocessing. For example, PCP can be applied to visualize the correlation/interaction among several crash descriptors including: cars involved, day/month effects, incident type, and road condition Pack et al. (2009); Wongsuphasawat et al. (2009); Pack (2010). Additionally, the trellis plot was used by Cottrill and Thakuriah (2010) to visualize variations in the number of crashes by different census tracts. On the upper end of the analytical spectrum, visualizations are preceded with the application of projection methods to reduce the problem's dimensionality. Examples include: (a) van Huysduynen et al. (2015) where cluster analysis and multidimensional scaling were used to produce a 2-dimensional (2D) plot of the relationship between the different constructs and types of drivers examined in the study; (b) Das et al. (2018) who utilized multiple correspondence analysis (MCA) to present a proximity map of key factors contributing to wrong-way driving in a 2D space; (c) Liu et al. (2017) where the multivariate time-series data capturing the driver behavior were reduced to a 3D feature space using deep learning techniques, and then visualized using a driving color map.

3.3.2 Dimension reduction

In the previous subsection, we highlighted how projection methods can be used to reduce the data dimensionality and assist in its visualization. Here, we discuss how dimension reduction techniques can be used to prepare the data for the predictive modeling stage. In general, there are three main goals for dimension reduction: (a) feature selection, where important variables are identified and selected; (b) feature extraction/generation, where the variable set is projected into lower subspace without losing significant information and; and (c) clustering, where similar observations are grouped together. Since researchers could combine these approaches in their analysis, we classified dimension reduction methods according to their goals.

Feature selection

One of the recommended steps before the use of statistical and machine learning models is to identify and use only the variables/features deemed important for the analysis since this (Sawalha and Sayed 2006): (a) avoids over-fitting, (b) reduces the computational complexity in the analysis, and (c) leads to better prediction performance. This step is often referred to as variable or feature selection. In the context of crash prediction models, variable selection plays an important role since there are many potential predictors (e.g., traffic, weather, road geometry related variables) which may have effect on the probability of a crash. In addition, in order to capture the spatial and temporal effects of these variables, new variables need to be introduced in the model. For instance, Shi and Abdel-Aty (2015) developed a crash prediction model where each traffic-related variable is collected prior to the crash from two upstream and two downstream sensors. This means that the information for each traffic variable is divided across four variables, and that these variables contain some redundant information within them. In such cases, feature/variable selection will improve model performance (Hassan and Abdel-Aty 2013; Hossain and Muromachi 2013; Yu and Abdel-Aty 2013; You et al. 2017; Basso et al.

2018). For the sake of conciseness, hereafter we use the term feature selection to denote feature and variable selection methods.

Feature selection methods can be classified into three groups: filter, wrapper, and embedded methods (Chandrashekar and Sahin 2014). In the filter methods, the process of selecting a subset of features is independent from the statistical and machine learning model used, i.e., a subset of features will be selected according to an algorithm (e.g., Pearson correlation or Mutual Information Criterion), and then the selected features will be the inputs to the explanatory/predictive model. Advantages of filter methods include: (a) simplicity, (b) computational efficiency, (c) speed, and (d) reduction of the risk of over-fitting. However, they can ignore the dependency between features and do not guarantee the selection of an optimal set of features (Saeys et al. 2007; Chandrashekar and Sahin 2014). In contrast, wrapper methods consider the prediction performance of the classifier (while accounting for the dependencies/interactions between features) and subsets the feature space using heuristic searching algorithms such as genetic algorithms (Goldberg and Holland 1988) and particle swarm optimization (Kennedy and Eberhart 1995). While they can improve performance when compared to filter methodologies, they are computationally inefficient. In addition, they also do not guarantee optimality and may over-fit (Saeys et al. 2007; Chandrashekar and Sahin 2014). To avoid such problems, feature selection is a part of the model training process in embedded approaches, which makes them the preferred approach in many crash risk modeling scenarios. Random forest (RF) was widely used in the literature as a feature selection method and to determine variable importance (Xu et al. 2013a; You et al. 2017; Basso et al. 2018). For more information about the feature selection methods and their applications, we refer the reader to Guyon and Elisseeff (2003); Saeys et al. (2007); Jović et al. (2015).

Feature extraction

Feature extraction methods offer an alternative approach to dimension reduction by projecting input space to a more efficient dimension space. The projection can combine input variables, reduce the problem complexity, and present a useful abstraction of the data (Khalid et al. 2014).

Thus, feature extraction differs from feature selection as the focus is not on dropping unimportant variables, but rather to combine the information across the variables through a mathematical transformation. Principal Component Analysis (PCA) is the most commonly used feature extraction method in the crash prediction literature (Nagendra and Khare 2003; Lee et al. 2003; Li et al. 2007; Caliendo et al. 2007; Guo and Fang 2013; Lee et al. 2018). Through an orthogonal transformation, PCA transforms the original variables into a set of linearly uncorrelated variables (i.e., principal components, PCs). Typically, the variation in the data can be explained with a few PCs, which reduces the dimensionality of the problem with minor loss of information. The determination of the number of PCs to retain is often determined through a scree plot or a threshold for the eigenvalues (Cook 2018). Since PCA was originally designed for numeric variables that can be linearly combined, there are several extensions to PCA which do not require such assumptions. These include: (a) probabilistic PCA Tipping and Bishop (1999), (b) non-linear PCA (Khalid et al. 2014), and (c) kernel-based PCA (Schölkopf et al. 1997). These methods have also been implemented extensively in the literature (Khalid et al. 2014).

Clustering

Contrary to feature selection and extraction, clustering is an unsupervised machine learning method that attempts to group observations together with the goals of maximizing the similarity within a cluster (i.e., minimizing distance between observations) and minimizing the similarity between clusters (i.e., maximizing the distance between cluster centers/centroids) (Berkhin 2006; Rai and Singh 2010). Clustering approaches can be divided into: partitioning-based, hierarchical-based, density based, grid-based, and model-based methodologies (Berkhin 2006; Fahad et al. 2014).

Crash risk modeling datasets have a number of characteristics that make clustering a viable and useful approach for dimension reduction. For example, if you consider traffic datasets, the goal is typically to understand the impact of traffic conditions on crash likelihood, which is typically achieved through: (a) classifying traffic into different states, and then (b) evaluating the impact of each traffic state (e.g., congested or not congested) on the crash likelihood

(Theofilatos and Yannis 2014). Historically, step (a) was achieved through an analysis of traffic flow characteristics (e.g., see (Hall et al. 1993; Kerner and Rehborn 1996; Wu 2002)). A limitation of such an approach is that the modeling can be influenced by researchers' biases and perceptions. Alternatively, one can use an assumption-free, data-driven approach to identify how observations can be clustered. Tsai et al. (2015) showed how clustering can be used to identify logical, but hard to model, groupings of the data. Applications of clustering include, but are not limited, to: (a) traffic categorization (Golob and Recker 2004; Xu et al. 2012; Tsai et al. 2015), (b) identifying accident clusters (Steenberghen et al. 2004; Xie and Yan 2013; Shen et al. 2019), and (c) grouping of weather conditions (Kwon and Park 2016). To demonstrate how an optimal number of clusters (k^*) can be obtained, we provide a detailed example in the supplementary materials where we use $k-$means clustering and the elbow method to determine the k^* clusters for traffic data.

3.4 Explanatory/predictive models for crash risk

This section focuses on two aspects: the risk factors that affect crash risk and statistical/machine learning models. In the risk factors part, we specifically consider the effects of fatigue, distracted driving, and environmental variables including traffic and weather on traffic safety. For the statistical part, we will review how some of the research that has been done to analyze those factors and build predictive models.

3.4.1 Risk factors for traffic safety

Roshandel et al. (2015) discussed five sets of factors that affect crash risk: (a) behavioral characteristics of the driver—e.g., impairment, fatigue, distractions; (b) vehicle condition; (c) traffic conditions—e.g., traffic speed, density and variation in speed between vehicles; (d) geometric characteristics of the road, i.e., type of road, number of lanes, curvature, nearby ramps/intersections, etc.; and (e) weather conditions—e.g., rain, visibility, ice/sleet/snow, etc.

Sleep and fatigue

Early work on the study of fatigue and the risk of adverse outcomes such as crashes relied on sample surveys of drivers. For example, Crum et al. (2001) conducted face-to-face interviews with approximately 500 truck drivers at five rest stops on interstates spread across the United States. The three outcomes were "close calls," "perception of fatigue," and "crash involvement." All of these were based on driver recall from survey responses. They identified three sets of variables that could affect drivers' fatigue, with self-reported measures. These measures included truck driving environments, economic pressures, and carrier support for safety. Three specific variables, all from the truck driving environment category, were identified as influencing fatigue, including: (a) drive regular or irregular shifts; (b) short or long load wait time; and (c) start the work week tired (or not). Crum et al. (2001) ran a regression analysis with these factors as predictors, with each of the responses described above. The first variable (drive regular or irregular shifts) was measured by determining how many six-hour times periods the drivers routinely drove. They found that starting the work week tired was a significant predictor for all three outcome measures described above. Long wait times were positively associated with close calls and self-perception of fatigue. Paradoxically, the number of time periods driven per day was negatively associated with close calls.

In another early study, Crum and Morrow (2002) conducted a stratified sample of trucking companies based on their safety record. They selected a sample from each of three strata defined as the bottom quartile (poorest safety performers), the middle two quartiles, and the top quartile (the highest safety performers). After taking a sample of carriers within each stratum they sent seven questionnaires to be filled out by various employees in the company, including the executive, the safety director, two dispatchers and three drivers. They also arranged focus groups within each company. Using the same three sets of variables as in Crum et al. (2001) they concluded that the most significant variable in predicting fatigue was "starting the work-week tired." Other significant factors were "difficulty finding a place to rest" and "shipper and receiver scheduling practices and requirements."

Garbarino et al. (2016) conducted a cross-sectional study of truck drivers in Italy to determine the risk factors for accidents and near misses. Data on sleep apnea, sleep debt, daytime sleepiness, frequency of naps, and frequency of rest breaks, as well as the accident responses were conducted from survey questionnaires and medical exams. They found that obstructive sleep apnea, sleep debt, and excessive daytime sleepiness were positively correlated with accidents; these yielded odds ratios of 2.32, 1.45, and 1.73, respectively. Naps and rest breaks were negatively associated with accidents, having odds ratios of 0.59 and 0.63 respectively. All of these odds ratios had confidence intervals that excluded the null value of 1.0.

With automatic data collection systems that can detect events like accidents, hard-breaks (sudden deceleration caused by braking), lane departures, and others. Mollicone et al. (2019) studied hard braking as safety critical events, which are highly correlated with crashes (Dingus et al. 2006). Their model used a predicted fatigue model of McCauley et al. (2009) and McCauley et al. (2013) to develop a Poisson regression model having the number of hard brakes as the response. The predictor variables included the predicted fatigue and six variables for the time of day. They found that there is an increasing and concave up relationship between the predicted fatigue and the relative risk of a hard brake.

In a recent study, Stern et al. (2019) reviewed the research related to fatigue of commercial motor vehicle drivers. Because of the difficulty of running a controlled experiment by imposing treatments, most research designs are observational studies, that is, they compare the effects of variables that are observed, not imposed. One exception to this is a *randomized encouragement design* where drivers are randomized to receive some sort of incentive to apply some treatment, but are not forced to do so. If an effect is observed, we would conclude that it is due to the incentive, not necessarily to the actual treatment. Many studies use a cohort design or a case-control study. In a cohort design, a number of drivers is identified and studied across time. In a case-control study, a number of cases (e.g., crashes) are identified and are matched with controls; focus is then placed on the differences between the cases and controls. Both cohort studies and case-control studies can be useful in assessing safety.

Recently, Bowden and Ragsdale (2018) developed an optimization algorithm for driver scheduling. The algorithm, denoted FAST (Fatigue Avoidance Scheduling Tool) was designed to minimize the trip duration subject to a minimum fatigue level along with other constraints, such as the maximum driving hours under United States law. The algorithm assumes the three process model of alertness (TPMA) developed by Åkerstedt and Folkard (1995) and Åkerstedt et al. (2004).

Distracted driving

Other researchers have looked at the effect of distracted driving. The problem of mobile phone usage and distracted driving has been noticed by the World Health Organization (Organization et al. 2011). They noted that world-wide use of cell phones has increased by up to 11% in the past 5 to 10 years. Their data suggest that cell phone usage increases the chance of a crash by a factor of four, and this is similar for hand-held phones and hands-free devices. Young et al. (2007) noted that at the time, about one fourth of all crashes (trucks and personal vehicles combined) were due at least in part to distractions, particularly mobile phones and navigational systems. They reviewed much of the literature available at the time of their writing. Wilson and Stimpson (2010) reviewed trends in distracted driving accidents and noted that deaths due to distracted driving had increased 28% from 2005 to 2008 when the rate was nearly 6000 deaths per year.

Olson et al. (2009) studied distracted driving in 203 commercial drivers. The data involved 4452 critical events, such as crashes, near-crashes, and unintentional lane departures, along with 19,888 time periods that involved no special events. The found that 71% of all crashes and 46% of near crashes involved drivers who were engaged in tasks not related to driving. Overall, 60% of critical events occurred while the driver was performing non-driving tasks. Klauer et al. (2014) conducted a study in which 42 young drivers (16.3 to 17.0 years of age) who had just received their driver's license and 109 experienced drivers were studied. Here the unit of measurement is the driver. Equipment, such as accelerometers and cameras, were used to detect distracted motion while driving. They found that distracting events like eating or cell phone dialing or texting led to an increased risk of accident, with odds ratios often exceeding 3.0.

In terms of safety optimization, the choice here is clear. Distracted driving, such as hands-on cell phone use and texting, should not be allowed. From a general public perspective, these have translated into driving laws in many countries as well as have been translated into company policies for many commercial transportation firms. In addition, there are several smart phone-based applications that disables texting while driving and/or encourage safe driving behavior. From a commercial driving perspective, there are wearable technologies (e.g., headsets embedded with sensors that are linked to a smartphone application) that are used by professional drivers that provide voice-alerts when their mirror-check rate deviates from a pre-set standard. This information is also shared with dispatchers to schedule rest-breaks as an intervention. While these smart-phone applications/technologies seem promising, there is not a large body of literature that examines the effectiveness of these interventions.

Weather, traffic conditions, and road geometry

In Sections 3.4.1 and 3.4.1, we have discussed driver-related factors. In many cases, the crash likelihood and severity can be impacted by non-driver/external factors. Variables/features capturing weather (e.g., temperature, precipitation, wind speed, humidity, and visibility), traffic conditions (e.g., traffic flow, occupancy, density, and volume), and road geometry (e.g., elevation, curvature, road surface, and the number of lanes) represent the main external factors that impact the crash likelihood and severity (Xu et al. 2013b; Theofilatos and Yannis 2014; Roshandel et al. 2015). Note that these factors should not be considered in isolation since their interactions are complex and can significantly change the crash likelihood. Thus, in this subsection, we highlight three relevant studies that have investigated the combined effect of such factors on crash risk.

Ahmed et al. (2012) investigated the effect of the interaction between road geometric features, real-time weather parameters, and traffic data on crash likelihood. Using a Bayesian logistic regression framework, the authors developed two models for snowy and dry seasons. Based on their models and case study, their results showed that in, both models, the main effects and at least one interaction term were significant. The authors showed that the crash risk during the snowy season was two times that of the dry season. Furthermore, the authors suggested that the crash risk likelihood may also be influenced by the interaction effects between the snowy, icy, or slushy road surface conditions with road segments involving steep grades.

In another study, Yu et al. (2013) conducted their study on a 15-mile segment of the I-70 interstate in Colorado. The authors utilized: (a) 30 Remote Traffic Microwave Sensor (RTMS) sensors to extract real-time traffic data; (b) six weather stations for obtaining real-time weather data; and (c) the Roadway Characteristics Inventory (RCI) for obtaining descriptors of road geometry. Different scenarios were considered in the study based on the season and crash type. The results showed that the adverse weather condition combined with critical roadway conditions (e.g., steep slopes) can increase the crash likelihood significantly. Further, single vehicle (SV) and multiple vehicle (MV) models shared some common significant predictors such as precipitation and average speed. Furthermore, in the SV model, the significant variables were more related to weather conditions and vehicle speed. On the other hand, MV crashes were more affected by traffic-related variables.

Wang et al. (2019a) studied several of the factors that could lead to high risk traffic conditions. They considered traffic, weather, road geometry, and some behavioral aspects, such as trip generation and social demographics. These variables were taken as the characteristics of the region surrounding the crash, not the individuals involved in the crash. They used a case-control design with a 10:1 ratio of non-crashes to crashes. They used support vector machines (SVM) for variable selection and Bayesian logistic regression for inference. They found that the percentage of home-based work production, which includes commuters, was the only behavioral characteristic that had a significant effect on the risk of accident.

Xu et al. (2013b) developed crash prediction models at different levels of crash severity. Three levels of crash severity were considered: fatal/incapacitating injury crashes (KA), non-incapacitating/possible injury crashes (BC), and property-damage-only crashes (PDO). Results showed that under different crash severity levels, the effect of environmental variables is different. For example, in the all crashes model (KA, BC and PDO), adverse weather conditions would increase crash risk. However, under the injury crashes model (KA and BC), adverse weather conditions had the opposite effect which indicated that it could possibly reduce the likelihood that a crash would result in injuries and fatalities (possibly due to uncaptured changes in driver behaviors). Note also that the significant traffic-related variables are different in these two models which indicates that the interaction of the external variables would result in different level of crash risk and severity.

3.4.2 Statistical modeling

Retrospective case-control studies are usually analyzed using logistic regression or other classification models. Since crashes are very rare compared with non-crashes, matching a case (crash) with one or more controls (non-crashes) is considered; this matching is then accounted for in the analysis. In other situations, non-crashes are unmatched; a set of controls is selected to mimic the aggregate of conditions of the crashes. Many studies are unclear about matching and whether (and also how) the matching was taken into account in the analysis. Since non-crashes are much more common than crashes, it is common to take several times as many non-crashes as crashes. Ratios as high as 10:1 are found.

Theofilatos and Yannis (2014) surveyed previous research on the relationships between these factors and traffic crashes. Some of the commonalities in the conclusions were that safety was a nonlinear function of traffic flow, speed limits were a factor, and precipitation was related to accident frequency, although the effect on severity is unclear. Roshandel et al. (2015) conducted a review and meta analysis of previous traffic safety studies and found that four variables are likely contributors to accident likelihood. These include speed variation around the crash site (odds ratio = 1.226), speed difference (odds ratio = 1.032), average traffic volume (odds ratio = 1.001) and average speed (odds ratio = 0.952).

Shi and Abdel-Aty (2015) used a matched control design to study rear end crashes. They matched 243 crashes with 962 non-crashes, a ratio of about 4:1. They used a random forest for variable selection, and then used a Bayesian approach for logistic regression. They found that peak hour, high volume upstream (from the accident), low speed downstream, and high congestion index downstream were significant factors for rear end crashes. Pande and Abdel-Aty (2006a) also studied exclusively rear end crashes in an unmatched case-control study. They found 2179 rear-end crashes, but only 1620 with full data, in a period of five years and selected a random sample of 150,000 of the roughly 363 million possibilities for the controls. They used classification and regression trees (CART) to discriminate low and high risk situations. Their approach could classify a situation as high-risk about 75% of cases where there was an accident, with approximately a 33% positive rate. Since crashes were rare events, one can conclude that their false positive rate was $\approx 33\%$.

In a later study, Pande et al. (2011) studied rear end crashes in a case-control study. They used a 5:1 ratio of non-crashes to crashes and used a random forest for variable selection and a multilevel perception neural network for inference. They found that occupancy downstream and average speed upstream were significant.

Theofilatos et al. (2018) studied traffic safety on a multi-lane belt-line highway in Athens, Greece, where there were 17 crashes and 91,118 non-crashes. In one model they use all data, and in another model they use a random sample of the non-crashes. They assume a logistic regression model and in one model they use a penalized maximum likelihood approach, called the Firth method, which uses all of the data. In another approach, they use a bias correction method to estimate parameters in the logistic regression, and for this they use a subset of the data. They find that average speed has a negative effect on crashes. The proportion of trucks on the road was considered but not found to be significant.

Lin et al. (2015) studied traffic safety on a corridor of Interstate 64 in Virginia, USA. Their study used a matched case-control design. They propose a frequent pattern (FP) tree which they use for variable selection. For inference on which variables are significant they use a k nearest neighbors algorithm and a Bayesian network. They conclude that the "accident risk prediction models based on FP tree variable selection outperform the models based on all variables ..." They also suggest using 10-minute intervals is more efficient than 5-minute intervals. Finally, they conclude that the Bayesian network model works well, yielding a false alarm rate of 0.38 and a sensitivity of 0.61.

Sun and Sun (2015) used a matched case-control design with a ratio of 5:1 to implement a Markov model involving the traffic states upstream and downstream. For example, if one upstream and one downstream segment is considered, then an expressway segment may be in the state FF (free flow upstream and free flow downstream); this leads to a four-state Markov chain. They also consider two upstream and two downstream conditions, leading to a nine-state

Markov chain. The transition probabilities were estimated using a dynamic Bayesian network model. Their model with nine states had a crash accuracy of 0.764 with a false alarm rate of 0.237. In addition to their work on the Bayesian network, they found an interesting nonlinear relationship between speed and risk, which they show in the second figure of their paper.

The effect of weaving, that is, traffic entering the expressway and merging while other traffic is exiting, was studied by Wang et al. (2015) in a case-control study of 125 crashes and 1250 non-crashes, a 10:1 ratio. They applied a multilevel Bayesian logistic regression model with weaving segments (that is, sections of the expressway where entering and exiting traffic had to merge) as random effects. These random effects were incorporated into the model as random intercepts. They found that the speed at the beginning of the weaving segment, difference in speed between the beginning and end, and the log of traffic volume were significant effects in these weaving segments. Wang et al. (2017) approached the traffic safety problem from two perspectives. One involved the crash frequency. This took as the sampling unit a section of the expressway and the number Y_i of accidents as the response. The other approach applied the usual logistic regression, taking the sampling unit as an expressway/time period slice and the indicator variable y_{ij} which is 1 for crash and 0 for a non-crash. The first approach leads to Poisson regression and the second approach leads to the usual logistic regression. The innovative contribution of their method is to combine, or integrate, these two models. This effectively uses two sources of data. Their integrated model includes the Poisson rate in the logistic regression model, yielding a multi-level model. They find that the integrated model performs better yielding a higher receiver operating characteristic (ROC) curve.

There are a lot of aspects of crash prediction models that can be studied, including model setting, specification, and validation, but those are beyond the scope of this review. Details of statistical models can be found in previously published reviews by Lord and Mannering (2010); Mannering and Bhat (2014); Abdulhafedh et al. (2017); Ambros et al. (2018); Yannis et al. (2016).

3.5 Conclusions

Given the tremendous loss of life and property directly attributed to motor vehicle incidents on one hand, and significant advances in relevant data availability on the other, it is natural that data analytics is viewed as having great potential for contributing to solving these problems. A successful effort in this direction necessarily has to rely on a combination of data collection, descriptive analytics, predictive/explanatory modeling, and optimization. At the same time, each piece separately can be a significantly nontrivial problem on its own. Hence, development of a mature data-driven decision support tool incorporating all of these stages "from scratch" is probably beyond the scope or ability of any single researcher. This is especially true since there is not a conscious effort in pulling all of these areas together with the goal of informing practical decision-making. The most significant gap that we have identified, is in the translation of outcomes/insights from predictive/explanatory models (which aim to help us better understand and quantify crash risk) into prescriptive optimization models (which aim to inform route/path selection, driver assignment, etc.). Perhaps, a partial underlying reason is the absence of readily available convenient data sources and/or data processing tools.

In this review, we highlighted a promising opportunity to develop advanced analytical methods for safety-enabled transportation. The following areas represent the main avenues for progress (ordered according to the sections in this review):

(A) The availability of historical, real-time and forecasted weather and traffic data, as well as the potential to collect driver performance data, means that the accessibility of data is no longer a major factor preventing progress in this area. However, a lack of a unified repository and the reluctance of sharing code/models by our research community leads to a fairly high overhead cost of developing such models (since every researcher has to develop many data collection techniques from scratch);

(B) Descriptive analytics tools are widely used in the preprocessing of driving-related data. Since the applicability of a particular preprocessing technique (e.g., visualization and clustering) often depends on the specific problem, the challenge here is to determine which method is the most suitable. Sharing best practices by creating reproducible documents (e.g., R Markdown and Jupyter notebook) represents one avenue for making the process more efficient for researchers and practitioners alike.

(C) Statistical methods for risk evaluation are well-researched and consider a wide range of factors. At the same time, it must be noted that (in some cases) these studies follow a similar pattern of a case-controlled study based on a single road segment data. In our view, there is an opportunity for a statistical analysis of a larger scale since:

   (i) real-time or near-real-time data are more widely available now;

   (ii) the computational advancements in the recent years can allow for parallelizing/computing risk across the entire road network or at the very least for all major highways and interstates;

   (iii) the insights from these relatively small road segments may not be generalizable to the entire road network; and

   (iv) it is unclear how drivers (regular commuters or commercial) can utilize these insights to make more informed decisions about their time-of-travel, path and/or route selection.

In our estimation, a more comprehensive/interdisciplinary approach to crash risk modeling is needed. The research questions should not be limited to only better understand the factors contributing to crash risk, but to also consider how the output from the research can be utilized by commuters and commercial drivers. This is especially important since, despite the technological advancements in sensing technologies and development of public policies that tackle distracted driving/ cell phone usage, the rate and counts of motor vehicle injuries and fatalities have remained alarmingly high.

3.6 Supplementary materials

In an effort to bridge the gap between the crash prediction literature and the hazmat/optimization literature, we have made all our source code used for (a) scraping crash-related data, (b) preprocessing of such-data; (c) descriptive analytics (i.e., visualizing traffic/weather/crash data and/or clustering); and (d) explanatory modeling available on a GitHub repository https://github.com/caimiao0714/TrafficSafetyReviewRmarkdown. To facilitate the consumption of this code, we host a website showcasing how the code can be used and depicting some of its results. The website was constructed using an R Markdown file (Xie et al. 2018), which is stored on the following GitHub Page https://caimiao0714.github.io/TrafficSafetyReviewRmarkdown/. We hope that the supplementary materials provided in this manuscript help promote "open data science" practices in our research community.

Chapter 4

Predicting unsafe driving events among commercial truck drivers: Lessons learned from the surveillance of 20 million driving miles using IoT sensors

4.1 Introduction

Road crashes are a widely-recognized global public health issue. According to the World Health Organization (2020), road injuries result in 1.35 million annual deaths and an estimated 20-50 million non-fatal injuries. Moreover, road crashes cause massive societal economic losses, which can be divided into (Wijnen and Stipdonk 2016): (a) human costs, (b) medical costs, (c) production losses, (d) property damages, (e) administrative costs, and (f) environmental costs. The World Health Organization (2018c) estimates that the total societal cost of road injuries corresponds to approximately 3% of a given country's gross domestic product (GDP). In the U.S., it is estimated that the total value of societal harm from vehicle crashes exceeds $830 billion annually (Blincoe et al. 2015; Wijnen and Stipdonk 2016), corresponding to 6% of U.S. GDP (Wijnen and Stipdonk 2016).

Large commercial vehicles (e.g., large trucks) are often involved in the most severe crashes. Large trucks and buses account for only 4% of registered vehicles in the U.S. (IACP 2018); however, they are involved in 13.2% of fatal crashes (FMCSA 2020). More alarmingly, the crash rates of such vehicles have increased over the past decade or so despite the technological safety innovations. In 2008, the fatal crash involvement rate of large trucks has increased from 1.32 to 1.48 per 100 million miles driven between 2008 and 2016 (NHTSA's National Center for Statistics and Analysis 2019). In addition, the numbers of large trucks involved in fatal, injury, and property damage crashes have increased by 1%, 5%, and 14%, respectively, from 2017 to 2018 (FMCSA 2018) – the most recent year where statistics are available.

In our estimation, the existing research dedicated to understanding road crash (and/or trucking) risk factors and their precursors can be divided into three major streams: (a) crash frequency models, (b) real-time crash prediction, and (c) naturalistic driving studies (NDSs). The overarching goal of *crash frequency* studies is to identify factors affecting crash risk such that appropriate interventions can be made (e.g., changing the design of a road/intersection). These studies collect data (e.g., crash count, weather conditions, annual average daily traffic (AADT, etc.) pertaining to a predetermined road segment over a relatively large time period (often measured in months), and then a statistical modeling technique is typically used to model the association between these factors and crash risk/frequency (see Lord and Mannering 2010; Saeed et al. 2019; Ziakopoulos and Yannis 2020, for an overview of possible modeling techniques). With recent advances in sensing and information technologies, more granular data on weather and traffic conditions are now available and can be processed in near real-time by existing computing infrastructure. Hence, studies in the second stream have attempted to identify/investigate factors that have immediate effect (next 10-30 minutes) on crash likelihood and then prescribe immediate preventive actions (e.g., variable speed limits) (see, among others, Theofilatos and Yannis 2014; Shi and Abdel-Aty 2015; Mehdizadeh et al. 2020). In these studies, the unit of analysis is a short roadway segment, where the collected data include real-time weather and traffic as well as cross-sectional data describing roadway geometry and type. Statistical and machine learning models are widely used to predict crash likelihood over a relatively short time span (often measured in minutes). With the choice of road segment as the unit of analysis, the first two streams are limited since: (a) they do not consider the effect of drivers' characteristics/behaviors (e.g, sleepiness, distraction and/or fatigue) on crash risk; (b) the number of observed crashes is typically small, which limits the power and inference of the developed models; and (c) the conclusions from these models may not be generalized to other road segments where traffic patterns, weather conditions and/or state policies are different.

Naturalistic driving studies (NDSs) have been proposed to overcome the limitations in the first two streams. The primary purpose of an NDS is to collect driver-based data such that the relationship between driver behavioral factors and crash risk can be better understood (Dingus et al. 2006; Olson et al. 2009; Eenink et al. 2014; Victor et al. 2015). Furthermore, such studies allow for the collection of vehicle-based kinematic safety critical events (SCEs), which include hard brakes, activation of forward collision system, and/or activation of rolling stability systems. Numerous studies have shown positive associations between SCEs and crashes (e.g., Pande et al. 2017; Gitelman et al. 2018; Cai et al. 2021). As a result, SCEs can be used as proxies for crash risk since they are more frequent to rare crash-events, and hence, can facilitate modeling by allowing for larger statistical power. While the existing NDSs have provided excellent insights to inform driver-behavior regulations (Victor et al. 2015), their impact on transforming trucking operations is limited because most NDSs were conducted on non-commercial drivers and in a few testing areas (Cai et al. 2021) and because the analyses have typically focused on the association between behavioral factors and crash risk. Existing NDS-based models have not examined the likelihood of a crash or SCE for a given time window by using both behavioral factors and driving conditions (e.g., speed limits, types of roads, weather conditions, etc.) The implications of these limitations are two-fold. First, it is unclear whether the findings can translate to commercial drivers who are heavily regulated and have superior driving training. Second, from a trucking operator's perspective, it is unclear how to capitalize on insights from NDSs to assign schedules or routes that can minimize crash risk.

The overarching goal of this study is to develop predictive models that can be used to quantify the impact of trip (e.g., day of the week and time), driver (e.g., age, cumulative driving time, gender and history of SCE), traffic and weather conditions on the likelihood of observing at least one SCE over a 30-minute driving window. We use SCEs as crash surrogates based on the findings of Cai et al. (2021) that showed that headway, hard breaks, activation of the rolling stability, and forward collision mitigation systems are positively associated with increased crash risk based on data collected from 30,000+ trucks and 2 billion miles of driving. To address our overarching goal, we capitalize on SCE data routinely collected by a large U.S. trucking company to examine the following research questions:

(A) Can SCEs be predicted using easy-to-access driving-related data sources? Here, we examine whether date-time, driver, traffic, weather, and past SCE driver records can inform the prediction of SCE occurrence over a 30-minute window.

(B) If the SCEs can be predicted, what is the relative importance of the various predictors? Here, we attempt to understand which factors can distinguish trips for which at least one SCE has been recorded versus trips where no SCEs were reported. This question is pertinent since existing studies have not considered the association between SCE occurrence and driving conditions for window sizes greater than a few seconds.

(C) Can the prediction models be generalized to new drivers and future time periods? This question tackles the inherent assumptions behind how machine learning models are trained. Specifically, if were only interested in question (A), then a random partitioning of the available data would be sufficient. However, we posit that practitioners would also be interested in understanding if the developed models can be used to predict SCE occurrence for drivers whose data were not used in model training, e.g., new drivers, and for future time periods without daily/weekly retraining of the model.

The remainder of the paper is organized as follows. We describe the data collection protocol in Section 4.2. We present the methodological framework used to address our overarching goal and its associated three research questions in Section 4.3. In Section 4.4, the prediction and variable importance results are presented. This is followed by a discussion and the analysis of results and its potential contributions in both literature and the trucking community in Section 4.5. To encourage the adoption of our work by industrial practitioners and facilitate future research, we provide the link for the repository containing our code and analysis in the *Supplementary Materials* Section.

4.2 Data description

This study capitalizes on data routinely collected by a large U.S.-based trucking company, providing transportation, delivery and logistics services to customers in North America. Unlike most NDSs, the vehicle-based data were collected by a trucking company, and not by the research team. The data were originally collected for regulatory compliance, routine performance monitoring, and driver assistance purposes. The company's trucking data were anonymized prior to its access by the research team. Furthermore, the study protocol and data usage for research purposes were approved by the Institutional Review Board of Saint Louis University.

To address our three research questions, we utilized data collected by the company from April 1, 2015 to March 31, 2016 (and first reported on by Cai et al. (2021)). The company data set is comprised of: (a) *drivers' characteristics*, which provides an anonymized driver's ID, age, gender, business unit, etc.; (b) *trajectory data*, which were collected by on-board GPS trackers, providing the vehicle's location and speed; and (c) *SCEs*, which captured kinematic-based unsafe driving events from the Bendix® Wingman® Advanced™ monitoring systems equipped on each truck. We supplemented the trucking company's datasets with weather information, which captured the precipitation probability, precipitation intensity, wind speed, and visibility conditions for each GPS ping of the truck (i.e., using both the location and time-stamp to depict the associated weather conditions). Moreover, we utilized the truck's mean speed and standard deviation as proxies for traffic conditions during each trip. The rationale was to account for the importance of traffic conditions on crash risk (see the review of Theofilatos and Yannis 2014, for a detailed introduction), while overcoming the lack of a national and freely available dataset capturing traffic experienced by our drivers.

4.2.1 Drivers' characteristics

For this study, we randomly sampled 500 out of the 15,707 regional truck drivers in the dataset. Regional drivers are typically on duty for five or more days, return home on a weekly basis, and provide delivery in a geographic area that may include nearby states. Note that: (a) regional drivers were selected for our analysis since we surmised that their experience in driving in varying traffic conditions and road types can make our developed prediction models more comprehensive and generalizable; and (b) we limited our analysis to 500 drivers to make the computations more tractable and to facilitate the collection of the associated weather data (which was limited in speed and volume due to querying restrictions).

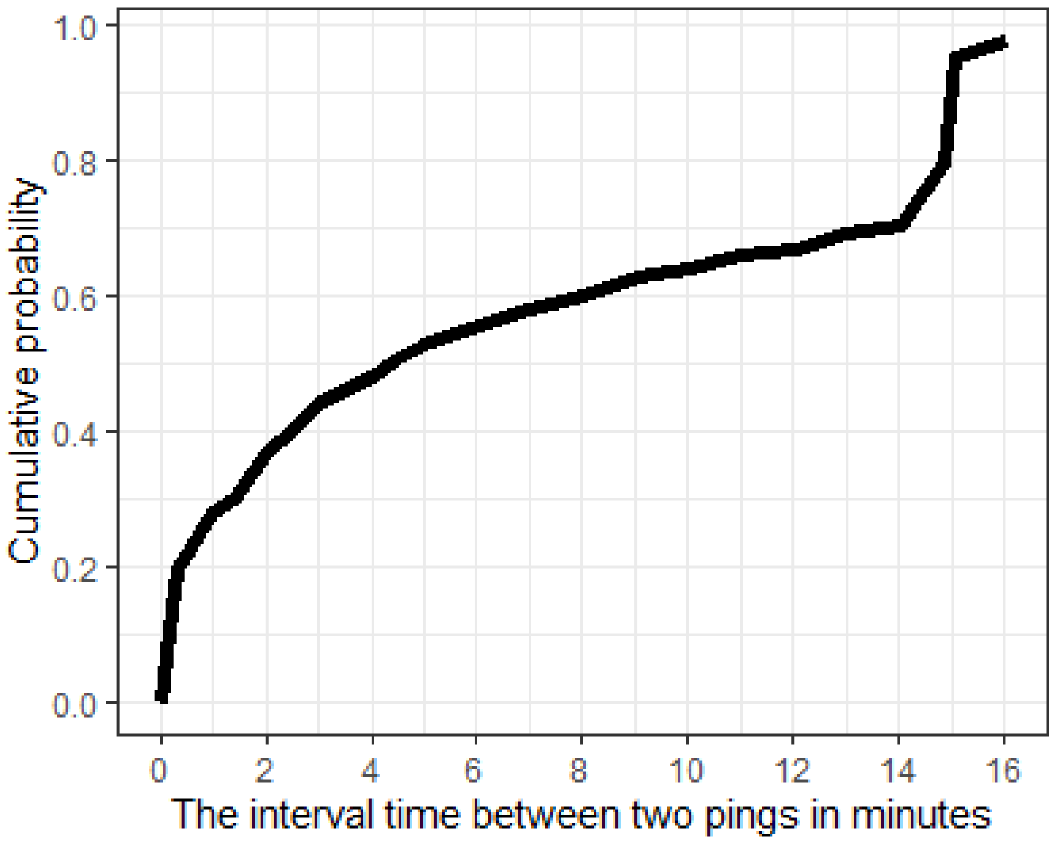
Over the course of the data collection period, the selected drivers drove from 1 to $2,100$ hours, covering 20 to $98,000$ miles. Approximately, 85% of the selected drivers drove at least 250 hours for a total of 10,000+ miles. We have eliminated four drivers due to an unrealistically high rate of reported SCEs relative to their miles driven (likely due to recording errors). Hereafter, all the reported data will correspond to only those generated by the 496 drivers, whose characteristics are summarized in Table 4.1.

Table 4.1: A summary of driver characteristics including their average age ± SD, number of drivers per gender, and business unit (with their % in parentheses).

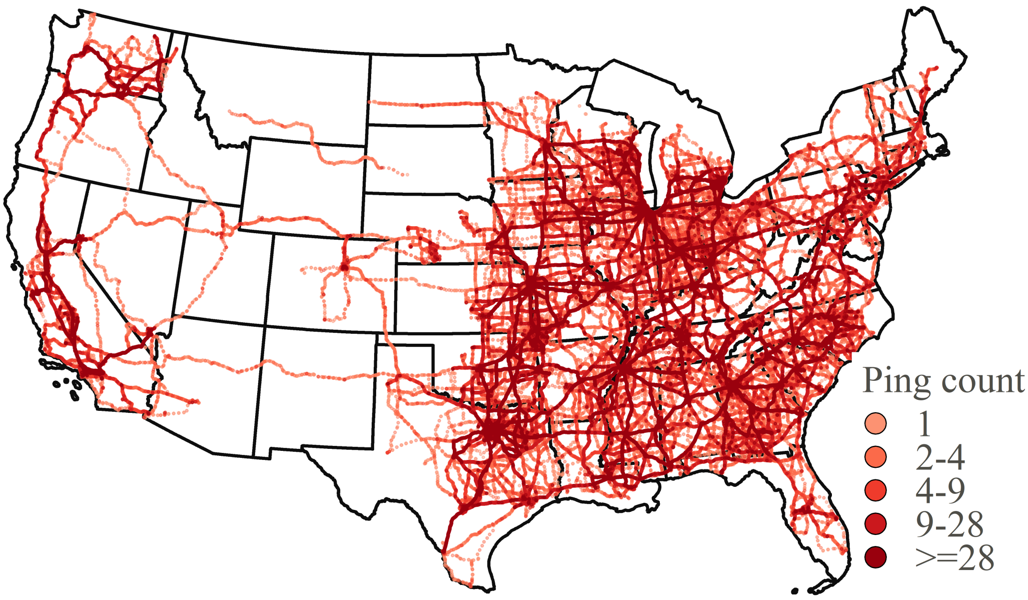
Variable	Statistic
Age:	
Range	21 to 76 years
Mean age ± SD	48.0 ± 11.8
Gender: (%)	
Male	456 (91.9%)
Female	36 (7.3%)
Unknown	4 (0.8%)
Business unit: (%)	
Intermodal	496 (100%)

4.2.2 Trajectory data

For each truck, intermittently collected real-time driving *pings* were obtained. Hereafter we use the term *pings* to these data records, each including: (a) date and time, (b) GPS coordinates (latitudes and longitudes with five decimal precision), (c) GPS quality, (d) speed, and (e) the driver's unique identification code (which we can use to link the driver to the characteristics presented in Section 4.2.1). In Figure 4.1(a), we show the cumulative density function for the interval time between two pings, which varied between a couple of seconds to approximately 16 minutes. The median interval time was 4.41 minutes, and more than 95% of the pings had interval times < 15 minutes. We show the associated geographic distribution of active (i.e., truck speed > 0) pings for the 496 sampled drivers in Figure 4.1(b). Note that a darker color indicates a higher number of pings. Observe that despite the sampling of drivers, the reported pings span most of continental U.S., generally following the U.S.'s population density and the distribution of major highways. Hence, we posit that our sample is representative of truck driving throughout the continental U.S.



(a) Cumulative density function of time between two pings.



(b) The geographic location of active pings in our dataset.

Figure 4.1: An overview of the pings data generated by our 496 drivers from April 1, 2015 to March 31, 2016.

4.2.3 Safety critical events

An SCE was recorded whenever a pre-specified kinematic threshold was exceeded while driving. The recording of an SCE event was performed using the Bendix® Wingman® Advanced™ monitoring system, which was equipped on each truck, and recorded the exact date-time, latitude and longitude, driver, and the type of SCE. Four different types of SCEs were recorded (Cai et al. 2021):

- Headway, which denoted an instance when an unsafe gap of ≤ 2.8 seconds between leading and trailing vehicles was maintained for ≥ 118 seconds (Grove et al. 2015);
- Hard brakes, which captured instances when the truck decelerated at a rate of ≥ 9.5 miles per hour per second.
- Activation of the rolling stability system, which assisted drivers by applying brake pressure (and potentially applying trailer pressure) to align the vehicle when the Bendix® monitoring system's thresholds were approached (Bendix® 2007); and
- Activation of the forward collision mitigation system, which intervened to avoid or mitigate forward collisions by the truck (Grove et al. 2015).

Over the course of the study period, the 496 drivers had a total of 9,032 SCEs, which were divided into: (a) 3,944 headway events, (b) 3,588 hard brakes, (c) 869 initiations of the forward-collision mitigation system, and (d) 631 initiations of the rolling stability system.

Note that (Cai et al. 2021), who used the same dataset, has previously established that these SCEs are significantly positively correlated with actual crashes as well as injuries. Consequently, this observation justifies the use of these specific SCEs as surrogates for traffic incidents.

4.2.4 Weather data

To capture the weather conditions encountered by each driver we collected data on: (a) precipitation probability, (b) precipitation intensity, (c) visibility, and (d) wind speed. These variables were selected since they have have been shown to be significant in several crash prediction models (see e.g., Theofilatos and Yannis 2014; Lin et al. 2015; Wang et al. 2015; Cai et al.

2021). We queried the historical values for the four weather variables from the *DarkSky API* (Dark Sky Company 2020) by inputting the timestamp, latitude, and longitude of each active ping.

4.3 Methods

To address our overarching goal and three research questions, we propose a 6-step machine learning framework. In the first step, the ping data are prepared/transformed by filtering noise, determining *stay point* locations (such that origin and destinations of each trip can be estimated), and segmenting each trip into ≤ 30-minute intervals. In the second step, two features are engineered from the SCE dataset: (a) a binary outcome variable is computed for each trip based on the SCE dataset, where $y = 1$ if at least one SCE is observed, and it is set to zero otherwise; and (b) an independent/predictor variable capturing the number of observed SCEs for each driver-trip over seven previous days (SCELag7). We have considered other definitions (3-, 5-, 10- and 14-day alternatives) but have observed that the model prediction accuracy generally increases when longer periods are considered, but this effect is not substantial. For a practical application, the selection of SCELag depends on the availability of the data. With the computation of these two variables, our four disparate data sources (ping, drivers, SCE, and weather) are combined into a tabular format. In the third step, the tabular data are divided into three sets of training and testing samples, corresponding to the random, driver-based, and time-based sampling strategies discussed in Section 4.1. In the fourth step, each of the three training sets are divided into 5-folds to facilitate the assessment of the fit of the trained models, and sub-sampling strategies are used to handle the data imbalance problem resulting from the frequency mismatch between the 0s (majority class) and 1s (minority class since the occurrence of an SCE is relatively rare) when training the models. In the fifth step, nine machine learning algorithms are trained, and they are evaluated/interpreted in the sixth step of our framework. We depict our six-step machine learning framework in Figure 4.2. The details for each step are presented in the subsections below.

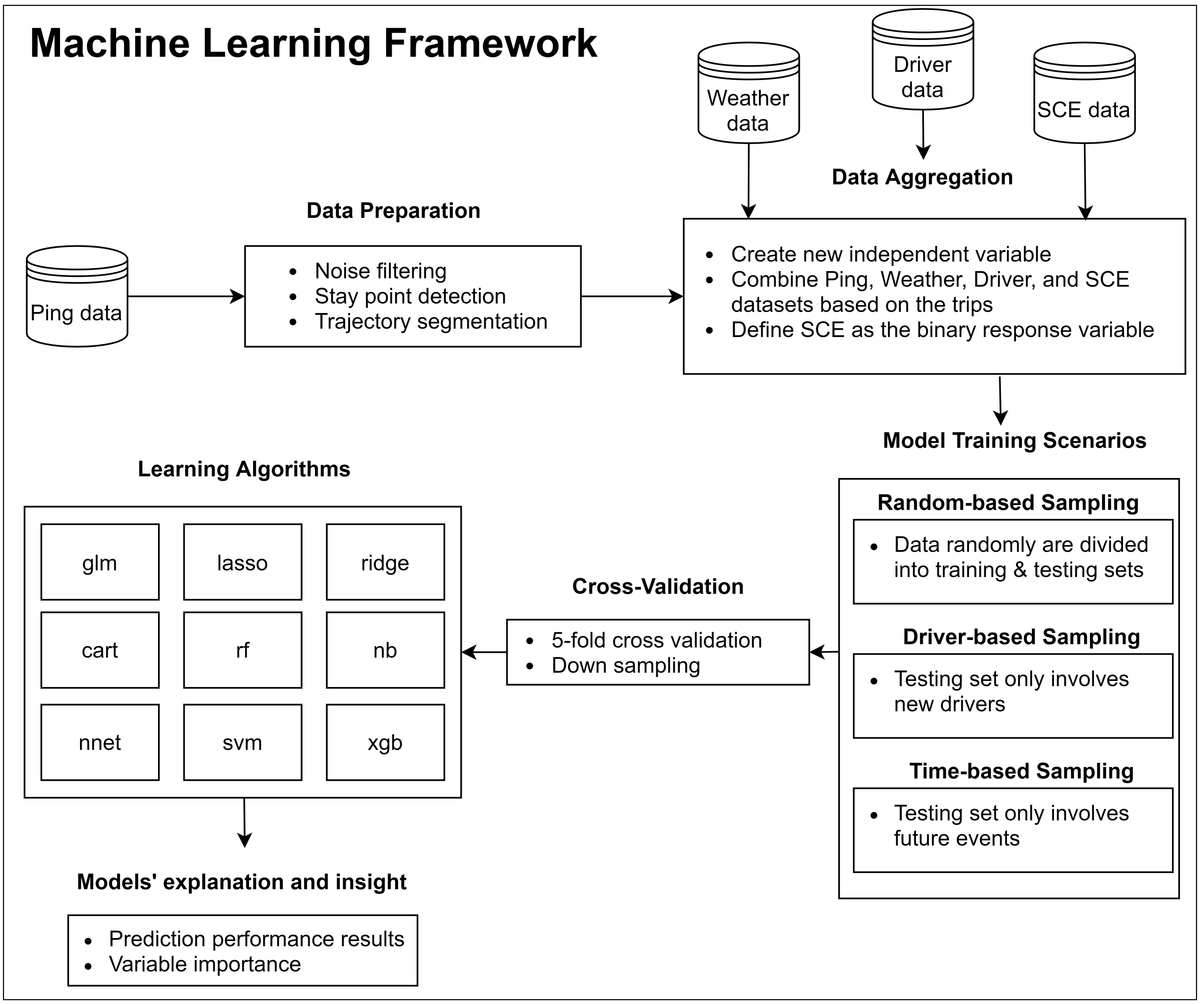


Figure 4.2: An overview of the proposed 6-step machine learning framework.

4.3.1 Data preparation

Noise filtering The first component in data preparation is to remove noisy/low-quality pings. Based on the GPS quality indicator, $\sim 98\%$ of the pings were of "good quality." Furthermore, we removed pings where the speed was over 80 miles per hour or the returned dates were outside of our analysis window (April 1, 2015 to March 31, 2016).

Stay point detection Our ping data were not limited to moving pings, but also included pings where the trucks were stationary. From an SCE occurrence perspective moving pings are more valuable than stationary pings since SCEs are triggered based on kinematic thresholds, i.e., no SCEs can be observed when the truck is stationary (or near stationary when parking, loading, and/or unloading). Based on the GPS trajectories, stay point detection algorithms are used to detect stop episodes from GPS trajectories with gaps or infrequent ping data (e.g., see Hwang et al. 2017; Wu et al. 2018; Li et al. 2008). In this study, we adopted the algorithm of Yuan et al. (2012), which uses both a time-based and a distance-based threshold to detect movements. In our analysis, we set the time-based threshold to 30 minutes and the distance-based threshold to 0.5 miles, i.e., if a truck spends greater than 30 minutes within a distance of 0.5 miles, these pings will be considered within a stay point.

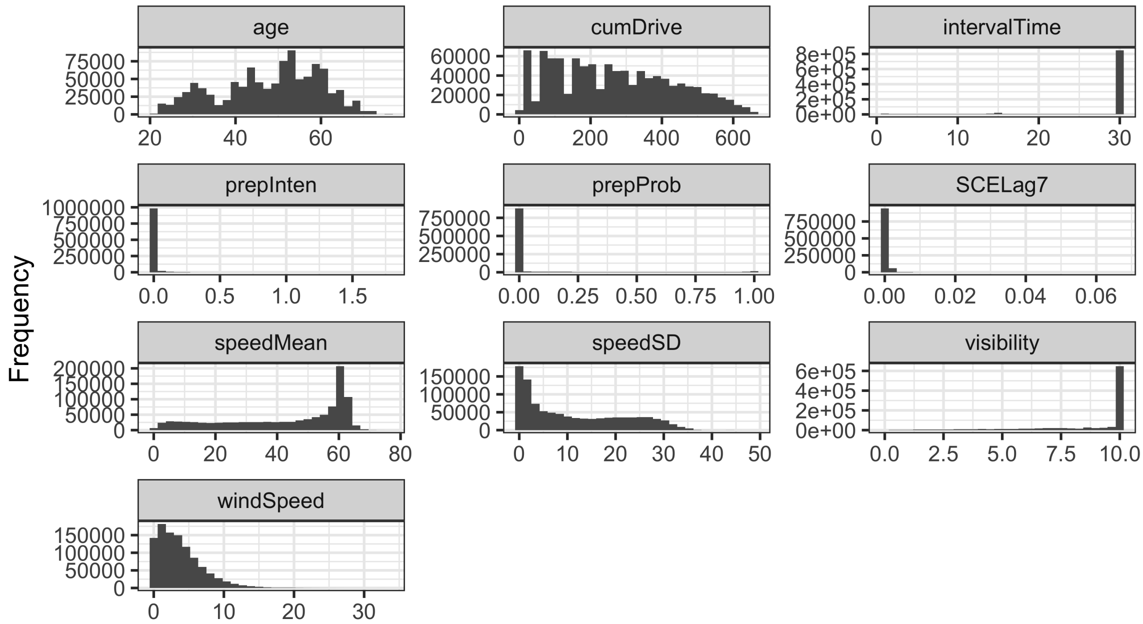
Trajectory segmentation After applying the stay point detecting algorithm, we grouped the actively moving pings into trips. A trip is defined to include the sequence of moving pings between two stay points. However, the lengths of these trips varied between a few minutes to a maximum of eight hours. This heterogeneity undermines both the high resolution of our data and its statistical validity. Specifically, a trip of eight hours can include hundreds of pings, and aggregating all these pings into a single observation (since we are using the trip as our unit of analysis) will result in a significant loss of information. Furthermore, the developed machine learning models will not be actionable since, for longer trips, practitioners are likely more interested in detecting the pattern of SCEs within a trip, i.e., whether the events occur in early or later stages of a trip, or there is no pattern in the distribution of events within a trip. On the other hand, if a trip becomes too short the prediction model would not be practical since there is no enough time to perform preventive actions (e.g, rest breaks). Recent studies in the crash prediction field (e.g, You et al. 2017; Sun and Sun 2015; Sun et al. 2014; Hassan and Abdel-Aty 2013) have used time intervals between 5 to 30 minutes before the crash to develop their crash prediction models. In this study, we cut trips into 30-minute intervals. For example, a 67-minute trip would be divided into three intervals (two 30-minute intervals and one 7-minute interval).

4.3.2 Data aggregation

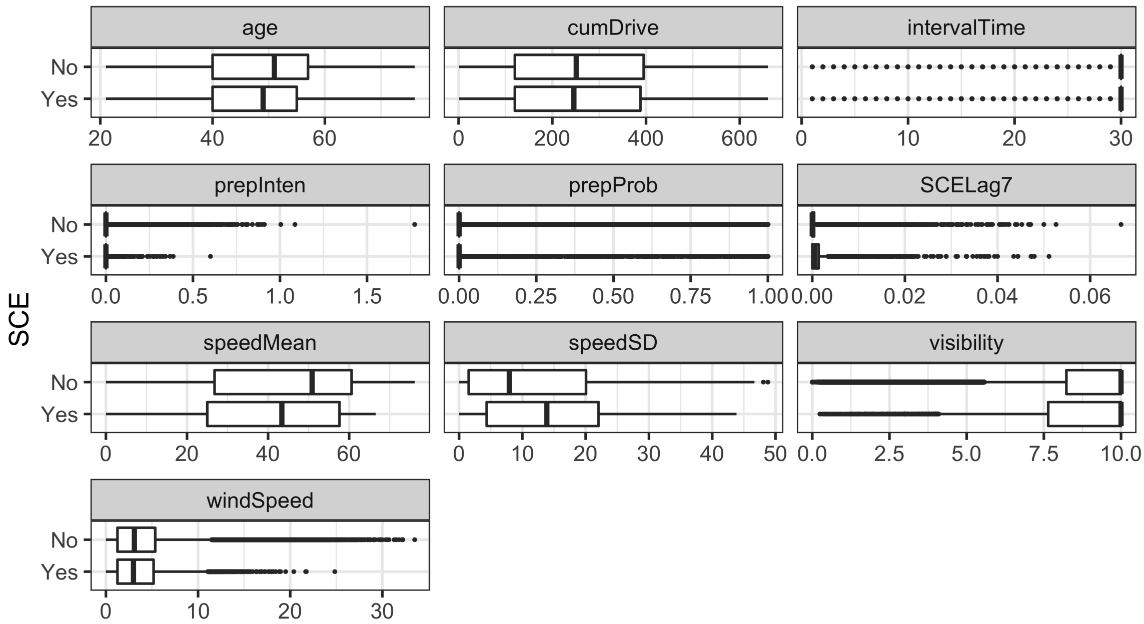
The process of aggregating ping data into trips was described in Section 4.3.1. In this step, the true value for the binary outcome/response variable to be used in the machine learning models is computed. We have set $y = 1$ if at least one SCE of any type is reported in a given trip, and $y = 0$ otherwise. For each trip, we recorded the values for 15 predictor variables, which are categorized and defined in Table 4.2. Furthermore, we summarize their obtained values and their distribution by the value of the response variable in Figure 4.3. Note that box plots by SCE overlap for a given continuous predictor, which provides evidence that the prediction problem is non-trivial. Furthermore, several of the predictors are highly skewed. While some machine learning models can handle such skewed data, we centered and scaled (i.e., $z-$transformed) each predictor so that the inputs across our different machine learning models are consistent. This process of centering and scaling the predictors is commonly used in machine learning research and practice (see James et al. (2013) for a detailed introduction).

Table 4.2: Definition of the predictors

Predictor	Definition	Type
	Block 1: trip related	
dayOfWeek	day of the week	categorical
holiday	whether or not the trip is within a holiday	binary
hourDayCat	time of the day's category (rush1, mid day, rush2, night)	categorical
intervalTime	trip duration	continuous
weekend	whether or not the trip is on weekend	binary
	Block 2: driver related	
age	driver's age	continuous
cumdrive	cumulative driving time within each shift	continuous
gender	driver's gender	binary
	Block 3: traffic related	
speedMean	average speed during the trip	continuous
speedSD	average standard deviation of speed during the trip	continuous
	Block 4: weather related	
prepInten	average precipitation intensity during each trip	continuous
prepProb	average probability of precipitation during each trip	continuous
visibility	average visibility during each trip	continuous
windSpeed	average wind speed during each trip	continuous
	Block 5: sensor related	
SCELag7	number of SCEs recorded for a given driver in the past 7 days divided by their total hours driven during that period	continuous



(a) Histograms of continuous predictors



(b) Box plots of continuous predictors by SCE occurrence

Figure 4.3: An exploratory analysis of the trip-based data.

As a final step prior to training the machine learning models, we investigated the correlation between the different variables. The goal here is to examine whether the dimension of the aggregated data can be reduced, which would increase the efficiency of training different machine learning models. Based on Table 4.2, our variables have three different types: continuous, binary/dichotomous, and polytomous/categorical. Hence, the standard Pearson correlation approach cannot be utilized among non-continuous predictors. To account for the mixed data types, we have utilized the approach of Revelle et al. (2010) who recommended the use of: (a) *Pearson correlation coefficient* for continuous predictors, (b) *tetrachorics* for dichotomous variables (c) *polychorics* for the polytomous variables, (d) and the polyserial/biserial correlations for the mixed variables (e.g., correlation between SCELag7, continuous, and weekend, binary). The results from the correlation analysis are presented in Figure 4.4.

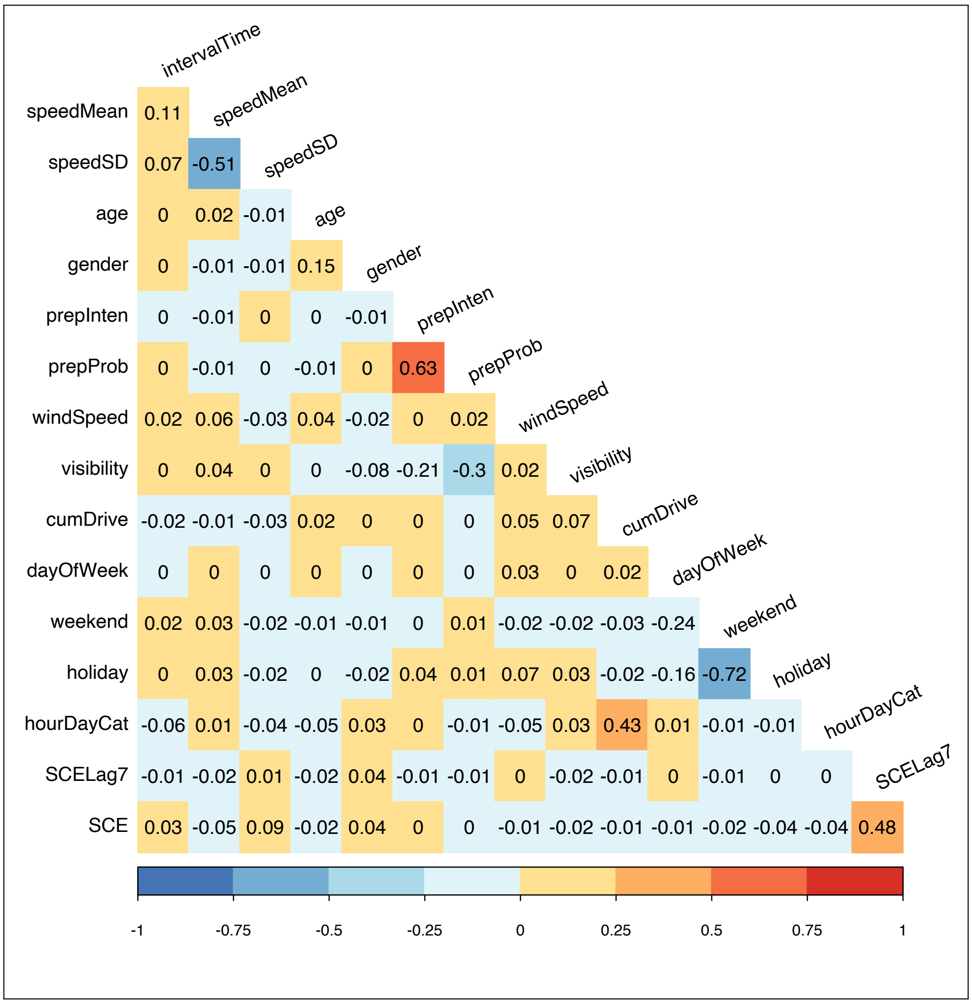


Figure 4.4: A correlation plot of the predictors.

Based on Figure 4, there are two important observations to be made. First, the correlation between each of the predictors and our response variable (SCE) is low, with an absolute value $|\leq 0.1|$, with the exception of SCELag7 which had a correlation coefficient of 0.48. One implication from such observation is the potential utility of non-linear machine learning models given the small correlation values. Second, all the correlations $< |0.90|$. The following three pairwise correlations were in the $0.50 < |\rho| < 0.75$: speed mean and speed standard deviation, precipitation intensity and probability, and holiday and weekend. We chose to retain all three pairs since: (a) their absolute values were less than the default cutoff in the `R caret` package of 0.90 (Kuhn 2020); (b) Tsai et al. (2015) showed differences in traffic flow patterns between weekends and holidays; and (c) Cai et al. (2021) showed that both precipitation intensity and probability are important in modeling the association between crashes and safety critical events for truck drivers.

4.3.3 Model training scenarios

As with any predictive analytics application, the statistical and machine learning models are trained using one dataset and assessed using a separate holdout dataset that is unseen during model training (Shmueli 2010). While it is customary to randomly split the available data, other non-random ways may be more appropriate in some cases. In this paper, we have considered three sampling scenarios. In the first scenario (*random-based sampling*), we randomly assign trips into the training and testing sets such that the proportion of events (i.e. $y = 1$) in both datasets is the same. This sampling strategy allows us to address research questions 1 and 2. On the other hand, the purpose of the other two sampling strategies is to mimic how the developed models can be used in practice. Hence, in the second *driver-based* scenario, we separate the training and testing data based on the drivers, i.e., each driver's trips are assigned to either the training or testing set. This way, we can evaluate the generalizability of the prediction, i.e., whether the models are significantly dependent on the subset of drivers used for training, or whether the prediction accuracy is transferable to other drivers, not seen by the model during training. This experiment is referred to as *driver-based sampling*. Finally, we sample training and testing sets chronologically, i.e., all testing trips occur after all training trips. This way we evaluate the potential of the models to be used in predicting future events. This experiment is referred to as *time-based sampling*. For all the sampling scenarios, we divide the dataset into training and testing/holdout sets with 80/20 ratio.

4.3.4 Cross validation and subsampling

Cross validation In this paper, we have employed a five-fold cross validation to tune hyper parameters and select the best model from the training set for each learning algorithm. Both five- and ten-fold versions have been widely used in the literature (e.g., see Fernández-Delgado et al. 2014; Wainer 2016). As stated in James et al. (2013)[p. 184], "there is a bias-variance trade-off associated with the choice of k in k-fold cross-validation. Typically, given these considerations, one performs k-fold cross-validation using $k = 5$ or $k = 10$, as these values have been shown empirically to yield test error rate estimates that suffer neither from excessively high bias nor from very high variance." We have chosen $k = 5$ in our analysis since it is more computationally efficient than the 10-fold cross validation approach. Tuning and optimizing hyper parameters in machine learning algorithms is an important step especially given unbalanced data. Two popular methods can be identified: (a) random search where the values for the hyper parameters are randomly sampled from the hyper parameters domain; and (b) grid search where the values for the each combination of the hyper parameters are selected from a specified grid. We have used random search method in this study since, at least according to some reports, it has demonstrated better performance and/or computationally efficiency (Bergstra and Bengio 2012; Khayyer et al. 2021).

Subsampling While we have used SCEs in place of actual crashes due to their higher occurrence frequency, they are still relatively rare. In our dataset, only 0.75% of trips contained at least one SCE. Hence, we have examined several standard subsampling methods to overcome the *data imbalance problem* (Batista et al. 2004; Wang et al. 2019b), which can lead to the development of naive models where the model (almost) always predicts the majority class. Based on our preliminary analysis, we have selected *down sampling* strategy, where the majority class is down sampled to equal to the size of the minority class since: (a) we have observed no significant differences were observed between *down sampling* and more advanced sub-sampling approaches methods, and (b) it is the most efficient, from a computational perspective, approach to handle the data imbalance problem.

4.3.5 Machine learning algorithms

It should be noted that while each learning method has its own advantages and disadvantages, it is generally impossible to predict *a priori* which will perform best for a particular application domain, let alone a particular dataset. Consequently, it is often advisable to try multiple methods and observe the prediction performance. Machine learning models can be grouped into the following three categories (Dolatsara et al. 2020):

(A) *Statistical models*, considered as parametric models. In these models there is an assumption about the form of prediction model. Logistic and other generalized linear models (glm) are often used. Also, ridge, and lasso regression, and naive Bayes (nb) are well-known statistical model which were widely used in the literature (e.g., Theofilatos 2017; Wu et al. 2018; Wang et al. 2019a).

(B) *Single (data-driven) classifiers*, which unlike the statistical models are considered as non-parametric models and have no assumptions about the form of the prediction models. There are a wide variety of these algorithm that have been used in the literature such as the classification and regression tree (cart), support vector machine (svm), and neural network (nnet) (e.g., Pande and Abdel-Aty 2006b; Hossain and Muromachi 2011; You et al. 2017; Parsa et al. 2020).

(C) *Ensemble approaches*, where the idea is to aggregate singular models together with the goal of reducing the overall bias or variance. Random Forest (rf) and XGBoost (xgb) are well-known ensemble models that were used in different applications (e.g., Theofilatos 2017; Zhang and Zhan 2017; Parsa et al. 2020).

From an interpretation perspective, statistical models are more explainable. However, the performance of machine learning models can be superior in some applications, especially when the two classes are not linearly separable. Therefore, we examine nine popular algorithms (glm, lasso, ridge, cart, rf, nb, nnet, svm, xgb) capturing statistical methods, single classifiers, and ensemble methods. These models have been shown to produce excellent predictive performance in a number of applications, e.g., traffic safety (Silva et al. 2020), transplantation (Dolatsara et al. 2020), physical fatigue prediction (Maman et al. 2020), and stock market movements

(Weng et al. 2018). Hence, in our estimation, they are representative of models that would have both good predictive performance and are relatively computationally efficient (i.e., do not need a GPU to run effectively).

In our implementation, we have used the `caret` package (Kuhn et al. 2008) in the **R** programming language, and utilized a random search of 20 combinations of the tunable parameters for each tunable model (i.e., all the models with the exception of the logistic regression model which has no tunable parameters). For more details, the reader is referred to our code in the supplementary materials.

4.3.6 Models' evaluation and insight

Performance metrics In this paper, we report five performance measures for each machine learning model. Note that, for a binary response variable, there are 2×2 possible outcomes from the application of a predictive model. These outcomes are typically summarized using a *confusion matrix*, which we schematically present in Table 4.3.

Table 4.3: Confusion Matrix for a classification problem with a binary response variable

		Predicted outcomes	
		0	1
Actual class	0	true negative (TN)	false positive (FP)
	1	false negative (FN)	true positive (TP)

Based on this confusion matrix, we can compute four metrics: (a) *accuracy*, which measures the number of observations (both positive and negative) that were correctly classified; (b) *sensitivity*, proportion of the true positive observations among the ones predicted to be positive; (c) *specificity*, which measures the proportion of the true negative observations among the ones predicted to be negative; and (d) *Gmean*, which measures the geometric mean between sensitivity and specificity, and consequently gives a balanced measure of model performance on the majority and minority classes. Mathematically, these measures can be computed as follows:

$$\textit{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{4.1}$$

$$\textit{Sensitivity} = \frac{TP}{TP + FN} \tag{4.2}$$

$$\textit{Specificity} = \frac{TN}{TN + FP} \tag{4.3}$$

$$\textit{Gmean} = \sqrt{\textit{Sensitivity} \times \textit{Specificity}} \tag{4.4}$$

Our fifth metric can be obtained from the receiver operating characteristics (ROC) curve, where *sensitivity* is plotted on the y-axis versus $1-$ *specificity* on the x-axis. The area under the curve (AUC) captures how well the model predicts actual 0s as 0s and actual 1s as 1s. AUC is our fifth metric. All five metrics can vary between 0 and 1.

Variable importance and model interpretation The study of variable importance can help explain the outputs from machine learning models by quantifying the impact of each predictor variable on the model's predictive performance. That being said, variable importance can be computed based on either the training or test/holdout dataset (see Molnar 2020, Section 5.5.2 for a detailed introduction on each approach). The fundamental difference between both approaches stems from the definition/purpose of computing feature importance (Molnar 2020). To allow for different interpretations and to evaluate the consistency of our predictors (an indirect measure for assessing over-fitting), we will consider both approaches. For the training set analysis, we have utilized the standard variable importance methodology from the **R** `caret` package (Kuhn et al. 2008), which attempts to measure the contribution of each predictor on each class of the response variable through ROC curve analysis, and then ranks the importance of each predictor by assigning them a relative score between 0 and 100. As an alternative method, we have used a sequential approach to perform variable importance on the testing set, where we have divided the predictors into five blocks (defined in Table 4.2) and examined the change in prediction accuracy as variables from a given block are made available to the models. We started with block 1 as our baseline since this trip-based information would be readily available for practitioners. We then sequentially added blocks 2 through 5. The rationale for this approach was based on: (a) our hypothesis that if researchers/practitioners have access to one predictor in a given block, they would likely be able to compute and/or query the remaining predictors within the block, and (b) examining all different permutations would be computationally expensive (given that we have 15 variables and 9 machine learning model).

4.3.7 Statistical software used and computational infrastructure

All data cleaning, statistical modeling, and visualization were performed in the statistical computing environment **R** 3.6.0 (R Core Team 2019) using the Ohio Supercomputer Center, with a virtual machine of 28 CPU cores, and 112 GB of RAM. We used the `caret` (Kuhn et al. 2008) and `caretEnsemble` packages (Deane-Mayer and Knowles 2019) and their dependencies to train the models and perform variable importance analysis. Note that we estimated the driving distances from the trajectory data using the haversine method which was computed by `distHaversine()` function in `geosphere` package (Hijmans 2019). To facilitate the adoption of our methodology, we provide a link of a compiled **R** Markdown file, containing the code for data aggregation and model building, and their associated results, in the supplementary materials section.

4.4 Results

4.4.1 The baseline/random-sampling scenario

In Table 4.4, we present the test/holdout results obtained from the application of the nine machine learning models to our random-sampling scenario, which is designed to address our first research question: whether SCEs can be predicted using the kind of data available. First, observe that all models make informative prediction, since AUC values range from 0.693 (cart) to 0.765 (xgb). Note also that the models exhibit larger variation if we focus on the specificity (0.663 to 0.805) and sensitivity (0.534 to 0.684). Interestingly, all models achieve higher specificity relative to their sensitivity results, which means that all models have higher errors when predicting trips with SCEs compared to those without. It can be explained by considering that 99.24% of our observations are without SCEs. Overall, it can argued that xgb model can be considered to have the best performance, since it achieves the highest AUC, sensitivity and Gmean, and has relatively low difference between sensitivity and specificity. Compare it to the nb model, which produced the best accuracy and specificity, but very low sensitivity. The confusion matrix for the xgb model is presented in Table 4.5.

Table 4.4: The predictive performance of the machine learning models on the holdout dataset for the random sampling scenario. We use **bold** text to highlight the best reported value for a given metric.

Metric/Model	cart	glm	lasso	nb	nnet	rf	ridge	svm	xgb
AUC	0.693	0.723	0.722	0.740	0.752	0.752	0.710	0.745	**0.765**
Accuracy	0.662	0.745	0.736	**0.803**	0.711	0.688	0.711	0.696	0.700
Sensitivity	0.641	0.548	0.554	0.534	0.652	0.679	0.579	0.647	**0.684**
Specificity	0.663	0.747	0.738	**0.805**	0.711	0.688	0.712	0.696	0.700
Gmean	0.651	0.640	0.639	0.656	0.681	0.683	0.634	0.671	**0.692**

Table 4.5: Confusion Matrix for the best classifier (xgb)

		Predicted outcomes	
		No SCE	SCE
Actual class	No SCE	141,589	60,562
	SCE	487	1,056

To further confirm the validity of our results and conclusions, following the machine learning literature we also performed additional verification steps. We first compared the results from the cross validation stage with those reported for the holdout dataset. From this check, we observed that there are no significant differences in each model's performance, which indicates that there is no strong evidence of over-fitting. Then, we trained and deployed three greedy ensemble algorithms to predict the occurrence of SCEs in a trip similar to the procedure detailed for the nine machine learning models in Section 4.3.5. The reader is referred to our R Markdown document in the supplementary materials for the exact implementation details. The AUC results for the three ensembles were 0.748, 0.765, and 0.765, i.e., they do not improve upon the xgb model's performance. Finally, as was noted earlier, during our preliminary experiments, we have also considered alternative definitions for the SCELag variable. Table 6 summarizes the results observed. We report AUC and Gmean only for the xgb model here for the sake of brevity. As discussed earlier, while a larger lag window marginally improves the model performance, this improvement is not substantial and in practice must be balanced with data availability and collection concerns.

Table 4.6: xgb performance based on different lag values in the definition of SCELag variable

Metric	SCELag3	SCELag7	SCELag10	SCELag14
Gmean	0.672	0.682	0.693	0.708
AUC	0.749	0.765	0.769	0.780

Based on these observations, we posit that using a larger grid for tuning the models and/or investigating the use of more advanced machine learning models would likely result in a similar prediction performance to our existing models. Further improvements are likely only possible if we were to explore additional independent variables for our analysis, which is outside the scope of this study.

4.4.2 Variable importance

Figure 4.5 shows the relative importance, scaled between 0 and 100, of each predictor on the training set for xgb. From the figure, we can observe that (a) the sensor and driver-based SCELag7 variable is the most important predictor, followed by the two traffic-related proxy variables; (b) precipitation intensity, precipitation probability, holiday indicator, and day of the week have limited importance (the latter two variables may be explained by the weekend indicator variable); and (c) at least one variable from the five predictor sets, defined in Table 4.2, had an importance score ≥ 9.7, which provides evidence to the multidimensional nature of SCE occurrence/prediction. Note that the results for the other models were generally similar. The interested reader is referred to the supplementary results to view these results (which are omitted here for the sake of conciseness).

In the second variable importance approach, we retrained the nine machine learning models using five strategies, where we incrementally increase the set of possible predictors available to each model using the blocks defined in Table 4.2. Once the models are trained using the incremental strategy, we have examined the predictive performance (measured using AUC) of

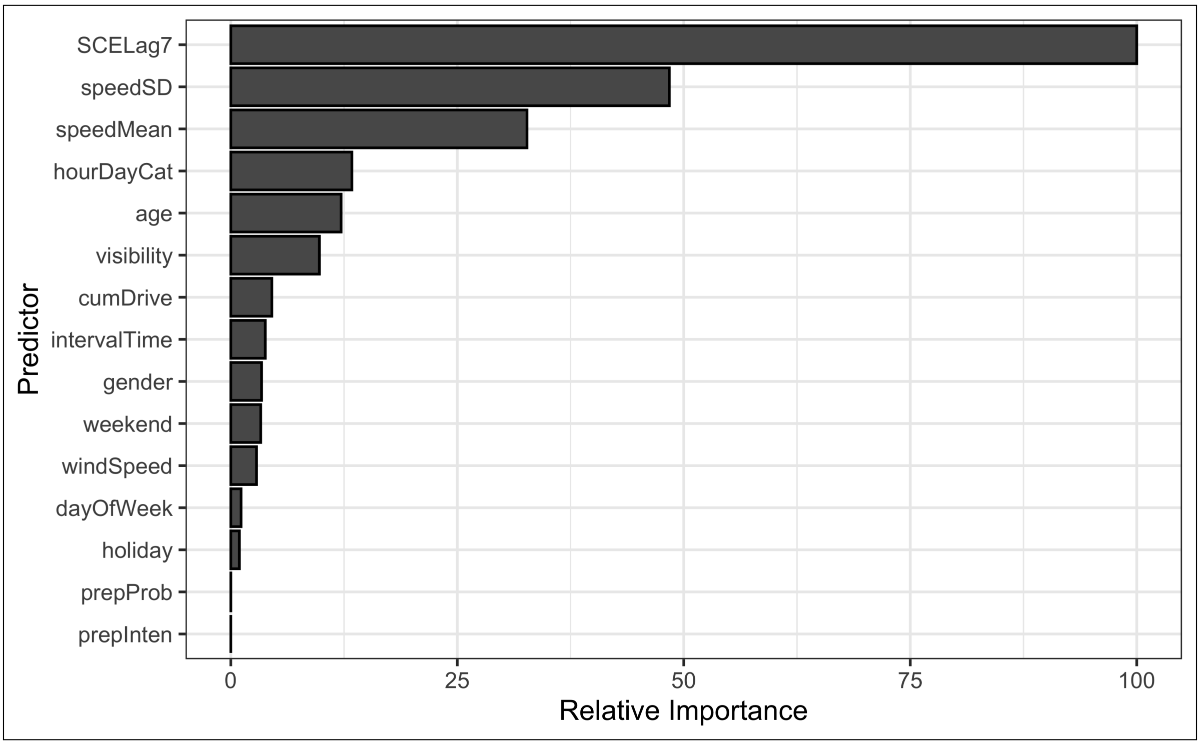


Figure 4.5: The scaled/relative importance of each predictor based on the trained xgb model.

these models on the holdout dataset. The obtained results are presented in Table 4.7. The addition of all blocks results in an increase in the AUC for all models, with the exception of block 4 which contained the weather-related predictors. This observation is consistent with Figure 4.5, which showed that the precipitation probability and intensity are the least important predictors. Hence, the inclusion of these variables in our models may result in more noise, which reduces the predictive performance of the models. Furthermore, the SCELag7 variable seems to be the most important since the inclusion of the fifth block improves the models' AUC by an average of 11.16%.

Table 4.7: The incremental change in AUC for each model based on limiting the set of possible predictors to those contained in the pre-specified set of blocks

	cart	glm	lasso	nb	nnet	rf	ridge	svm	xgb
Block 1	0.545	0.555	0.554	0.554	0.555	0.550	0.555	0.548	0.552
Blocks $1-2$	0.575	0.556	0.557	0.575	0.579	0.594	0.551	0.555	0.628
Blocks $1-3$	0.607	0.586	0.586	0.637	0.638	0.665	0.587	0.642	0.681
Blocks $1-4$	0.591	0.590	0.591	0.625	0.641	0.666	0.590	0.634	0.670
Blocks $1-5$	0.693	0.723	0.722	0.740	0.752	0.752	0.710	0.745	0.765

4.4.3 The driver- and time-based sampling scenarios

Table 4.8 presents the results for the driver-based and chronological (time-based) sampling scenarios. Compared to the base (random sampling) scenario, both the driver- and time-based sampling scenarios result in a slight decrease in prediction performance. This is not surprising, since both intuitively result in higher dissimilarity between training and testing sets, and hence higher perdition errors on the testing set. At the same time, the reduction in performance is relatively minor, especially for the best performing methods (xgb, svm, rf and nnet); for example, the xgb's AUC changes from 0.765 to 0.746 (driver) and 0.747 (chronological). From a trucking operator's perspective, this result means that a trained xgb model can be used for new drivers. Similarly, the model can be deployed and applied to future observations, i.e., it can be deployed without the need for daily/weekly retraining (since we held out more than 2 months of data in the chronological sampling scenario).

Table 4.8: The predictive performance of the machine learning models on the holdout dataset for the driver-based and chronological-based sampling scenarios. We use **bold** text to highlight the best reported value for a given scenario and metric.

	cart	glm	lasso	nb	nnet	rf	ridge	svm	xgb
Driver-based sampling scenario									
AUC	0.694	0.705	0.705	0.724	0.735	0.736	0.697	0.727	**0.746**
Accuracy	0.708	0.765	0.775	**0.802**	0.740	0.721	0.742	0.744	0.743
Sensitivity	0.597	0.495	0.495	0.517	0.582	**0.613**	0.506	0.570	0.606
Specificity	0.709	0.767	0.771	**0.804**	0.741	0.722	0.744	0.745	0.744
Gmean	0.651	0.616	0.620	0.645	0.657	0.665	0.613	0.652	**0.671**
Chronological sampling scenario									
AUC	0.679	0.707	0.710	0.719	0.742	0.736	0.701	0.722	**0.747**
Accuracy	0.653	0.731	0.744	**0.770**	0.700	0.684	0.713	0.699	0.702
Sensitivity	0.596	0.538	0.548	0.541	0.655	0.652	0.560	0.621	**0.658**
Specificity	0.654	0.733	0.746	**0.772**	0.701	0.685	0.714	0.700	0.703
Gmean	0.624	0.628	0.639	0.647	0.677	0.669	0.633	0.659	**0.680**

4.5 Discussion and conclusions

4.5.1 Summary of the main contributions

This study provides robust evidence that the occurrence of SCEs can be predicted over the span of 30-minute time windows based on merging routinely collected NDS and kinematic data from 20+ million miles driven by 496 commercial truck drivers with variables/proxies capturing trip, driver, weather, and traffic characteristics. Using a machine learning framework, we demonstrated the following:

(A) Statistical and machine learning models can predict the occurrence of SCEs in 30-minute trips/intervals, with the best model (xgb) achieving an AUC of 0.765 in the random sampling scenario. This result is excellent considering that the prediction problem is complex since (i) trips containing ≥ 1 SCE are relatively rare; (ii) the distribution of each continuous predictor seems to be similar, for trips where no and at least one SCE were observed, based on the box plots in Figure 4.3; and (iii) we did not directly observe behavioral factors (e.g., fatigue, sleepiness, and/or inattention) and traffic parameters (e.g., traffic flow, construction zones, road geometries, etc.) which are important per the retrospective/real-time crash risk modeling literature (Mehdizadeh et al. 2020), but are difficult to account for in an NDS study.

(B) The relative importance of the predictor variables was generally consistent based on two distinct variable importance techniques. Specifically, we showed that the kinematic (SCELag7) predictor was the most important, followed by the proxies for traffic-conditions (speedMean and speedSD), while the precipitation-related variables (prepInten and prepProb) were the least important. Hence, we also demonstrated that additional insights can be generated from traditionally black-box machine learning models through utilizing state-of-the-art methods in interpretable machine learning/explainable artificial intelligence (Molnar 2020).

(C) The models can be used for both new drivers and future time periods based on our driver-based, and time-based training/sampling scenarios. Specifically, the AUC for the best model was within 0.02 for our best model (xgb) when compared to the base (random) sampling scenario. From an implementation perspective, this difference is practically insignificant.

In addition to the reported results, we believe that the innovative use of non-random sampling, for assessing the generalizability of machine learning models, has not been considered prior in the transportation safety literature. In our estimation, such sampling strategies can be tailored to other transportation research to assist in evaluating predictive models' performance when deployed.

4.5.2 Contributions to the trucking industry

Our results can directly inform dispatching and routing applications and lead to the development of behavioral-based technological interventions to improve the safety of both commercial truckers and commuters. We have shown that the developed prediction models can be used to identify high-risk driving situations. Consequently, trucking operators and dispatchers can capitalize on our models to explain the risks to their drivers and/or inform their routing and rest-break scheduling policies (Hu et al. 2020; Mehdizadeh et al. 2020). Furthermore, with the question of *how would our models perform when deployed* in mind, we have shown that our models can be used for new drivers and can be updated using a batch process (i.e., every 1-2 months instead of being continuously retrained with new data). Hence, we estimate that operations research analysts at trucking companies can utilize our predictive modeling framework to inform their prescriptive dispatching and routing models by minimizing/mitigating high-risk driving situations.

4.5.3 Limitations and future research opportunities

This study has a number of limitations and data gaps that should be highlighted. First, the analyses presented throughout the study capitalize on observational and secondary data that were not originally collected for SCE predictions. The implications of this limitation are four-fold: (a) the research team had no input on driver and sensor system selection, (b) due to the data collection issues, the ping data seems sparse for some trips which cause the vehicle speed mean and standard deviation to not be an accurate estimation for traffic, (c) no causal relationships can be established between the SCEs and the predictors, and (d) potentially important predictors pertaining to traffic and weather conditions were not recorded (and had to be queried and/or estimated using external data sources). Second, our analysis herein was limited to regional truck drivers. It is unclear whether the insights obtained from this analysis can be translated to regular commuters and commercial drivers who have a significantly different driving experience (e.g., local truck and bus drivers). Third, we used only 30-minutes as the cut-off window for the trajectory segmentation. The impact of window-size has not been examined by the research team (for the purposes of the analyses performed in this paper) and hence, it is unclear whether the results would significantly change with other choices for window size.

Finally, it is worth noting that if deployed directly, the proposed models would need to rely on forecasted values for some of the predictors. The trip, driver and sensor-related predictors are fixed (i.e., have no uncertainty for a given 30-minute interval). Additionally, weather-related factors can be accurately obtained from weather forecasts since they are generally reliable for a 30-minute ahead time-window. The two remaining factors: speed mean and standard deviation, while they can be estimated based on trucking organization's routing and scheduling models, are less amenable for accurate forecasting. Note that we chose to use these as surrogates for traffic-related factors, since there are not any freely-available sources for accessing such data for 5+ million GPS pings (Mehdizadeh et al. 2020). The observed relative importance of speedMean and speedSD reflects how much traffic-related features are important. Consequently, in a practical setting, it may be preferable to instead use explicit forecasts of traffic factors (e.g., delay due to traffic, traffic volume, etc.), that are available through a number of services for commercial use.

In our estimation, there are two main streams of research that can capitalize on our framework and findings. First, our work can be extended to investigate the impacts of combining driver types (e.g., non commercial drivers), driver's attention, direct measures of traffic flow, and/or actual depictions of road surface conditions with the predictors investigated on the predictive performance of the trained machine learning models. We posit that the inclusion of additional predictors is likely to be more beneficial than examining a larger grid for tuning our machine learning models and/or examining more complex machine learning models (e.g., deep learning techniques). Our hypothesis is based on our observation that the ensemble models used to assess our models' performance did not improve upon our xgb model. Second, it is important to examine how prescriptive dispatching, routing and/or rest-break scheduling models can integrate the results from our models. Specifically, the probability of an SCE over the next 30-minutes based on predicted values for our independent variables can be easily computed from any of our models. With this in mind, optimization/prescriptive analytic models are needed for data-driven actions/decisions (e.g., rest-break, stop, decrease/increase speed) to minimize SCE/crash risk. Our predictive models developed in this paper can help address this critical gap (Hu et al. 2020; Mehdizadeh et al. 2020) by estimating the parameter values needed for the prescriptive models.

Online supplementary materials

As a supplement to this manuscript, we provide an **R** Markdown document that details the process by which we have obtained the machine learning results on our dataset. Specifically, the **R** Markdown document is comprised of five sections: (a) preliminaries, which provide the detailed information of the packages used for analyzing the data and their versions; (b) descriptive analysis of the dataset through statistical summaries and exploratory data analysis visualizations; (c) the setup and implementation procedure for the predictive models; (d) procedure used to compute variable importance; and (e) concluding remarks. Each section provides both the code used and its corresponding results. We host the HTML document generated by knitting the **R** Markdown at `https://mehdizadeamir.github.io//sce_predictions`. We placed our code/Markdown in the public domain through a CC0 - "No Rights Reserved" license to encourage both the research and practitioner communities to build upon our analysis.

Acknowledgments

This work was supported in part by the National Science Foundation (CMMI-1635927 and CMMI-1634992); the Ohio Supercomputer Center (PMIU 0138 and PMIU 0162); the University of Cincinnati Education and Research Center Pilot Research Project Training Program;

the Transportation Informatics Tier I University Transportation Center (TransInfo); a Google Cloud Platform research grant for data management; and a Dark Sky grant for extended API access of weather data.

Chapter 5

Investigating the robustness of the predication models

5.1 Introduction and motivation

In the previous chapter, we have shown that for the regional drivers, SCEs can be forecasted 30 minutes in advance, using machine learning techniques and dependent variables capturing the driver's characteristics, weather conditions, and day/time categories, where an area under the curve up to 76% can be achieved. Moreover, the predictive performance remains relatively stable when tested on new (i.e., not in the training set) drivers and a future two-month time period. However, we did not investigate the possibility of using the same prediction models that we trained for the regional drivers to predict unsafe driving situations for other type of drivers. Note that in our data, there are three types of truck drivers (Cai et al. 2020): (a) local drivers who provide service in at most 200-mile radius and should be returned home in the same day, (b) regional drivers who provide service in wider regions compare to the local drivers which might involve surrounding states, and (c) over-the-road drivers who provide service for long distances and are on the road for several days or weeks. We hypothesize that each type of driver has a distinct working environment that might affect the precursors that cause unsafe driving situations. Further, we think the different geographical areas might affect the likelihood of a safety critical event and need to be investigated due to their associated weather conditions, road types, and driving regulations. Furthermore, the transferability of the models was not tested in the previous chapter. Hence, we aim at investigating the temporal and spatial transferability of the models that we trained and developed for predicting SCEs. Our last goal for this chapter is to examine the possibility of developing crash prediction models to identify precursors for each SEC's type which as described in Section 4.2.3 are:

- Headway, which denoted an instance when an unsafe gap of $\leq$ 2.8 seconds between leading and trailing vehicles was maintained for $\geq$ 118 seconds (Grove et al. 2015);
- Hard brakes, which captured instances when the truck decelerated at a rate of $\geq$ 9.5 miles per hour per second.
- Activation of the rolling stability system, which assisted drivers by applying brake pressure (and potentially applying trailer pressure) to align the vehicle when the Bendix® monitoring system's thresholds were approached (Bendix® 2007); and
- Activation of the forward collision mitigation system, which intervened to avoid or mitigate forward collisions by the truck (Grove et al. 2015).

To sum up, the goals of this chapter are:

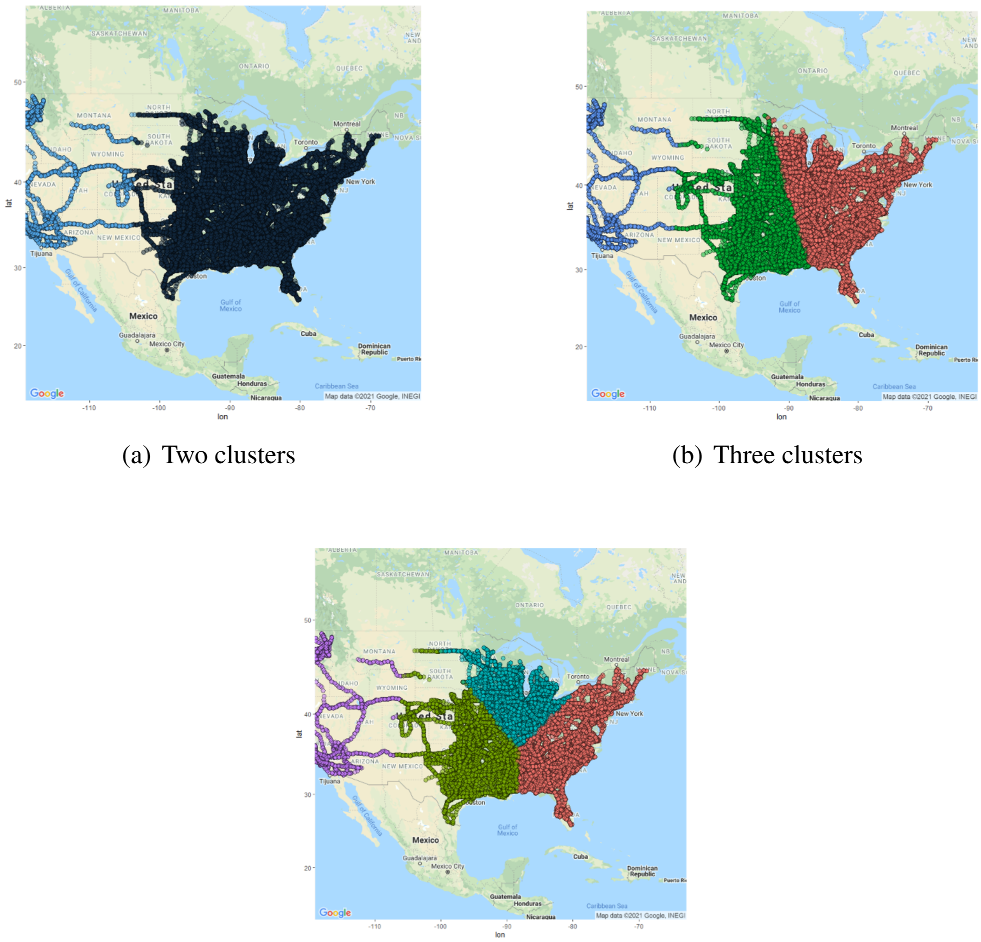
- Investigating the tranferability of the prediction models
- Investigating the effects of drivers' type and geographical regions on the likelihood of safety critical events.
- Developing Multi-class prediction models to be able to predict SCE's type.

5.1.1 Transferability of the prediction models

Spatial Here, we want to see the performance of the trained prediction models in different geographical locations. In chapter four, we trained and tested the prediction models with no restrictions on the geographical locations of the data. Now, we designed several experiments to investigate the spatial transferability of the prediction models in more detail. As the first step, we grouped similar data point based on their GPS locations. To do so, we used the K-mean clustering algorithm as shown in Figure 5.1.

Then, we trained the models on one cluster and tested the models on other clusters. For example, when we have two clusters (Figure 5.1.a), we trained the models on data colored with dark blue and then tested the models on data with red colors. Table 5.1 shows the results.

We repeated the same process while choosing three as the number of clusters (Figure 5.1.b). Table 5.2 shows the results.



(a) Two clusters

(b) Three clusters

(c) Four clusters

Figure 5.1: Clustering results based on latitude and longitude of data.

Table 5.1: The predictive performance of the machine learning models when grouping data into two clusters.

Metric/Model	cart	glm	lasso	nb	nnet	rf	ridge	xgb
AUC	0.786	0.801	0.801	0.751	0.820	0.801	0.790	**0.824**
Accuracy	0.655	**0.743**	0.714	0.722	0.674	0.654	0.705	0.633
Sensitivity	0.778	0.706	0.721	0.673	0.771	0.774	0.714	**0.801**
Specificity	0.653	**0.743**	0.714	0.723	0.673	0.653	0.705	0.631
Gmean	0.713	**0.724**	0.717	0.698	0.721	0.711	0.710	0.711

Finally, we conducted the same experiment while selecting four clusters (Figure 5.1.c) to group data. Table 5.3 shows the results.

According to Table 5.1, Table 5.2, and Table 5.3, we have the following observations:

Table 5.2: The spatial transferability of the machine learning models when grouping data into three clusters.

	cart	glm	lasso	nb	nnet	rf	ridge	xgb
				Train: Red, Test: Green				
AUC	0.672	0.682	0.680	0.700	0.719	0.709	0.670	**0.725**
Accuracy	0.713	0.781	0.779	**0.802**	0.770	0.730	0.740	0.740
Sensitivity	**0.573**	0.433	0.435	0.449	0.516	0.560	0.471	0.570
Specificity	0.713	0.784	0.781	**0.804**	0.771	0.731	0.741	741
Gmean	0.639	0.583	0.583	0.601	0.629	0.640	0.590	**0.650**
				Train: Red, Test: Blue				
AUC	0.787	0.801	0.800	0.695	0.818	0.806	0.781	**0.827**
Accuracy	0.653	0.771	**0.777**	0.771	0.704	0.666	0.733	0.693
Sensitivity	0.784	0.687	0.677	0.590	0.760	**0.785**	0.682	0.773
Specificity	0.652	0.772	**0.778**	0.773	0.703	0.665	0.734	0.692
Gmean	0.715	0.728	0.726	0.675	0.731	0.723	0.708	**0.732**

Table 5.3: The spatial transferability of the machine learning models when there are four clusters.

	cart	glm	lasso	nb	nnet	rf	ridge	xgb
				Train: Green, Test: Blue				
AUC	0.677	0.666	0.669	0.688	0.707	0.690	0.655	**0.713**
Accuracy	0.713	0.739	0.729	**0.785**	0.731	0.743	0.706	0.703
Sensitivity	0.544	0.466	0.473	0.464	0.547	0.521	0.496	**0.595**
Specificity	0.714	0.741	0.731	**0.787**	0.732	0.745	0.708	0.704
Gmean	0.623	0.587	0.588	0.604	0.633	0.623	0.592	**0.647**
				Train: Green, Test: Red				
AUC	0.682	0.670	0.674	0.687	0.709	0.694	0.657	**0.715**
Accuracy	0.737	0.767	0.765	**0.815**	0.767	0.783	0.710	0.755
Sensitivity	0.511	0.442	0.455	0.432	0.491	0.474	0.483	**0.522**
Specificity	0.739	0.769	0.766	**0.817**	0.769	0.785	0.711	0.757
Gmean	0.615	0.583	0.591	0.594	0.615	0.610	0.586	**0.629**
				Train: Green, Test: Purple				
AUC	0.799	0.790	0.795	0.697	**0.822**	0.805	0.771	0.819
Accuracy	0.714	0.786	0.771	**0.795**	0.761	0.746	0.746	0.756
Sensitivity	0.738	0.665	0.687	0.558	0.735	**0.743**	0.652	0.733
Specificity	0.713	0.787	0.772	**0.798**	0.762	0.747	0.747	0.756
Gmean	0.726	0.724	0.729	0.667	**0.748**	0.745	0.698	0.744

- The results are consistent with the previous chapter meaning that SCEs are predictable even the geographical locations of the models are varied.
- In general, the predictive models showed good performance while trained and tested on different geographical locations.
- XGBoost model had the best predictive performance.
- SCEs are more predictable on the west side of the USA.

To further investigate why the predictive performance in the west side of the USA is higher than the east side, we trained the models on 496 regional drivers that we had in the previous chapter. However, we excluded the drivers that drove on the west side and kept it for the testing set. The results are shown in Table 5.4.

Table 5.4: The predictive performance of the machine learning models on regional drivers in the west side of the USA

Metric/Model	cart	glm	lasso	nb	nnet	rf	ridge	xgb
AUC	0.800	0.833	0.829	0.848	0.847	0.846	0.795	**0.853**
Accuracy	0.667	0.748	0.761	**0.771**	0.679	0.669	0.684	0.688
Sensitivity	0.811	0.739	0.728	0.800	0.822	0.800	0.739	**0.844**
Specificity	0.666	0.748	0.761	**0.771**	0.678	0.667	0.683	0.686
Gmean	0.735	0.743	0.744	**0.786**	0.746	0.731	0.711	0.761

We can observe that the the predictive performance in Table 5.4 is higher than Table 4.4. It indicates that SCEs in the west part of the USA are more predictable no matter where the predictive models were trained.

Temporal The previous chapter showed that the predictive models work well enough when data are provided in 30-min intervals. Another aspect of a robust predicate model is to be flexible in terms of time. More specifically, we want to ensure that we can use our predictive models with different time intervals as input. To answer above question, we train the the model in 30-min intervals and then we test the model on data with different time intervals.

Table 5.5 shows the results for XGboost for the first experiment. Note that we picked the XGBoost model because it proved to have the best performance in the previous chapter.

Table 5.5: The temporal predictive performance of the XGBoost model.

Time-Interval (minutes)	AUC	Sensitivity	Specificity
20	0.792	0.776	0.644
30	0.764	0.708	0.672
40	0.754	0.682	0.680
50	0.741	0.677	0.674
60	0.703	0.608	0.675
70	0.714	0.640	0.666
80	0.713	0.632	0.680
90	0.708	0.625	0.664
100	0.718	0.642	0.674
110	0.702	0.622	0.678
120	0.717	0.655	0.670

As we can see in Table 5.5, the predictive performance of the XGBoost model approximately remains consistent in all time intervals. The resultss= indicated that our machine learning framework has temporal transferability. This aspect is crucial because it makes our models more practically usable.

5.1.2 Driver's type

We have 42.0%, 43.9%, and 8.6% local, regional, and over-the-road drivers, respectively, in the dataset that the company provided. We aim to investigate whether or not we can have similar prediction performance with over-the-road (OTR) drivers as regional drivers. To do so, we sampled 500 drivers from OTR drivers datasets and took the following steps:

- Extract weather condition data for OTR data from DarkSky API. (Dark Sky 2019)
- Perform the data preparation steps that were explained in the previous chapter for OTR drivers
- Utilize the machine learning framework which was developed in the previous chapter to predict SCEs for the ORT drivers.

Table 5.6 shows the prediction performance of the ML framework on the ORT drivers. As we can see, the models perform well on the new driver type. It is worth mentioning that even the performance of the models is better when we compare the results with the regional drivers.

Table 5.6: The prediction performance of the ml framework on the ORT drivers.

Metric/Model	cart	glm	lasso	nb	nnet	rf	ridge	svm	xgb
AUC	0.808	0.845	0.840	0.857	0.864	0.856	0.785	0.858	**0.867**
Accuracy	0.889	0.911	0.913	0.901	0.897	0.886	0.807	**0.918**	0.890
Sensitivity	0.687	0.608	0.594	0.675	0.680	0.696	0.583	0.621	**0.690**
Specificity	0.890	0.913	0.914	0.902	0.898	0.887	0.808	**0.919**	0.891
Gmean	0.782	0.745	0.737	0.781	0.782	0.786	0.686	0.756	**0.784**

5.1.3 Prediction of SCE type

As the last step, we want to see the possibility of predicting SCE's type. We believe that the consequence of each SCE type is not the same. Some of them might cause a severe safety issue (e.g., accident), and some of them might cause a minor safety issue (e.g., hard break). Hence, it is essential to be able to identify the precursors that cause each SCEs. In this sub-section, we aim at developing a prediction model to predict SCE's type. We use the same data with the same preprocessing steps that we had in the previous chapter. Next, we took the following steps:

- Develop a prediction model for each specific SCE type.
- Develop a multi-class prediction model where a single classifier is used to predict the class of the predicted value.
- Develop a two-stage model where the first model acts as a binary classifier. Then, if the predicated value is in class "1", the second model aims that identifies its type.

As mentioned before, we have four types of SCE in our data: a) Headway (HW), b) Hard Break (HB), 3) Activation of the rolling stability system (RS), and d) Activation of the forward collision mitigation system (CM). However, due to the limited number of RS and CM, we aggregated them into one class (CMRS). Table 5.7 shows the results for predicting each specific SCE type.

Next, we developed a multi-class prediction model to predict SCE's type. However, due to its poor results, we designed a two-stage prediction model. For the binary classifier, we used XGBoost due to its performance in the previous chapter. Also, a neural network classifier was used for the second stage. Table 5.8 shows the results.

Table 5.7: The predictive performance of the machine learning models on each specific SCE type.

	cart	glm	lasso	nb	nnet	rf	ridge	svm	xgb
HW									
AUC	0.762	0.813	0.813	0.814	0.816	0.816	0.799	0.810	**0.823**
Accuracy	0.849	0.888	0.887	0.814	0.873	0.888	0.877	0.888	**0.897**
Sensitivity	0.590	0.565	0.556	**0.650**	0.591	0.553	0.553	0.543	0.539
Specificity	0.850	0.889	0.888	0.815	0.874	0.889	0.878	0.889	**0.898**
Gmean	0.708	0.708	0.703	**0.727**	0.719	0.701	0.697	0.695	0.696
HB									
AUC	0.758	0.728	0.726	0.781	0.778	0.774	0.725	0.767	**0.789**
Accuracy	0.831	0.741	0.730	**0.878**	0.754	0.867	0.732	0.774	0.868
Sensitivity	0.501	0.546	0.552	0.415	**0.652**	0.436	0.554	0.575	0.466
Specificity	0.832	0.741	0.731	**0.879**	0.754	0.868	0.732	0.775	0.869
Gmean	0.645	0.636	0.635	0.604	**0.701**	0.615	0.637	0.667	0.636
CMRS									
AUC	0.770	0.756	0.760	0.794	0.801	0.819	0.756	0.788	**0.829**
Accuracy	0.808	0.781	0.791	0.750	0.819	**0.911**	0.788	0.822	0.876
Sensitivity	0.586	0.541	0.532	**0.673**	0.627	0.500	0.527	0.591	0.545
Specificity	0.809	0.781	0.791	0.750	0.819	**0.911**	0.788	0.822	0.876
Gmean	0.689	0.650	0.649	0.710	**0.717**	0.675	0.645	0.697	0.691

Table 5.8: The prediction performance of the ml framework on the ORT drivers.

Class/Metric	Sensitivity	Specificity	Balanced Accuracy
CMRS	0.437	0.931	0.686
HB	0.438	0.863	0.651
HW	0.481	0.904	0.693
None	0.699	0.676	0.687

According to Table 5.7 and Table 5.8, we have the following observations:

- Each type of SCE is predictable individually.
- When predicting HW, and CMRS, the models have similar performance. However, the results for predicting HB is a bit lower compare to CMRS and HW classes.
- XGBoost had the best performance.
- Multi-class prediction had poor performance
- Two-stage model improved the performance of the multi-class prediction model significantly.

- The sensitivity results of each class in the two-stage model is considered low.

5.1.4 Conclusion

In this chapter, the robustness of the prediction models developed in the previous chapter was investigated. First, we showed that the predictive has both spatial and temporal transferability. Second, the models were tested for different types of truck drivers. The results showed that the performance of the predictive models remains consistent. These two findings are crucial due to their importance in utilizing the ML framework in the trucking industry. Finally, the possibility of predicting each type of SCE was examined. Our results showed that a two-stage prediction model could achieve a balanced accuracy of around 0.65 for each class of SCEs.

Chapter 6

Summary and Future Research

This chapter summarizes the contributions of this dissertation and discusses the limitations and potential future studies.

In this dissertation, data analytics methods for supply chain risk management are studied. We mainly focus on the risk associated with transportation and manufacturing. The contributions of this dissertation are two-folds: In the first part (Chapter Two), we consider the problem of employing job rotation schemes to improve worker safety in a manufacturing setting by combining optimization methods with novel modeling techniques developed in occupational safety community. The work is based on a recently proposed fatigue-failure model for musculoskeletal disorders (MSD) risk evaluation. This part aims to minimize the likelihood of workers getting injured in manufacturing sections so that the overall safety of a supply chain would increase. Results suggest that the effect of job rotation is highly dependent on the composition of the job pool, and the inclusion of jobs with higher risk results in a drastic decrease in the effectiveness of rotation for reducing overall worker risk. Job rotation alone does not appear to be an effective means of redistributing injury risk between low-risk and high-risk jobs when considering the fatigue-failure perspective. The increase in risk for those in formerly low-risk situations as a result of a rotation scheme can exceed the decrease in risk for a worker in a high-risk situation. Job rotation will thus be counterproductive in terms of overall injury risk in such scenarios. When high-risk jobs are present, the best recommendation remains the use of ergonomic principles to redesign such jobs to reduce injury risk. However, if all jobs in the rotation are relatively low risk, according to our model, workers may derive benefits from rotation such as decreased boredom, increased skill development, and increased motor variability without significant increases to injury risk. One limitation is our assumption that anatomical sites are independent and that there are no constraints on rotation structure (e.g. each worker can perform any amount of each job). By assuming a 'typical' worker and not including personal characteristics (e.g. gender, body mass index, level of experience/training) in the proposed model, the specificity to any particular occupational group is limited. Individual differences and abilities, particularly as they relate to subsequent susceptibility or resistance to MSD injuries, should be considered in future work as the workforce is both aging and increasingly obese. Also, we considered MSDs as the sole safety criterion and the severity of injury was not considered. Of course, other criteria (e.g. occupational noise, exposure to psychosocial risk factors, production rate and the resulting return on investment) come into consideration when deciding whether to adopt job rotation. These effects should certainly be considered in any practical evaluation of job rotation plans. We believe, however, that our findings have relevance to those who may use job rotation for other risks and/or purposes, since, as evidenced by our findings, any rotation scheme has the potential to inadvertently lead to increased physical MSD risk, regardless of potential positive effects of other factors.

The second part (Chapters Three, Four, and Five) aims to improve transportation safety in a supply chain. We primarily consider truck safety as one of the most used transportation methods in a manufacturing-related supply chain. To do so, we first aim to reduce the start-up burden of data collection and descriptive analytics for statistical modeling and route optimization of risk associated with motor vehicles. From a data-driven bibliometric analysis, we show that the literature is divided into two disparate research streams: (a) predictive or explanatory models that attempt to understand and quantify crash risk based on different driving conditions, and (b) optimization techniques that focus on minimizing crash risk through route/path-selection and rest-break scheduling. Translation of research outcomes between these two streams is limited. To overcome this issue, we present publicly available high-quality data sources (different study designs, outcome variables, and predictor variables) and descriptive analytic techniques (data summarization, visualization, and dimension reduction) that can be used to achieve safer-routing and provide code to facilitate data collection/exploration by practitioners/researchers. Then, we review the statistical and machine learning models used for crash risk modeling. We show that (near) real-time crash risk is rarely considered, which might explain why the optimization models have not capitalized on the research outcomes from the first stream. As the next step, we focus on improving the safety of truck drivers by predicting the likelihood of a safety-critical event (SCEs) in the next 30 minutes. We show that SCEs can be predicted 30 min in advance, using machine learning techniques and dependent variables capturing the driver's characteristics, weather conditions, and day/time categories, where an area under the curve (AUC) up to 76% can be achieved. Moreover, the relative importance of the predictor variables was generally consistent based on two distinct variable importance techniques. Specifically, we showed that the kinematic (SCELag7) predictor was the most important, followed by the proxies for traffic-conditions (speedMean and speedSD), while the precipitation-related variables (prepInten and prepProb) were the least important. Hence, we also demonstrated that additional insights can be generated from traditionally black-box machine learning models through utilizing state-of-the-art methods in interpretable machine learning/explainable artificial intelligence (Molnar 2020). Furthermore, the models can be used for both new drivers and future time periods based on our driver-based, and time-based training/sampling scenarios. Specifically, the AUC for the best model was within 0.02 for our best model (xgb) when compared to the base (random) sampling scenario. From an implementation perspective, this difference is practically insignificant.

This dissertation has several limitations that should be highlighted. First, more advanced devices (e.g., wearable sensors) can be used to evaluate workers' safety more accurately in manufacturing systems. Also, there is a need to utilize AI methods in manufacturing systems to make decisions considering the safety criteria. Second, driver-behavioral features (e.g., fatigue, sleepiness) and advanced methods to quantify these features need to be used to have a more comprehensive evaluation of road and driver safety. Third, road geometry featured were not included in our research. Fourth, more investigations are needed to improve multi-class prediction model performance to reveal the precursors of different crash types. Despite these limitations, the findings described in this dissertation highlight some of the understudied risk sources that affect supply chins' overall performance. It also provides practical tools to mitigate these risks and suggest preventive actions. It is my hope that others will expand upon this work and continue to address safety-related outcomes that affect global supply chains.

References

NHTSA. FARS speeding data visualization. United States Department of Transportation, `https://www.nhtsa.gov/press-releases/usdot-releases-2016-fatal-traffic-crash-data`, 2018. [Online; last accessed 7-Sept-2018].

Wei Chen, Fangzhou Guo, and Fei-Yue Wang. A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):2970–2984, 2015.

Nasrin Asgari, Ehsan Nikbakhsh, Alex Hill, and Reza Zanjirani Farahani. Supply chain management 1982–2015: a review. *IMA Journal of Management Mathematics*, 27(3):353–379, 2016.

Hartmut Stadtler, Hartmut Stadtler, Christoph Kilger, Christoph Kilger, Herbert Meyr, and Herbert Meyr. *Supply chain management and advanced planning: concepts, models, software, and case studies*. Springer, 2015.

Premaratne Samaranayake. A conceptual framework for supply chain management: a structural integration. *Supply Chain Management: An International Journal*, 2005.

Faisal Aqlan and Sarah S Lam. Supply chain optimization under risk and uncertainty: A case study for high-end server manufacturing. *Computers & Industrial Engineering*, 93:78–87, 2016.

William Ho, Tian Zheng, Hakan Yildiz, and Srinivas Talluri. Supply chain risk management: a literature review. *International Journal of Production Research*, 53(16):5031–5069, 2015.

George A Zsidisin. A grounded definition of supply risk. *Journal of purchasing and supply management*, 9(5-6):217–224, 2003.

Scott C Ellis, Raymond M Henry, and Jeff Shockley. Buyer perceptions of supply disruption risk: A behavioral view and empirical assessment. *Journal of operations management*, 28 (1):34–46, 2010.

Andrew Neal and Mark A Griffin. A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *Journal of applied psychology*, 91(4):946, 2006.

Dov Zohar and Gil Luria. The use of supervisory practices as leverage to improve safety behavior: A cross-level intervention model. *Journal of safety research*, 34(5):567–577, 2003.

JE Tharaldsen, Espen Olsen, and Torbjørn Rundmo. A longitudinal study of safety climate on the norwegian continental shelf. *Safety Science*, 46(3):427–439, 2008.

Unto Varonen and Markku Mattila. The safety climate and its relationship to safety practices, safety of the work environment and occupational accidents in eight wood-processing companies. *Accident Analysis & Prevention*, 32(6):761–769, 2000.

Sharon Clarke. The relationship between safety climate and safety performance: a meta-analytic review. *Journal of occupational health psychology*, 11(4):315, 2006.

Dov Zohar. A group-level model of safety climate: testing the effect of group climate on microaccidents in manufacturing jobs. *Journal of applied psychology*, 85(4):587, 2000.

D Romero and J Stahre. Towards the resilient operator 5.0: The future of work in smart resilient manufacturing systems. In *54th CIRP Conference on Manufacturing Systems*, 2021.

Toshiyuki Inagaki et al. Adaptive automation: Sharing and trading of control. *Handbook of cognitive task design*, 8:147–169, 2003.

Sean Gallagher and John R Heberger. Examining the interaction of force and repetition on musculoskeletal disorder risk: a systematic literature review. *Human factors*, 55(1):108–124, 2013.

G Mossa, F Boenzi, S Digiesi, G Mummolo, and VA Romano. Productivity and ergonomic risk in human based production systems: A job-rotation scheduling model. *International Journal of Production Economics*, 171:471–477, 2016.

Bureau of Labor Statistics. Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work, 2015. `https://www.bls.gov/news.release/pdf/osh2.pdf`, 2015. [Online; accessed 19-December-2018].

Sang-Young Yoon, Jeonghan Ko, and Myung-Chul Jung. A model for developing job rotation schedules that eliminate sequential high workloads and minimize between-worker variability in cumulative daily workloads: Application to automotive assembly lines. *Applied ergonomics*, 55:8–15, 2016.

Divya Srinivasan and Svend Erik Mathiassen. Motor variability in occupational health and performance. *Clinical biomechanics*, 27(10):979–993, 2012.

Svend Erik Mathiassen. Diversity and variation in biomechanical exposure: what is it, and why would we like to know? *Applied ergonomics*, 37(4):419–427, 2006.

Jonas Sandlund, Divya Srinivasan, Marina Heiden, and Svend Erik Mathiassen. Differences in motor variability among individuals performing a standardized short-cycle manual task. *Human movement science*, 51:17–26, 2017.

Maria Luiza Caires Comper, Jack Tigh Dennerlein, Gabriela dos Santos Evangelista, Patricia Rodrigues da Silva, and Rosimeire Simprini Padula. Effectiveness of job rotation for preventing work-related musculoskeletal diseases: a cluster randomised controlled trial. *Occup Environ Med*, 74(8):543–544, 2017.

Rosimeire Simprini Padula, Maria Luiza Caires Comper, Emily H Sparer, and Jack T Dennerlein. Job rotation designed to prevent musculoskeletal disorders and control risk in manufacturing industries: A systematic review. *Applied ergonomics*, 58:386–397, 2017.

Priscilla C Leider, Julitta S Boschman, Monique HW Frings-Dresen, and Henk F van der Molen. Effects of job rotation on musculoskeletal complaints and related work exposures: a systematic literature review. *Ergonomics*, 58(1):18–32, 2015.

Alexander Vinel, Amir Mehdizadeh, Mark C Schall Jr, Sean Gallagher, and Richard F Sesek. An optimization framework for job rotation to better assess the impact on overall risk. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 843–847. SAGE Publications Sage CA: Los Angeles, CA, 2018.

Sean Gallagher and Mark C Schall Jr. Musculoskeletal disorders as a fatigue failure process: evidence, implications and research needs. *Ergonomics*, 60(2):255–269, 2017.

W Brent Edwards. Modeling overuse injuries in sport as a mechanical fatigue phenomenon. *Exercise and sport sciences reviews*, 46(4):224–231, 2018.

Sean Gallagher, Richard F Sesek, Mark C Schall Jr, and Rong Huangfu. Development and validation of an easy-to-use risk assessment tool for cumulative low back loading: The lifting fatigue failure tool (lifft). *Applied ergonomics*, 63:142–150, 2017.

Sean Gallagher, Mark C Schall Jr, Richard F Sesek, and Rong Huangfu. An upper extremity risk assessment tool based on material fatigue failure theory: The distal upper extremity tool (duet). *Human factors*, 60(8):1146–1162, 2018.

Alena Otto and Olga Battaïa. Reducing physical ergonomic risks at assembly lines by line balancing and job rotation: A survey. *Computers & Industrial Engineering*, 111:467–480, 2017.

Thomas Zahnert. The differential diagnosis of hearing loss. *Deutsches ärzteblatt international*, 108(25):433, 2011.

Alena Otto and Armin Scholl. Reducing ergonomic risks by job rotation scheduling. *OR spectrum*, 35(3):711–733, 2013.

JA Diego-Mas, S Asensio-Cuesta, MA Sanchez-Romero, and MA Artacho-Ramirez. A multi-criteria genetic algorithm for the generation of job rotation schedules. *International Journal of Industrial Ergonomics*, 39(1):23–33, 2009.

Brian J Carnahan, Mark S Redfern, and Bryan Norman. Designing safe job rotation schedules using optimization and heuristic search. *Ergonomics*, 43(4):543–560, 2000.

Wipawee Tharmmaphornphilas, Benjamin Green, Brian J Carnahan, and Bryan A Norman. Applying mathematical modeling to create job rotation schedules for minimizing occupational noise exposure. *Aiha Journal*, 64(3):401–405, 2003.

Wipawee Tharmmaphornphilas and Bryan A Norman. A quantitative method for determining proper job rotation intervals. *Annals of operations research*, 128(1-4):251–266, 2004.

Krisada Asawarungsaengkul and Suebsak Nanthavanij. Design of optimal noise hazard control strategy with budget constraint. *International Journal of Occupational Safety and Ergonomics*, 12(4):355–367, 2006.

J Bhadury and Z Radovilsky. Job rotation using the multi-period assignment model. *International Journal of Production Research*, 44(20):4431–4444, 2006.

Sorawit Yaoyuenyong and Suebsak Nanthavanij. Hybrid procedure to determine optimal workforce without noise hazard exposure. *Computers & Industrial Engineering*, 51(4):743–764, 2006.

Serap Ulusam Seçkiner and Mustafa Kurt. A simulated annealing approach to the solution of job rotation scheduling problems. *Applied Mathematics and Computation*, 188(1):31–45, 2007.

Wipawee Tharmmaphornphilas and Bryan A Norman. A methodology to create robust job rotation schedules. *Annals of operations research*, 155(1):339–360, 2007.

Krisada Asawarungsaengkul and Suebsak Nanthavanij. Heuristic genetic algorithm for workforce scheduling with minimum total worker-location changeover. *International Journal of Industrial Engineering: Theory, Applications and Practice*, 15(4):373–385, 2008a.

Krisada Asawarungsaengkul and Suebsak Nanthavanij. Optimization approach to hazard prevention budgeting problem. *International Journal of Industrial Engineering: Theory, Applications and Practice*, 15(4):330–340, 2008b.

Serap Ulusam Seçkiner and Mustafa Kurt. Ant colony optimization for the job rotation scheduling problem. *Applied Mathematics and Computation*, 201(1-2):149–160, 2008.

Sorawit Yaoyuenyong and Suebsak Nanthavanij. Heuristic job rotation procedures for reducing daily exposure to occupational hazards. *International Journal of Occupational Safety and Ergonomics*, 14(2):195–206, 2008.

MB Aryanezhad, AS Kheirkhah, V Deljoo, and SMJ Mirzapour Al-e hashem. Designing safe job rotation schedules based upon workers' skills. *The International Journal of Advanced Manufacturing Technology*, 41(1-2):193–199, 2009.

Sabina Asensio-Cuesta, JA Diego-Mas, L Canós-Darós, and C Andrés-Romano. A genetic algorithm for the design of job rotation schedules considering ergonomic and competence criteria. *The International Journal of Advanced Manufacturing Technology*, 60(9-12):1161–1174, 2012a.

Ashkan Ayough, M Zandieh, and H Farsijani. Ga and ica approaches to job rotation scheduling problem: considering employee's boredom. *The International Journal of Advanced Manufacturing Technology*, 60(5-8):651–666, 2012.

Sabina Asensio-Cuesta, Jose Antonio Diego-Mas, Lázaro Vicente Cremades-Oliver, and María-Carmen González-Cruz. A method to design job rotation schedules to prevent work-related musculoskeletal disorders in repetitive work. *International Journal of Production Research*, 50(24):7467–7478, 2012b.

Szu-Hao Huang and Ying-Cheng Pan. Ergonomic job rotation strategy based on an automated rgb-d anthropometric measuring system. *Journal of Manufacturing Systems*, 33(4):699–710, 2014.

JooBong Song, Chaiwoo Lee, WonJung Lee, Sangwoo Bahn, ChanJu Jung, and Myung Hwan Yun. Development of a job rotation scheduling algorithm for minimizing accumulated work load per body parts. *Work*, 53(3):511–521, 2016.

Salvatore Digiesi, Francesco Facchini, Giorgio Mossa, and Giovanni Mummolo. A rula-based optimization model for workers' assignment to an assembly line. In *XVII International Scientific Conference on Industrial Systems (IS'17)*, pages 8–13, 2017.

Shib Sankar Sana, Holman Ospina-Mateus, Fabian Gazabón Arrieta, and Jaime Acevedo Chedid. Application of genetic algorithm to job scheduling under ergonomic constraints in manufacturing industry. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–28, 2018.

Sue Hignett and Lynn McAtamney. Rapid entire body assessment (reba). *Applied ergonomics*, 31(2):201–205, 2000.

P Brinckmann, N Johannleweling, D Hilweg, and M Biggemann. Fatigue fracture of human lumbar vertebrae. *Clinical biomechanics*, 2(2):94–96, 1987.

Sean Gallagher, William S Marras, Alan S Litsky, and Deborah Burr. Torso flexion loads and the fatigue failure of human lumbosacral motion segments. *Spine*, 30(20):2265–2273, 2005.

H Schechtman and DL Bader. In vitro fatigue of human tendons. *Journal of biomechanics*, 30 (8):829–835, 1997.

Tishya AL Wren, Derek P Lindsey, Gary S Beaupré, and Dennis R Carter. Effects of creep and cyclic loading on the mechanical properties and failure of human achilles tendons. *Annals of biomedical engineering*, 31(6):710–717, 2003.

Gail M Thornton, Timothy D Schwab, and Thomas R Oxland. Cyclic loading causes faster rupture and strain rate than static loading in medial collateral ligament at high stress. *Clinical Biomechanics*, 22(8):932–940, 2007.

David B Lipps, Edward M Wojtys, and James A Ashton-Miller. Anterior cruciate ligament fatigue failures in knees subjected to repeated simulated pivot landings. *The American journal of sports medicine*, 41(5):1058–1066, 2013.

Jennifer H Shepherd and Hazel RC Screen. Fatigue loading of tendon. *International journal of experimental pathology*, 94(4):260–270, 2013.

Robert Fourer, David M Gay, and Brian W Kernighan. A modeling language for mathematical programming. *Management Science*, 36(5):519–554, 1990.

William S Marras, Steven A Lavender, Sue E Leurgans, Sudhakar L Rajulu, S W Gary Allread, Fadi A Fathallah, and Sue A Ferguson. The role of dynamic three-dimensional trunk motion in occupationally-related. *Spine*, 18(5):617–628, 1993.

Richard Frank Sesek. Evaluation and refinement of ergonomic survey tools to evaluate worker risk of cumulative trauma disorders. 2000.

World Health Organization. WHO — The Top 10 Causes of Death. `http://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death`, 2018a. [Online; accessed 23-February-2019].

National Highway Traffic Safety Administration, NHTSA. U.S. DOT Announces 2017 Roadway Fatalities Down. `https://www.nhtsa.gov/press-releases/us-dot-announces-2017-roadway-fatalities-down`, 2018. [Online; accessed 23-February-2019].

Insurance Institute for Highway Safety. Fatality Facts - IIHS. The Insurance Institute for Highway Safety and the Highway Loss Data Institute, `http://www.iihs.org/iihs/topics/t/general-statistics/fatalityfacts/overview-of-fatality-facts`, 2018. [Online; 23-February-2019].

World Health Organization. WHO — Road Traffic Injuries. `http://www.who.int/mediacentre/factsheets/fs358/en/`, 2018b. [Online; accessed 22-April-2018].

Lawrence Blincoe, Ted R Miller, Eduard Zaloshnja, and Bruce A Lawrence. The economic and societal impact of motor vehicle crashes, 2010 (revised). Technical report, 2015.

World Bank, 2018. GDP (current US $) — Data: United States. World Bank national accounts data, and OECD National Accounts data files. `https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US`, 2018. [Online; accessed 28-April-2018].

Erhan Erkut, Stevanus A Tjandra, and Vedat Verter. Hazardous materials transportation. *Handbooks in Operations Research and Management Science*, 14:539–621, 2007.

Konstantinos N Androutsopoulos and Konstantinos G Zografos. A bi-objective time-dependent vehicle routing and scheduling problem for hazardous materials distribution. *EURO Journal on Transportation and Logistics*, 1(1-2):157–183, 2012.

Mark Abkowitz and Paul Der-Ming Cheng. Developing a risk/cost framework for routing truck movements of hazardous materials. *Accident Analysis & Prevention*, 20(1):39–51, 1988.

Athanasios Theofilatos and George Yannis. A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis & Prevention*, 72:244–256, 2014.

Saman Roshandel, Zuduo Zheng, and Simon Washington. Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. *Accident Analysis & Prevention*, 79:198–211, 2015.

Massimo Aria and Corrado Cuccurullo. bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4):959–975, 2017.

Eugene Garfield and Irving H Sher. Keywords plus™—algorithmic derivative indexing. *Journal of the American Society for Information Science*, 44(5):298–299, 1993.

Matthew S Thiese, Richard J Hanowski, Stefanos N Kales, Richard J Porter, Gary Moffitt, Nan Hu, and Kurt T Hegmann. Multiple conditions increase preventable crash risks among truck drivers in a cohort study. *Journal of occupational and environmental medicine*, 59(2):205, 2017.

Sharon Newnam, Ting Xia, Sjaan Koppel, and Alex Collie. Work-related injury and illness among older truck drivers in australia: a population based, retrospective cohort study. *Safety science*, 112:189–195, 2019.

Thomas A Dingus, Richard J Hanowski, and Sheila G Klauer. Estimating crash risk. *Ergonomics in Design*, 19(4):8–12, 2011.

Feng Guo. Statistical methods for naturalistic driving studies. *Annual Review of Statistics and Its Application*, 6:309–328, 2019.

Federal Highway Administration. Real-Time System Management. Department of Transportation, `https://ops.fhwa.dot.gov/511/index.htm`, 2016. [Online; accessed 03-August-2018].

Feng Guo, Sheila G Klauer, Jonathan M Hankey, and Thomas A Dingus. Near crashes as crash surrogate for naturalistic driving studies. *Transportation Research Record*, 2147(1):66–74, 2010.

Reinier J Jansen and SWOV Simone Wesseling. Harsh braking by truck drivers: A comparison of thresholds and driving contexts using naturalistic driving data. In *Proceedings of the 6th Humanist Conference, Disponibile al link https://bit. ly/2C2Bw3Z [Online*, 2018.

Daniel Mollicone, Kevin Kan, Chris Mott, Rachel Bartels, Steve Bruneau, Matthew van Wollen, Amy R Sparrow, and Hans PA Van Dongen. Predicting performance and safety based on driver fatigue. *Accident Analysis & Prevention*, 126:142–145, 2019.

Lai Zheng, Karim Ismail, and Xianghai Meng. Traffic conflict techniques for road safety analysis: open questions and some insights. *Canadian journal of civil engineering*, 41(7):633–641, 2014.

Carl Johnsson, Aliaksei Laureshyn, and Tim De Ceunynck. In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. *Transport reviews*, 38(6):765–785, 2018.

SM Sohel Mahmud, Luis Ferreira, Md Shamsul Hoque, and Ahmad Tavassoli. Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs. *IATSS research*, 41(4):153–163, 2017.

Ronald R Knipling. Naturalistic driving events: No harm, no foul, no validity. 2015.

Ronald R Knipling. Threats to scientific validity in truck driver hours-of-service studies. 2017.

Juan Guerrero-Ibáñez, Sherali Zeadally, and Juan Contreras-Castillo. Sensor technologies for intelligent transportation systems. *Sensors*, 18(4):1212, 2018.

Sherin Abdelhamid, Hossam S Hassanein, and Glen Takahara. Vehicle as a mobile sensor. *Procedia Computer Science*, 34:286–295, 2014.

Wikipedia contributors. Openstreetmap — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=OpenStreetMap&oldid=900226891`, 2019. [Online; accessed 5-June-2019].

Manuel JA Eugster and Thomas Schlesinger. osmar: Openstreetmap and r. *The R Journal*, 5 (1):53–63, 2013.

Simon P Washington, Matthew G Karlaftis, and Fred Mannering. *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC, 2010.

Nivan Ferreira, Jorge Poco, Huy T Vo, Juliana Freire, and Cláudio T Silva. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158, 2013.

Ben Croxford, Alan Penn, and Bill Hillier. Spatial distribution of urban pollution: civilizing urban traffic. *Science of the Total Environment*, 189:3–9, 1996.

Weiguo Han, Jinfeng Wang, and Shih-Lung Shaw. Visual exploratory data analysis of traffic volume. In Alexander Gelbukh and Carlos Alberto Reyes-Garcia, editors, *MICAI 2006: Advances in Artificial Intelligence*, pages 695–703, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-49058-6.

Ishteaque Alam, Mohammad Fuad Ahmed, Mohaiminul Alam, João Ulisses, Dewan Md Farid, Swakkhar Shatabda, and Rosaldo JF Rossetti. Pattern mining from historical traffic big data. In *IEEE Region 10 Symposium (TENSYMP), 2017*, pages 1–5. IEEE, 2017.

Lalit Sivanandan Nookala. Weather impact on traffic conditions and travel time prediction. M.S. Thesis. Department of Computer Science, University of Minnesota Duluth, `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.332.740&rep=rep1&type=pdf`, 2006. Online; last accessed 04-Sept-2018.

Susan Havre, Beth Hetzler, and Lucy Nowell. Themeriver: Visualizing theme changes over time. In *Information visualization, 2000. InfoVis 2000. IEEE symposium on*, pages 115–123. IEEE, 2000.

Hanqi Guo, Zuchao Wang, Bowen Yu, Huijing Zhao, and Xiaoru Yuan. Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *2011 IEEE Pacific Visualization Symposium (PacificVis)*, pages 163–170. IEEE, 2011.

Yao-Te Tsai, T Alhwiti, Stephen M Swartz, and Fadel M Megahed. Using visual data mining in highway traffic safety analysis and decision making. *Journal of Transportation Management*, pages 43–60, 2015.

Jiansu Pu, Siyuan Liu, Ye Ding, Huamin Qu, and Lionel Ni. T-watcher: A new visual analytic system for effective traffic surveillance. In *Mobile Data Management (MDM), 2013 IEEE 14th International Conference on*, volume 1, pages 127–136. IEEE, 2013.

Zhixiao Xie and Jun Yan. Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, 32(5):396–406, 2008.

Robin Lovelace, Jakub Nowosad, and Jannes Muenchow. *Geocomputation with R*. CRC Press, 2019.

Menno-Jan Kraak. Visualising spatial distributions. *Chapter*, 11:157–173, 1999.

Saffet Erdogan. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of turkey. *Journal of Safety Research*, 40(5):341–351, 2009.

Krist Wongsuphasawat, Michael Pack, Darya Filippova, Michael VanDaniker, and Andreea Olea. Visual analytics for transportation incident data sets. *Transportation Research Record: Journal of the Transportation Research Board*, 2138(1):135–145, 2009.

Siyuan Liu, Jiansu Pu, Qiong Luo, Huamin Qu, Lionel M Ni, and Ramayya Krishnan. Vait: A visual analytics system for metropolitan transportation. *IEEE Transactions on Intelligent Transportation Systems*, 14(4):1586–1596, 2013.

Wei Zeng, Chi-Wing Fu, Stefan Müller Arisona, and Huamin Qu. Visualizing interchange patterns in massive movement data. In *Proceedings of the 15th Eurographics Conference on Visualization*, EuroVis '13, pages 271–280, Chichester, UK, 2013. The Eurographs Association & John Wiley & Sons, Ltd. doi: 10.1111/cgf.12114. URL http://dx.doi.org/10.1111/cgf.12114.

Menno-Jan Kraak. The space-time cube revisited from a geovisualization perspective. In *Proceedings of the 21st International Cartographic Conference*, pages 1988–1996. Citeseer, 2003. Durban, South Africa, August, 2003.

Thomas Kapler and William Wright. Geotime information visualization. *Information Visualization*, 4(2):136–146, 2005.

Brett Romero. Traffic Accidents. http://brettromero.com/traffic-accidents-cyclists/, 2015. [Online; accessed 09-September-2018].

Max Galka. Traffic Accidents. http://metrocosm.com/map-us-traffic/, 2016. [Online; accessed 09-September-2018].

Christian Tominski, Heidrun Schumann, Gennady Andrienko, and Natalia Andrienko. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on visualization and Computer Graphics*, 18(12):2565–2574, 2012.

Michael L Pack, Krist Wongsuphasawat, Michael VanDaniker, and Darya Filippova. Ice–visual analytics for transportation incident datasets. In *Information Reuse & Integration, 2009. IRI'09. IEEE International Conference on*, pages 200–205. IEEE, 2009.

Caitlin D Cottrill and Piyushimita Vonu Thakuriah. Evaluating pedestrian crashes in areas with high low-income or minority populations. *Accident Analysis & Prevention*, 42(6):1718–1728, 2010.

Michael L Pack. Visualization in transportation: challenges and opportunities for everyone. *IEEE Computer Graphics and Applications*, 30(4):90–96, 2010.

Ding Chu, David A Sheets, Ye Zhao, Yingyu Wu, Jing Yang, Maogong Zheng, and George Chen. Visualizing hidden themes of taxi movement with semantic transformation. In *Visualization Symposium (PacificVis), 2014 IEEE Pacific*, pages 137–144. IEEE, 2014.

Hanneke Hooft van Huysduynen, Jacques Terken, Jean-Bernard Martens, and Berry Eggen. Measuring driving styles: a validation of the multidimensional driving style inventory. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 257–264, 2015. Nottingham, United Kingdom, September, 2015.

HaiLong Liu, Tadahiro Taniguchi, Yusuke Tanaka, Kazuhito Takenaka, and Takashi Bando. Visualization of driving behavior based on hidden feature extraction by using deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 18(9):2477–2489, 2017.

Subasish Das, Raul Avelar, Karen Dixon, and Xiaoduan Sun. Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident Analysis & Prevention*, 111:43–55, 2018.

Jarke J Van Wijk and Edward R Van Selow. Cluster and calendar based visualization of time series data. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*, pages 4–9. IEEE, 1999.

Ori Gudes, Richard Varhol, Qian Chayn Sun, and Lynn Meuleners. Investigating articulated heavy-vehicle crashes in western australia using a spatial approach. *Accident Analysis & Prevention*, 106:243–253, 2017.

Ziad Sawalha and Tarek Sayed. Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*, 33(9):1115–1124, 2006.

Qi Shi and Mohamed Abdel-Aty. Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies*, 58:380–394, 2015.

Hany M Hassan and Mohamed A Abdel-Aty. Predicting reduced visibility related crashes on freeways using real-time traffic flow data. *Journal of Safety Research*, 45:29–36, 2013.

Moinul Hossain and Yasunori Muromachi. A real-time crash prediction model for the ramp vicinities of urban expressways. *IATSS Research*, 37(1):68–79, 2013.

Rongjie Yu and Mohamed Abdel-Aty. Utilizing support vector machine in real-time crash risk evaluation. *Accident Analysis & Prevention*, 51:252–259, 2013.

Jinming You, Junhua Wang, and Jingqiu Guo. Real-time crash prediction on freeways using data mining and emerging techniques. *Journal of modern transportation*, 25(2):116–123, 2017.

Franco Basso, Leonardo J Basso, Francisco Bravo, and Raul Pezoa. Real-time crash prediction in an urban expressway using disaggregated data. *Transportation Research Part C: Emerging Technologies*, 86:202–219, 2018.

Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, 2014.

Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, 2007.

David E Goldberg and John H Holland. Genetic algorithms and machine learning. *Machine Learning*, 3(2):95–99, 1988.

James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks*, volume 4, pages 1942–1948. IEEE, 1995. Perth, Australia, 27 Nov-1 Dec 1995.

Chengcheng Xu, Wei Wang, and Pan Liu. A genetic programming model for real-time crash prediction on freeways. *IEEE Transactions on Intelligent Transportation Systems*, 14(2): 574–586, 2013a.

Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar):1157–1182, 2003.

Alan Jović, Karla Brkić, and Nikola Bogunović. A review of feature selection methods with applications. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on*, pages 1200–1205. IEEE, 2015.

Samina Khalid, Tehmina Khalil, and Shamila Nasreen. A survey of feature selection and feature extraction techniques in machine learning. In *Science and Information Conference (SAI), 2014*, pages 372–378. IEEE, 2014.

SM Shiva Nagendra and Mukesh Khare. Principal component analysis of urban traffic characteristics and meteorological data. *Transportation Research Part D: Transport and Environment*, 8(4):285–297, 2003.

Hoe C Lee, Don Cameron, and Andy H Lee. Assessing the driving performance of older adult drivers: on-road versus simulated driving. *Accident Analysis & Prevention*, 35(5):797–803, 2003.

Qu Li, Hu Jianming, and Zhang Yi. A flow volumes data compression approach for traffic network based on principal component analysis. In *2007 IEEE Intelligent Transportation Systems Conference*, pages 125–130. IEEE, 2007.

Ciro Caliendo, Maurizio Guida, and Alessandra Parisi. A crash-prediction model for multilane roads. *Accident Analysis & Prevention*, 39(4):657–670, 2007.

Feng Guo and Youjia Fang. Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*, 61:3–9, 2013.

Jaeyoung Lee, Mohamed Abdel-Aty, and Imran Shah. Evaluation of surrogate measures for pedestrian trips at intersections and crash modeling. *Accident Analysis & Prevention*, 2018.

R Dennis Cook. Principal components, sufficient dimension reduction, and envelopes. *Annual Review of Statistics and Its Application*, 5:533–559, 2018.

Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Kernel principal component analysis. In *International conference on artificial neural networks*, pages 583–588. Springer, 1997.

Pavel Berkhin. A survey of clustering data mining techniques. In *Grouping multidimensional data*, pages 25–71. Springer, 2006.

Pradeep Rai and Shubha Singh. A survey of clustering techniques. *International Journal of Computer Applications*, 7(12):1–5, 2010.

Adil Fahad, Najlaa Alshatri, Zahir Tari, Abdullah Alamri, Ibrahim Khalil, Albert Y Zomaya, Sebti Foufou, and Abdelaziz Bouras. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE Transactions on Emerging Topics in Computing*, 2(3): 267–279, 2014.

Fred L Hall, VF Hurdle, and James H Banks. Synthesis of recent work on the nature of speed-flow and flow-occupancy (or density) relationships on freeways. `https://trid.trb.org/view/1172887`, 1993. [Online; accessed 10-August-2019].

Boris S Kerner and Hubert Rehborn. Experimental properties of complexity in traffic flow. *Physical Review E*, 53(5):R4275, 1996.

Ning Wu. A new approach for modeling of fundamental diagrams. *Transportation Research Part A: Policy and Practice*, 36(10):867–884, 2002.

Thomas F Golob and Wilfred W Recker. A method for relating type of crash to traffic flow characteristics on urban freeways. *Transportation Research Part A: Policy and Practice*, 38 (1):53–80, 2004.

Chengcheng Xu, Pan Liu, Wei Wang, and Zhibin Li. Evaluation of the impacts of traffic states on crash risks on freeways. *Accident Analysis & Prevention*, 47:162–171, 2012.

Thérèse Steenberghen, T Dufays, Isabelle Thomas, and Benoît Flahaut. Intra-urban location and clustering of road accidents using gis: a belgian example. *International Journal of Geographical Information Science*, 18(2):169–181, 2004.

Zhixiao Xie and Jun Yan. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of Transport Geography*, 31:64–71, 2013.

Ling Shen, Jian Lu, Man Long, and Tingjun Chen. Identification of accident blackspots on rural roads using grid clustering and principal component clustering. *Mathematical Problems in Engineering*, 2019, 2019.

Oh Hoon Kwon and Shin Hyoung Park. Identification of influential weather factors on traffic safety using k-means clustering and random forest. In *Advanced Multimedia and Ubiquitous Engineering*, pages 593–599. Springer, 2016.

Michael R Crum, Paula C Morrow, Patricia Olsgard, and Philip J Roke. Truck driving environments and their influence on driver fatigue and crash rates. *Transportation Research Record*, 1779(1):125–133, 2001.

Michael R Crum and Paula C Morrow. The influence of carrier scheduling practices on truck driver fatigue. *Transportation Journal*, pages 20–41, 2002.

Sergio Garbarino, Paolo Durando, Ottavia Guglielmi, Guglielmo Dini, Francesca Bersi, Stefania Fornarino, Alessandra Toletone, Carlo Chiorri, and Nicola Magnavita. Sleep apnea, sleep debt and daytime sleepiness are independently associated with road accidents. a cross-sectional study on truck drivers. *PLoS One*, 11(11):e0166262, 2016.

Thomas A Dingus, Sheila G Klauer, Vicki Lewis Neale, Andy Petersen, Suzanne E Lee, Jeremy Sudweeks, Miguel A Perez, Jonathan Hankey, David Ramsey, Santosh Gupta, et al. The 100-car naturalistic driving study, phase ii - results of the 100-car field experiment. Technical Report DOT HS 810 593, United States. Department of Transportation. National Highway Traffic Safety Administration, April 2006.

Peter McCauley, Leonid V Kalachev, Amber D Smith, Gregory Belenky, David F Dinges, and Hans PA Van Dongen. A new mathematical model for the homeostatic effects of sleep loss on neurobehavioral performance. *Journal of Theoretical Biology*, 256(2):227–239, 2009.

Peter McCauley, Leonid V Kalachev, Daniel J Mollicone, Siobhan Banks, David F Dinges, and Hans PA Van Dongen. Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance. *Sleep*, 36(12): 1987–1997, 2013.

Hal S Stern, Daniel Blower, Michael L Cohen, Charles A Czeisler, David F Dinges, Joel B Greenhouse, Feng Guo, Richard J Hanowski, Natalie P Hartenbaum, Gerald P Krueger, et al. Data and methods for studying commercial motor vehicle driver fatigue, highway safety and long-term driver health. *Accident Analysis & Prevention*, 126:37–42, 2019.

Zachary E Bowden and Cliff T Ragsdale. The truck driver scheduling problem with fatigue monitoring. *Decision Support Systems*, 110:20–31, 2018.

Torbjörn Åkerstedt and Simon Folkard. Validation of the s and c components of the three-process model of alertness regulation. *Sleep*, 18(1):1–6, 1995.

Torbjörn Åkerstedt, Simon Folkard, and Christian Portin. Predictions from the three-process model of alertness. *Aviation, Space, and Environmental Medicine*, 75(3):A75–A83, 2004.

World Health Organization et al. Mobile phone use: a growing problem of driver distraction. 2011.

Kristie Young, Michael Regan, and M Hammer. Driver distraction: A review of the literature. *Distracted Driving*, 2007:379–405, 2007.

Fernando A Wilson and Jim P Stimpson. Trends in fatalities from distracted driving in the united states, 1999 to 2008. *American Journal of Public Health*, 100(11):2213–2219, 2010.

Rebecca L Olson, Richard J Hanowski, Jeffrey S Hickman, Joseph Bocanegra, et al. Driver distraction in commercial vehicle operations. Technical Report FMCSA-RRR-09-042, United States. Department of Transportation. Federal Motor Carrier Safety Administration, 2009.

Sheila G Klauer, Feng Guo, Bruce G Simons-Morton, Marie Claude Ouimet, Suzanne E Lee, and Thomas A Dingus. Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine*, 370(1):54–59, 2014.

Chengcheng Xu, Andrew P Tarko, Wei Wang, and Pan Liu. Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accident Analysis & Prevention*, 57: 30–39, 2013b.

Mohamed Ahmed, Mohamed Abdel-Aty, and Rongjie Yu. Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data. *Transportation Research Record: Journal of the Transportation Research Board*, (2280):51–59, 2012.

Rongjie Yu, Mohamed Abdel-Aty, and Mohamed Ahmed. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. *Accident Analysis & Prevention*, 50:371–376, 2013.

Ling Wang, Mohamed Abdel-Aty, Jaeyoung Lee, and Qi Shi. Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. *Accident Analysis & Prevention*, 122:378–384, 2019a.

Anurag Pande and Mohamed Abdel-Aty. Comprehensive analysis of the relationship between real-time traffic surveillance data and rear-end crashes on freeways. *Transportation Research Record*, 1953(1):31–40, 2006a.

Anurag Pande, Abhishek Das, Mohamed Abdel-Aty, and Hany Hassan. Estimation of real-time crash risk: are all freeways created equal? *Transportation research record*, 2237(1):60–66, 2011.

Athanasios Theofilatos, George Yannis, Pantelis Kopelias, and Fanis Papadimitriou. Impact of real-time traffic characteristics on crash occurrence: Preliminary results of the case of rare events. *Accident Analysis & Prevention*, 2018.

Lei Lin, Qian Wang, and Adel W Sadek. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. *Transportation Research Part C: Emerging Technologies*, 55:444–459, 2015.

Jie Sun and Jian Sun. A dynamic bayesian network model for real-time crash prediction using traffic speed conditions data. *Transportation Research Part C: Emerging Technologies*, 54: 176–186, 2015.

Ling Wang, Mohamed Abdel-Aty, Qi Shi, and Juneyoung Park. Real-time crash prediction for expressway weaving segments. *Transportation Research Part C: Emerging Technologies*, 61:1–10, 2015.

Ling Wang, Mohamed Abdel-Aty, and Jaeyoung Lee. Safety analytics for integrating crash frequency and real-time risk modeling for expressways. *Accident Analysis & Prevention*, 104:58–64, 2017.

Dominique Lord and Fred Mannering. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation research part A: policy and practice*, 44(5):291–305, 2010.

Fred L Mannering and Chandra R Bhat. Analytic methods in accident research: Methodological frontier and future directions. *Analytic methods in accident research*, 1:1–22, 2014.

Azad Abdulhafedh et al. Road crash prediction models: different statistical modeling approaches. *Journal of transportation technologies*, 7(02):190, 2017.

Jiří Ambros, Chris Jurewicz, Shane Turner, and Mariusz Kieć. An international review of challenges and opportunities in development and use of crash prediction models. *European transport research review*, 10(2):35, 2018.

George Yannis, Anastasios Dragomanovits, Alexandra Laiou, Thomas Richter, Stephan Ruhl, Francesca La Torre, Lorenzo Domenichini, Daniel Graham, Niovi Karathodorou, and Haojie Li. Use of accident prediction models in road safety management–an international inquiry. *Transportation research procedia*, 14:4257–4266, 2016.

Yihui Xie, Joseph J Allaire, and Garrett Grolemund. *R markdown: The definitive guide*. CRC Press, 2018.

World Health Organization. Road traffic injuries. `https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries`, 2020. [Last accessed on December 08, 2020].

Wim Wijnen and Henk Stipdonk. Social costs of road crashes: An international analysis. *Accident Analysis & Prevention*, 94:97–106, 2016.

World Health Organization. The top 10 causes of death. https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death, 2018c. [Last accessed on August 17, 2020].

IACP. Large trucks and buses. International Association of Chiefs of Police https://www.theiacp.org/resources/document/large-trucks-and-buses, 2018. [Online; accessed December 2, 2020].

FMCSA. Large trucks and buses by the numbers. U.S. Department of Transportation. https://www.fmcsa.dot.gov/ourroads/large-trucks-and-buses-numbers, 2020. [Online; accessed December 2, 2020].

NHTSA's National Center for Statistics and Analysis. 2017 data: large trucks. U.S. Department of Transportation. National Highway Traffic Safety Administration. Traffic Safety Facts. DOT HS 812 663. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812663, 2019. [Published online January 2019; accessed August 17, 2020].

FMCSA. Large Truck and Bus Crash Facts 2016. https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/safety/data-and-statistics/398686/ltbcf-2016-final-508c-may-2018.pdf, 2018. [Online; accessed 20-February-2019].

Tariq Usman Saeed, Thomas Hall, Hiba Baroud, and Matthew J Volovski. Analyzing road crash frequencies with uncorrelated and correlated random-parameters count models: An empirical assessment of multilane highways. *Analytic Methods in Accident Research*, 23: 100101, 2019.

Apostolos Ziakopoulos and George Yannis. A review of spatial approaches in road safety. *Accident Analysis & Prevention*, 135:105323, 2020.

Amir Mehdizadeh, Miao Cai, Qiong Hu, Mohammad Ali Alamdar Yazdi, Nasrin Mohabbati-Kalejahi, Alexander Vinel, Steven E Rigdon, Karen C Davis, and Fadel M Megahed. A review of data analytic applications in road traffic safety. part 1: descriptive and predictive modeling. *Sensors*, 20(4):1107, 2020.

Rob Eenink, Yvonne Barnard, Martin Baumann, Xavier Augros, and Fabian Utesch. Udrive: the european naturalistic driving study. In *Proceedings of Transport Research Arena*. IFSTTAR, 2014.

Trent Victor, Marco Dozza, Jonas Bärgman, Christian-Nils Boda, Johan Engström, Carol Flannagan, John D Lee, and Gustav Markkula. Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk. Technical Report SHRP 2 Report S2-S08A-RW-1, Transportation Research Board, 2015.

Anurag Pande, Sai Chand, Neeraj Saxena, Vinayak Dixit, James Loy, Brian Wolshon, and Joshua D Kent. A preliminary investigation of the relationships between historical crash and naturalistic driving. *Accident Analysis & Prevention*, 101:107–116, 2017.

Victoria Gitelman, Shlomo Bekhor, Etti Doveh, Fany Pesahov, Roby Carmel, and Smadar Morik. Exploring relationships between driving events identified by in-vehicle data recorders, infrastructure characteristics and road crashes. *Transportation Research Part C: Emerging Technologies*, 91:156–175, 2018.

Miao Cai, Mohammad Ali Alamdar Yazdi, Amir Mehdizadeh, Qiong Hu, Alexander Vinel, Karen Davis, Hong Xian, Fadel M Megahed, and Steven E Rigdon. The association between crashes and safety-critical events: Synthesized evidence from crash reports and naturalistic driving data among commercial truck drivers. *Transportation Research Part C: Emerging Technologies*, 126:103016, 2021.

K Grove, J Atwood, P Hill, G Fitch, A DiFonzo, M Marchese, and M Blanco. Commercial motor vehicle driver performance with adaptive cruise control in adverse weather. *Procedia Manufacturing*, 3:2777–2783, 2015.

Bendix®. Bendix® ABS-6 Advanced with ESP® Stability System - frequently asked questions to help you make an intelligent investment in stability. Bendix Commercial Vehicle Systems LLC, a member of the Knorr-Bremse Group. `https://www.bendix.com/media/documents/products_1/absstability/truckstractors/StabilityFAQ.pdf`, 2007. [Published March 2007; accessed December 08, 2020].

Dark Sky Company. Historical weather data. `https://darksky.net/forecast/40.7127,-74.0059/us12/en`, 2020. [Last accessed on August 17, 2020].

Sungsoon Hwang, Christian Evans, and Timothy Hanke. Detecting stop episodes from gps trajectories with gaps. In *Seeing Cities Through Big Data*, pages 427–439. Springer, 2017.

Yina Wu, Mohamed Abdel-Aty, and Jaeyoung Lee. Crash risk analysis during fog conditions using real-time traffic data. *Accident Analysis & Prevention*, 114:4–11, 2018.

Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, and Wei-Ying Ma. Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, pages 1–10, 2008.

Nicholas Jing Yuan, Yu Zheng, Liuhang Zhang, and Xing Xie. T-finder: A recommender system for finding passengers and vacant taxis. *IEEE Transactions on knowledge and data engineering*, 25(10):2390–2403, 2012.

Jian Sun, Jie Sun, and Peng Chen. Use of support vector machine models for real-time prediction of crash risk on urban expressways. *Transportation Research Record*, 2432(1):91–98, 2014.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

William Revelle, Joshua Wilt, and Allen Rosenthal. Individual differences in cognition: New methods for examining the personality-cognition link. In *Handbook of individual differences in cognition*, pages 27–49. Springer, 2010.

Max Kuhn. *caret: Classification and Regression Training*, 2020. URL https://cran.r-project.org/web/packages/caret/caret.pdf. R package version 6.0-86.

Galit Shmueli. To explain or to predict? *Statistical Science*, 25(3):289–310, 2010.

Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research*, 15(1):3133–3181, 2014.

Jacques Wainer. Comparison of 14 different families of classification algorithms on 115 binary datasets. *arXiv preprint arXiv:1606.00930*, 2016.

James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.

Armin Khayyer, Daniel F Silva, and Alexander Vinel. Predicting public transit arrival times: A hybrid deep neural network approach. *Journal of Big Data Analytics in Transportation*, pages 1–15, 2021.

Gustavo EAPA Batista, Ronaldo C Prati, and Maria Carolina Monard. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 6(1):20–29, 2004.

Chen Wang, Lin Liu, Chengcheng Xu, and Weitao Lv. Predicting future driving risk of crash-involved drivers based on a systematic machine learning framework. *International journal of environmental research and public health*, 16(3):334, 2019b.

Hamidreza Ahady Dolatsara, Ying-Ju Chen, Christy Evans, Ashish Gupta, and Fadel M Megahed. A two-stage machine learning framework to predict heart transplantation survival probabilities over time with a monotonic probability constraint. *Decision Support Systems*, page 113363, 2020.

Athanasios Theofilatos. Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. *Journal of safety research*, 61:9–21, 2017.

Anurag Pande and Mohamed Abdel-Aty. Assessment of freeway traffic parameters leading to lane-change related collisions. *Accident Analysis & Prevention*, 38(5):936–948, 2006b.

Moinul Hossain and Yasunori Muromachi. Understanding crash mechanisms and selecting interventions to mitigate real-time hazards on urban expressways. *Transportation research record*, 2213(1):53–62, 2011.

Amir Bahador Parsa, Ali Movahedi, Homa Taghipour, Sybil Derrible, and Abolfazl Kouros Mohammadian. Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis. *Accident Analysis & Prevention*, 136:105405, 2020.

Licheng Zhang and Cheng Zhan. Machine learning in rock facies classification: an application of xgboost. In *International Geophysical Conference, Qingdao, China, 17-20 April 2017*, pages 1371–1374. Society of Exploration Geophysicists and Chinese Petroleum Society, 2017.

Philippe Barbosa Silva, Michelle Andrade, and Sara Ferreira. Machine learning applied to road safety modeling: a systematic literature review. *Journal of Traffic and Transportation Engineering (English Edition)*, 2020.

Zahra Sedighi Maman, Ying-Ju Chen, Amir Baghdadi, Seamus Lombardo, Lora A Cavuoto, and Fadel M Megahed. A data analytic framework for physical fatigue management using wearable sensors. *Expert Systems with Applications*, page 113405, 2020.

Bin Weng, Lin Lu, Xing Wang, Fadel M Megahed, and Waldyn Martinez. Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications*, 112:258–273, 2018.

Max Kuhn et al. Building predictive models in R using the Caret package. *Journal of statistical software*, 28(5):1–26, 2008.

Christoph Molnar. *Interpretable Machine Learning*. 2020. `https://christophm.github.io/interpretable-ml-book/`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.

Zachary A. Deane-Mayer and Jared E. Knowles. *caretEnsemble: Ensembles of Caret Models*, 2019. URL https://CRAN.R-project.org/package=caretEnsemble. R package version 6.0-84.

Robert J. Hijmans. *geosphere: Spherical Trigonometry*, 2019. URL https://CRAN.R-project.org/package=geosphere. R package version 1.5-10.

Qiong Hu, Miao Cai, Nasrin Mohabbati-Kalejahi, Amir Mehdizadeh, Mohammad Ali Alamdar Yazdi, Alexander Vinel, Steven E Rigdon, Karen C Davis, and Fadel M Megahed. A review of data analytic applications in road traffic safety. part 2: prescriptive modeling. *Sensors*, 20(4):1096, 2020.

Miao Cai, Mohammad Ali Alamdar Yazdi, Amir Mehdizadeh, Qiong Hu, Alexander Vinel, Karen C Davis, Hong Xian, Fadel M Megahed, and Steven E Rigdon. The association between crashes and safety-critical events: synthesized evidence from crash reports and naturalistic driving data among commercial truck drivers. *Transportation Research Part C: Emerging Technologies*, under review(1), 2020.

Dark Sky. Dark Sky API. The DarkSky Company, LLC, https://darksky.net/dev, 2019. [Online; accessed 24-February-2019].