ARTIFICIAL INTELLIGENCE-AIDED PREDICTION OF BROKEN RAIL-CAUSED DERAILMENT RISK

By

KANG ZHOU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Civil and Environmental Engineering

Written under the direction of

Xiang Liu

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2020

ABSTRACT OF THE DISSERTATION

Artificial Intelligence-Aided Prediction of Broken Rail-Caused Derailment Risk

by KANG ZHOU

Dissertation Director:

Dr. Xiang Liu

Broken rails are the leading cause of freight train derailments in the United States. The American railroad industry annually spends billions of dollars on track inspection, maintenance and repair. Accurate prediction of broken rail risk is critical for the railroad industry to further improve operational safety, efficiency and the state of good repair.

This dissertation research focuses on predicting the risk of broken rail-caused derailment via Artificial Intelligence (AI) empowered by the fast-growing "big data" in the railroad industry, related to network-level track characteristics, maintenance activities, traffic and operation, as well as condition monitoring. The intended contributions of this research include:

- Development of a novel, customized Soft Tile Coding based Neural Network model (STC-NN) to predict the spatial-temporal probability of broken rail occurrence for any given time horizon. This proposed AI algorithm shows

superior performance over several alternative algorithms in terms of solution quality, computational efficiency, and modeling flexibility.

- An analysis of the relationship between the probability of broken rail-caused derailment and the probability of broken rail occurrence. New analyses are performed to understand how the probability of broken rail-caused derailment may vary with infrastructure characteristics, signal type, weather, and other factors.
- Development of an Integrated Broken Rail Derailment Risk Model for predicting location-centric broken rail-caused derailment risk on the network-level. Predicting and identifying "high-risk" locations can ultimately lead to significant safety improvement and cost savings.

The major conclusions of this research include:

- The proposed STC-NN algorithm can predict broken rail risk for any time period (from 1 month to 2 years), with better performance for near-term prediction than long-term prediction. The algorithm slightly outperforms Extreme Gradient Boosting, Logistic Regression, and Random Forest, and is also much more flexible.
- Appropriate network segmentation is important for prediction accuracy. Our proposed dynamic segmentation scheme shows a significant improvement over the fixed-length segmentation scheme.
- Segment length, traffic tonnage, number of rail car passes, rail weight, rail age, track curvature, presence of turnout, and presence of historical rail defects are all found to be among influencing factors for broken rail occurrence.

- Signaled track in the cold season has the lowest ratio of broken rail-caused derailments to broken rails, while non-signaled track in warm weather has the highest. Moreover, lower FRA track classes (Class 1, Class 2) have higher ratio of broken rail-caused derailments to broken rails, compared with higher track classes Class 3 and Class 4.
- A longer, heavier train traveling at a higher speed is associated with more cars derailed per broken rail-caused derailment.

This work uses enterprise-level big data for over 20,000 miles of track from a major freight railroad in the United States. The new methodology, algorithm, and analysis results can potentially be implemented for railroad rail asset management, in support of both short-term inspection and maintenance prioritization as well as long-term capital planning and resource allocation.

DEDICATION

Especially to My Parents, I love you.

ACKNOWLEDGEMENTS

The completion of this dissertation could not have been possible without the assistance from many people. Foremost, I would like to express my gratitude to my dissertation advisor, Dr. Xiang Liu for his constant support, guidance and patience throughout my research at Rutgers University. He has been always fully guiding me whenever I ran into challenges during my research path. His wisdom in my research area always extricates me from research dilemma. It is truly an honor and blessing to carry out my research with him as my advisor.

I am also deeply thankful for Dr. Franklin Moon, Dr. Peter J. Jin and Dr. Jingjin Yu serving on my committee. It has been a privilege for me to acquire their guidance, consideration, and encouragement.

Special appreciation is made to Dr. Yuan Wang for his great insights, assistance and encouragement as well as friendship beyond the completion of this dissertation. I am also grateful to Dr. Jihong Chen and Dr. Yun Bai for their guidance and support. I also would like to thank Mr. Zhipeng Zhang for his support in my research. My sincere thanks are extended to Civil Engineering department staffs, Gina Cullari and Linda Szary. Special thanks also go to all my friends and colleagues at Rutgers University for their help in this study: Zheyong Bian, Shuting Zha, and Pengyu Xie.

This research is based on a project sponsored by the Federal Railroad Administration of US Department of Transportation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background and Problem

In 2017, American freight railroads generated almost $26 billion in tax revenues, supported approximately 1.1 million jobs, and generated nearly $220 billion in annual economic activity and $71 billion in wages. Behind huge revenues, there are always potential accidents which damage the railroads' benefits. U.S. freight railroads spent over $660 billion in maintenance and capital expenditures between 1980 and 2017, and over $24.8 billion in capital and maintenance disbursements in 2017 alone (AAR, 2018). Although freight-train derailment rates in the U.S. have been reduced by 44% since 2010, derailment remains a common type of freight train accident in the U.S. According to accident data from the Federal Railroad Administration (FRA) of the U.S. Department of Transportation (USDOT), approximately 6,450 freight-train derailments occurred between 2000 and 2017, causing $2.5 billion worth of infrastructure and rolling stock damage (FRA Rail, 2017).

The FRA of USDOT classifies over 380 distinct accident causes into categories of infrastructure, rolling stock, human factor, signaling and others. The FRA subgroups of accident causes developed by Arthur D. Little (ADL), which combine similar cause codes

into groups based on expert assessment (ADL, 1996), are used in this dissertation. Based on the statistical analysis on the freight-train derailments that occurred on Class I mainline from 2000 to 2017, broken rails or welds have consistently been the leading cause in recent years of all freight-train derailments (Figure 1.1). As a result, broken-rail prevention and risk management have been a major activity for a long time for the railroad industry. In addition to the United States, other countries with heavy-haul railroad activity have also identified the crucial importance of broken rail risk management (Kumar, 2006a; Zarembski, 2009).

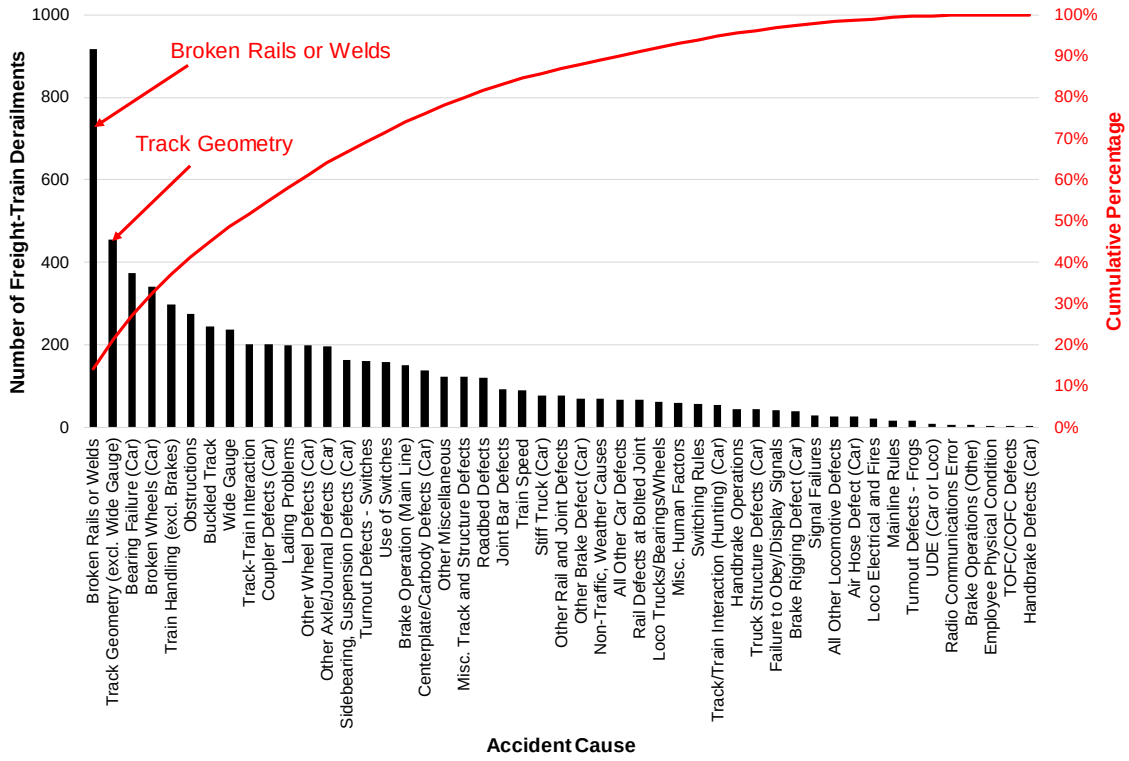


Figure 1. 1 Class I Railroad Mainline Freight-Train Derailment Frequency by Accident Cause Group, 2000 to 2017

Quantifying mainline broken-rail derailment risk and thus identifying the locations with high risk can allow railroads to improve resource allocations for safety management and maintenance optimization. The derailment risk would depend on the probability of the

occurrence of broken-rail derailment and the severity of broken rail-caused derailment that is defined as the number of cars derailed from a train. The number of cars derailed in freight-train derailments is related to several factors, including the train length, derailment speed, and proportion of loaded cars (Liu et al., 2013b).

1.2 Research Motivation

The railroad agencies have undertaken a major task to collect and process railway related data throughout the all railway network including track characteristics (e.g., rail profile information, rail laid information), traffic-related information (e.g., monthly gross tonnage, number of car passes), maintenance records (e.g., rail grinding or track ballast cleaning activities), the past defect occurrences, and many other data sources. In addition, Federal Railroad Administration (FRA) has collected railroad accident data since 1970s. These multi-source data provided the basis for understanding the potential factors that may affect the occurrence of broken rails as well as broken rail-caused derailments. However, there is still limited prior research that takes full advantage of these real-world data to address the relationship between factors and broken rail-caused derailment risk, while using the risk information to screen the network and identify higher-risk locations.

1.3 Research Significance

The U.S. freight railroad network is widely considered as the most dynamic freight systems in the world. The U.S. freight railroad agencies are private organizations that are responsible for their own maintenance and improvement operations. They invest a large proportion of revenues and resources on the maintenance and repair of railroad infrastructure. The optimal expenditure of revenue dollars or resources allocated to railroad

infrastructure becomes crucial for railroad industry which allows to achieve the best outcome with the limited resources. However, the broken rail risk or broken-rail derailment risk accounting for influencing factors is not explicitly quantified. It is expected from this dissertation that railroad agencies can be placed in a better position to allocate safety management resources.

1.4 Research Objective and Scope

The main research objective of this dissertation is to predict the location-specialized broken rail-caused derailment risk using Artificial Intelligence (AI) approaches, more specifically machine learning techniques. Train derailment risk analysis accounts for derailment probability and derailment-caused consequences simultaneously. Due to the low frequency of broken rail-caused derailments, it is desirable to estimate the probability of broken rail-caused derailments through the broken rail occurrence. The estimation of the probability of broken rail-caused derailment includes the conditional probability of train derailment given broken rail occurrence and the probability of broken rail occurrence.

The primary objective of this dissertation is to investigate broken rail (also known as “service failure”) prediction accounting for a set of track-related and operational factors based on railway-provided big data. To accomplish this objective, a number of statistical analyses and machine learning algorithms are conducted. The majority of data related to broken rails from one Class I freight railroad were from 2011 to 2016, covering over 20,000 track miles on mainline tracks.

This dissertation also estimates the conditional probability of train derailment given broken rail occurrence and the estimation of the severity of broken rail-caused train

derailment. Approximately, 4,000 mainline broken rails were identified between 2012 and 2016 on the studied railway company track, whereas during the same period, 25 mainline freight-train broken rail-caused derailments occurred (FRA, 2017). The severity of broken rail-caused derailment, which is defined as the number of cars (both loaded and empty) derailed per derailment in this dissertation, would be estimated based on train-related factors (e.g., train length, speed, tonnage). More specifically, the following objectives of this dissertation can be classified into the following:

1. Review the state of art of studies related to broken rail prediction and broken rail severity;
2. Develop broken rail prediction model using machine learning algorithms accounting for potential factors;
3. Conduct statistical analysis to estimate the probability of broken rail-caused derailments given the broken rail occurrence;
4. Develop a decision tree model to estimate the severity that results from broken rail-caused derailment associated with train-related factors.

1.5 Organization of the Dissertation

The research schematic diagram is shown in Figure 1.2. The dissertation contains seven chapters. A brief description of each chapter will be addressed herein:

Chapter 1 presents the research background and problem statement. The research objectives and scope are identified.

Chapter 2 conducts a comprehensive literature review related to this research scope, including a summary of contributing factors to broken rail prediction, models related to broken rail prediction, and a brief review on broken rail severity analysis.

Chapter 3 discusses the structure and the preprocessing of railroad “big data” from multiple resources to prepare the comprehensive dataset which will be used in the broken rail prediction model. Also, exploratory data analysis is conducted to identify the significant factors contributing to broken rails.

Chapter 4 compares alternative track segmentation schemes and identifies the importance of segmentation for improving the modeling accuracy.

As the main chapter, Chapter 5 develops a novel, customized machine learning model (Soft-Tile-Coding-Based Neural Network) to predict broken rail occurrence by time and location, accounting for a variety of influencing factors. In this chapter, feature engineering, feature transformation, and feature selection are involved.

Chapter 6 analyzes the statistical relationship between broken rail-caused derailment and broken rail occurrence. Conditional probability of derailment given broken rail is obtained. Also, a machine learning model is developed to estimate the severity of a broken rail-caused derailment. Finally, a broken rail-caused derailment risk analysis model is proposed that integrates broken rail occurrence prediction, the conditional probability of derailment given broken rail, as well as derailment severity.

Chapter 7 summarizes the conclusions from this dissertation and recommendation for future work.

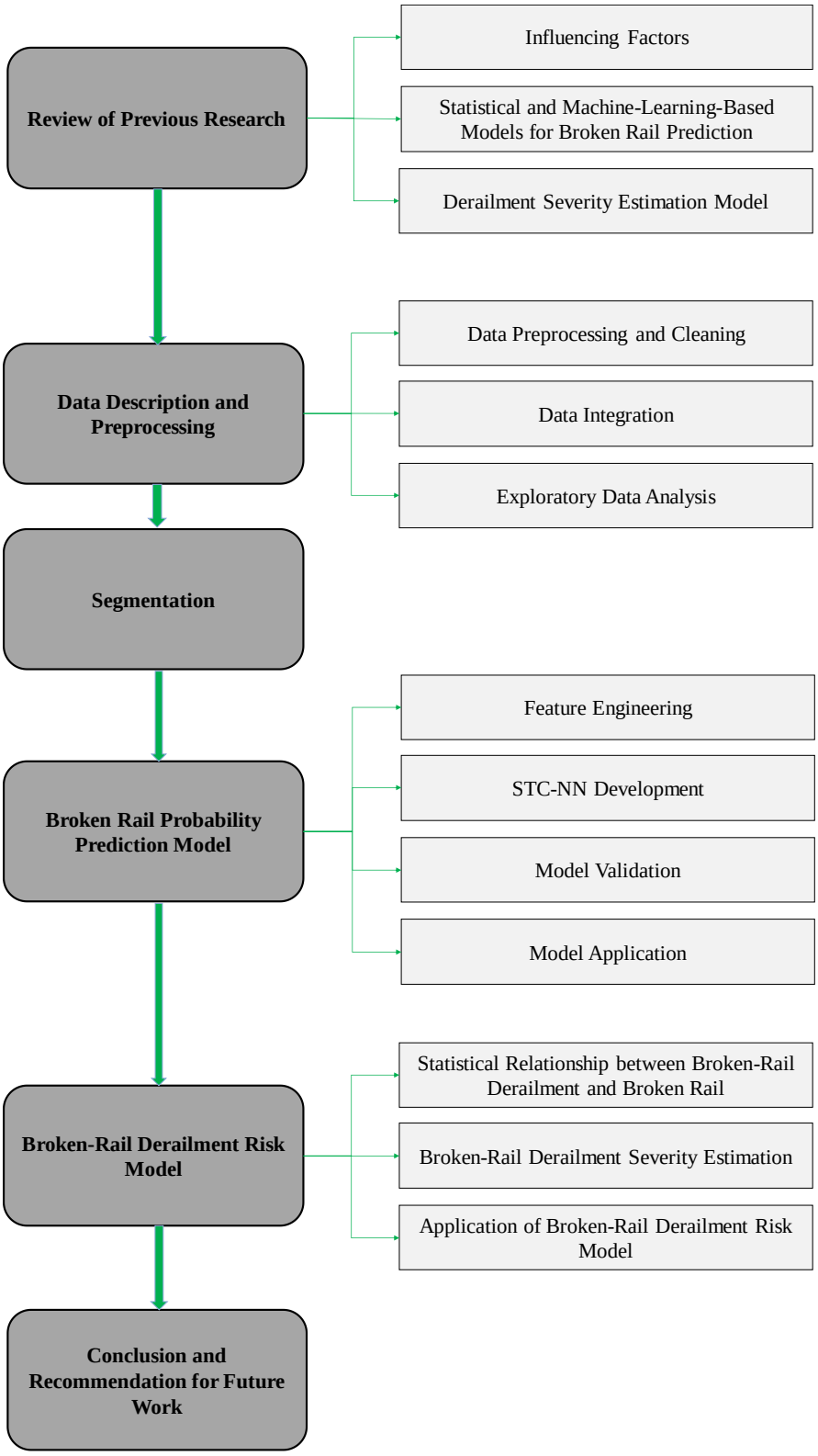


Figure 1. 2 Framework for This Dissertation

CHAPTER 2

LITERATURE REVIEW

This chapter provides an overview of previous broken rail research, focusing on influencing factors of broken rails, statistical and data analytic approaches and machine learning approaches to predicting broken rails, and derailment severity models. A review of mechanistic analyses of rail defects or broken rails can be found in da Silva et al. (2003) and Fischer et al. (2006).

2.1 Introduction

Broken rails originate from various types of rail defects, which can develop in any type of rails as the result of several types of stresses, including bending and shear stress, wheel-rail contact stresses, thermal stresses, and residual stresses (Cannon et al., 2003). These stresses could originate from the rail manufacturing process, cyclical loading, and impact from rolling stock, rail wear and plastic flow. They can exacerbate or produce defects over time, including worn-out rails, defective welds, internal defects, corrugation, and other rolling contact fatigue (RCF) defects, such as surface cracks, head checks, squats, spalling and shelling (Kumar, 2006a). Worn out rails are the result of lateral wear and vertical wear. Lateral wear occurs primarily on the gauge face of the higher rail of a curve. Vertical wear results from cyclical loading and rail grinding on the rail head running surface. Internal defects generally come from inherent flaws in the rail, such as transverse

and compound fissures (Schafer, 2008). Internal defects are commonly small and are only detectable above a certain size. Defects formed due to rolling contact fatigue can be divided into subsurface-initiated cracks and surface-initiated cracks. Subsurface-initiated cracks are often caused by metallurgical defects, which might turn into transverse defects in the rail head. Additionally, increased traffic density and axle load can cause surface-initiated cracks (Kumar, 2006a; Olofsson and Nilsson, 2002). Some surface-initiated cracks might turn into detail fractures from shelling/head checks. RCF surface-initiated cracks might decrease ultrasonic detection effectiveness, which further increases broken rail risk. In summary, there are several common causes that can result in broken rails, including:

- Inherent defects originating from the rail manufacturing process, such as faulty chemical composition, harmful segregation, piping, seams, laps, and guide marks.
- Defects due to a fault of the rolling stock, engine burns, skidding of wheel, severe braking, etc.
- Excessive corrosion of rails: excessive corrosion in the rail generally takes place due to weather conditions, the presence of corrosive salts such as chlorides, and constant exposure of the rails to moisture and humidity in locations near water columns, ashpits, tunnels etc. corrosion normally leads to the development of cracks in regions with a high concentration of stresses.
- Poorly maintained joints: poor maintenance of joints, such as improper packing of joint sleeper and loose fittings.
- Defects in welding of joints: these defects arise because of improper composition of the thermite weld.

- Improper maintenance of track: ineffective or improper maintenance of the track or delayed renewal of the track.
- Rail damages caused by derailments.

The scope of this literature review will focus on broken rails due to defects resulting from the rolling stock and abnormal traffic effects during the rail service.

2.2 Influencing Factors

Most rail defects are detected and treated before they deteriorate into a critical defect (Kumar, 2006a). Broken rails are linked to many factors affecting one or more processes of rail defect development, including defect initiation and/or propagation. This study divides contributing factors into five high-level categories, including 1) track layout, 2) rail characteristics, 3) track maintenance, 4) operational information, and 5) defect inspection history (Figure 2.1). The summary of impact of influencing factors on rail defects or broken rails is described in Table 2.1.

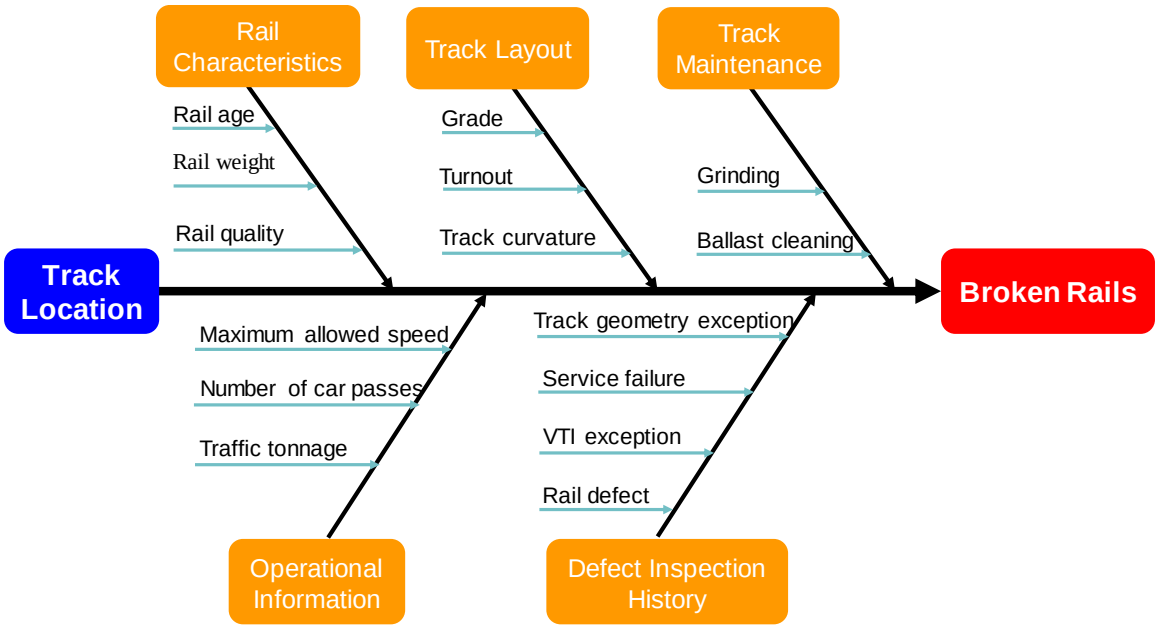


Figure 2. 1 Classification of Selected Contributing Factors

Table 2. 1 Literature by Influencing Factors

Factor	Observation	References
Rail age	Increased probability of rail defects associated with increased rail age.	Chattopadhyay and Kumar, 2009; Dick, 2001; Dick et al., 2002, 2003; Jeong, 2001; Roney and Ebersohn, 2001; Shyr and Ben-Akiva, 1996
Rail weight	As rail weight increases, rail defect probability decreases.	Dick, 2001; Dick et al., 2002, 2003; Hay, 1982; Shyr and Ben-Akiva, 1996
Track curvature	Curved track is associated with higher rail defect probability than tangent track, all else being equal	An et al., 2017; Dick, 2001; Dick et al., 2003

	No consistent conclusion is obtained. Although curved sections of rails have greater rail stress, they usually have a higher frequency of replacement than tangent rails.	Chattopadhyay and Kumar, 2009; Shyr and Ben-Akiva, 1996
Grade	Steep grades increase the risk of rail defect.	An et al., 2017; Dick, 2001; Stock and Pippan, 2011
Maximum allowed speed	Higher maximum allowed speed is associated with a higher probability of rail defect.	Corbin and Fazio, 1981; Dick, 2001; Dick et al., 2002, 2003; IHHA, 2001; Kassa et al., 2006; Kassa and Nielsen, 2008; Kumar, 2006a, 2006b; Reddy, 2004; Shyr and Ben-Akiva, 1996; Sun et al., 2011
	Higher maximum allowed speed is associated with better track geometry, counteracting the effect of higher dynamic wheel load.	Dick 2001; Dick et al., 2002, 2003; Shyr and Ben-Akiva, 1996
Axle load	Increases in axle loads cause more bending and shear stresses in the rail, which might increase	Algan and Gan, 2001; Brouzoulis, 2014; Clayton, 1996; Dick et al., 2003; Esveld, 2001; Farris, 1996;

	dynamic loadings and increase rail defect risk.	IHHA, 2001; Jablonski and Pelloux, 1992; Kim and Kim, 2002; Kumar, 2006a, 2006b; Reddy, 2004; Skyttebol et al., 2005; Zerbst et al., 2009a, 2009b
Traffic density	Higher traffic density causes an increase in rail defects, especially surface-initiated defects.	An et al., 2017; Brouzoulis, 2014; Dick, 2001; Dick et al., 2003; Jeong et al., 1997; Kim and Kim, 2002; Kumar, 2006a, 2006b
Annual wheel passes	Higher number of annual wheel passes is associated with higher rail defect risk.	Algan and Gan, 2001; Brouzoulis, 2014; Dick, 2001; Dick et al., 2003; Kim and Kim, 2002; Shyr and Ben-Akiva, 1996; Skyttebol et al., 2005
Track geometry exception	Presence of geometry exceptions increases probability of rail defects and reduces the life of a rail.	Ahlbeck, 1980; He et al, 2013, 2015; Jenkins et al., 1974; Reddy, 2004; Zarembski and Bonaventura, 2010; Zarembski et al., 2016
Turnout	Presence of turnouts increases the rail defect risk.	An et al., 2017; Dick et al., 2003; Kassa and Nielsen, 2008; Schupp et al., 2004; Sebes et al., 2006; Sun et al., 2011
Rail grinding	Rail grinding might delay the occurrence of rail corrugation	Burstow et al., 2002; Chattopadhyay et al., 2003;

	and reduce the probability of rail defects.	Chattopadhyay et al., 2005; Farris, 1996; Judge, 2000; Kalousek and Magel, 1997; Kumar, 2006a, 2006b; Magel and Kalousek, 2002; Reddy, 2004; Soeleiman and Rucinski, 1991; Shyr and Ben-Akiva, 1996; van den Bosch, 2002; Zarembski et al., 2005; Zarembski, 2005; Zarembski and Palese, 2010; Zhao et al., 2006, 2007a, 2007b
Ballast cleaning	Ballast cleaning reduces the risk of rail defects.	Lichtberger, 2005; Kumar, 2006a, 2006b
Temperature	There is a higher probability of broken rails in colder weather.	An et al., 2017; Chattopadhyay and Kumar, 2009; Dick, 2001; Garnham and Beynon, 1992; Jeong and Gordon, 2009; Jeong et al., 1997; Kumar, 2006a, 2006b; Lichtberger, 2005; Liu et al., 2013a; Muster et al., 1996; Skyttebol et al., 2005; Zerbst et al., 2009b

Joint or welded	Welded rails suffer less impact loading and have lower rail defect risk than joint rails.	Dick et al., 2003; Dick, 2001; Zong et al., 2013
Traction/brake section	Presence of a traction/brake section is prone to causing broken rails.	An et al., 2017
Inspection	More inspections decrease the risk of broken rails.	Dick, 2001; IHHA, 2001; Kumar, 2006a, 2006b; Orringer, 1990; Orringer et al., 1988; Sourget and Riollet, 2006
Lubrication	More lubrication decreases the risk of rail defects.	Thelen, 1996

Rail age: Some researchers (Chattopadhyay and Kumar, 2009; Shyr and Ben-Akiva, 1996) used cumulative million gross tons (MGT) to represent rail age. Alternatively, some others used the number of years that the rail is in place to measure rail age, due to the changes in rail manufacturing technologies (Dick, 2001; Dick et al., 2002, 2003). The lifetime of the track can be translated into the total passage of cyclical loading. Having encountered more cumulative traffic, less service time is left. Prior research has found that the risk of rail failure increases with rail age, due to accumulative fatigue load cycles on the rail, as well as older rail design techniques (Chattopadhyay and Kumar, 2009; Dick et al., 2003; Jeong, 2001; Roney and Ebersohn, 2001; Shyr and Ben-Akiva, 1996).

Rail weight: Rail weight is measured as weight of laid rail per unit length, such as pounds per yard in North America. It is an important factor in determining rail strength.

Rail strength can be considered as either resistance to permanent plastic bending or resistance to fracture (Jeong et al., 1998). Stress in rail is dependent on the rail cross-section and weight (Hay, 1982). Heavier rails have larger cross-sectional area and stronger stiffness, which are associated with higher moment of inertia. These can resist bending and shear stresses under the same axle loads. The risk of broken rails or rail defects decreases on segments of track with relatively heavy rail (Dick et al., 2003).

Track curvature: Curved track (horizontal geometry) experiences more lateral loads from wheel flanges (Dick, 2001) which result in more severe rail wear. Vehicles do not bend to the shape of the curvature while moving over a curve, which results in greater contact friction and stress. Excessive rail wear could damage the capability of the rail to support cyclical loading. Previous research showed that curved rails are more likely to have defects, all other conditions being equal (An et al., 2017; Chattopadhyay and Kumar, 2009; Dick, 2001; Dick et al., 2003; Shyr and Ben-Akiva, 1996). However, a higher frequency of replacement on curved sections of rail than tangent sections can counteract the effect of track curvature (Shyr and Ben-Akiva, 1996).

Grade: In North America, grade (vertical geometry) is expressed in terms of the number of feet of rise per 100 feet of horizontal distance. Steep elevations will increase the probability of broken rails due to increases in longitudinal stress on the rail resulting from both tractive and braking forces (An et al., 2017; Dick, 2001; Stock and Pippan, 2011).

Maximum allowed speed: The literature indicates that higher speed results in a higher probability of broken rails, due to speed-dependent dynamic loading. For example, Dick (2001) found that increase in speed causes an exponential increase in rail failures due

to increases in defects and crack. However, Shyr and Ben-Akiva (1996) found that higher maximum allowed track speed is correlated with better track geometry, thus counteracting the effect of a higher dynamic wheel load.

Axle load: Axle load is the total weight borne by the track from all wheels in contact. Skyttebol et al. (2005) showed that the time to failure of a rail depends on axle load. Impact axle load, caused by vertical acceleration, further increases the probability of rail defects. Zerbst et al. (2009a) found that irregularities in rails can significantly increase impact loading. Zerbst et al. (2009b) stated that impact axle loads can significantly increase bending and shear stresses in the rail and thus increase the probability of rail defects.

Traffic density: Annual traffic density is measured in millions of gross tons (MGT) per year. Higher traffic density causes an increase in rail defects, especially surface-initiated defects (Olofsson and Nilsson, 2002). Dick (2001) found that 75 percent of service failures occur on tracks with annual traffic levels above 100 MGT despite those tracks representing only 12 percent of all rails (Dick, 2001). Dick et al. (2003) further found that increases in traffic density have a significant impact on light rail.

Annual wheel passes: Dick et al. (2003) found that the probability of rail service failure increases as annual wheel passes increase and that, like traffic density, the effect is magnified on the lighter rail (lower rail weight).

Track geometry exception: Track geometry exceptions can be related to rail alignment, cross-level, curvature, overhead lines, gauge, profile, and warp (He et al. 2013, 2015). Jenkins et al. (1974) showed that the presence of geometry exception defects generates an increase in dynamic wheel/rail loads. Zarembski et al. (2016) studied the

relationship between geometry defects and rail defects based on data from two major freight railroads in the United States. They found that the presence of track geometry defects can reduce rail life by an average of 30 percent.

Prior vehicle-track interaction exception: A vehicle-track interaction measurement system assists in the early identification of vehicle dynamics that can lead to rapid degradation of track and equipment. If the measurement exceeds the defined threshold, it is called vehicle-track interaction exception. Vehicle-track interaction could represent the evolution of the vertical track level. The presence of vehicle-track interaction exception is associated with a higher probability of broken rails (Vale and Calcada, 2013).

Turnout: Turnouts are essential components of railway infrastructure, which provide flexibility to traffic operation. Turnout creates an environment of sharp rapid diverging for the train. Dynamic interaction between the train and turnout is more complex than that on tangent and curved tracks. The high impact loads can generate serious damage to the contact surfaces. (Xu et al., 2016). The dynamic interaction between the vehicle and turnout affects the rail wear in railway turnouts, which affects the dynamic interaction in reverse. Due to the abrupt changes in wheel-rail contact points, the profile wear aggravates the wheel-rail dynamic interaction and increases rail degradation as well. Dick et al. (2003) found that proximity to a turnout has a higher probability of a rail service failure. A number of models have been used to simulate the dynamic interaction at the turnout. For example, Sun et al. (2011) used the VAMPIRE model, Kassa et al. (2006) used the commercial multi-body system software GENSYS, Schupp et al. (2004) used the multi-body system simulation pack SIMPACK, and Sebes et al. (2006) used the semi-Hertizian. These

researchers have unanimously found that the presence of a turnout increases the risk of rail defects.

Rail grinding: Rail grinding removes surface defects and track surface irregularities to reshape the rail profile, thus improving rail integrity. It can be used to remove decarbonized layers on new rail, remove plastic flow metal, and remove and control the growth of short and long pitch corrugations, rail surface pitting, spalling, and shells (Kalousek et al., 1989, Cuervo et al., 2015). There are preventive grinding and corrective grinding. Preventive grinding aims to remove the surface irregularities when contact fatigue corrugations grow up to an average rail surface depth of 0.5mm. If crack was not fully ground off by preventive grinding, corrective grinding with multiple passes needs to be performed (Kalousek et al., 1989). Shyr and Ben-Akiva (1996) used a statistical model to show that rail defect rates decrease when grinding frequency increases. Van den Bosch (2002) found that preventative grinding increased rail life, reduced rail corrugation, and curbed the noise of wheel-rail interaction. Chattopadhyay et al. (2005) identified the optimal grinding interval accounting for safety and economics.

Ballast cleaning: Ballast cleaning aims to replace small worn ballast with new ballast. Failure to clean ballast can cause improper drainage, track support, and track position. Kumar (2006a, 2006b) stated that proper ballast cleaning could reduce track stress and thus reduce the rail defect risk.

Temperature: Thermal stresses can influence rail defects and rail failures, especially for continuously welded rails (CWR). Compressions in the rail occur under high temperatures, associated with thermal expansion of constrained rails. Thermal stresses

increase the risk of sudden brittle rail failure, especially in winter season when temperatures are low and the thermally induced tensile force is high. The occurrence of rail failures is affected by ambient temperature (Jeong et al., 1997; Skyttebol et al., 2005), as well as the difference between the rail neutral temperature and the operation temperature (Zerbst et al., 2009b). Dick (2001) found that in colder times of the year, rail defects are more likely to result in broken rails. Liu et al. (2013a) developed a model to quantify the effect of seasonal variation for risk-based inspection optimization. They found that broken rails occur at a higher rate in colder weather. Zerbst et al. (2009b) studied the optimal annual time to inspect rails due to temperature fluctuations by season.

Joint or welded: Segments of rail are either welded together or connected at a joint. Joint systems are favored in locations with low or moderate traffic density but high annual temperature fluctuations, because it allows more room for expansion and contraction (Kumar, 2006a, 2006b). However, rail joints also represent "weak points" in the track system, resulting in a high impact load (Zong et al., 2013), and thereby causing rail defects or failures. Continuous welded rails (CWR) reduce the risks associated with rail joints but increases the risk of the track buckling under high temperatures. Furthermore, the number of welds increases with successive repairs of rail defects and failures (Zhao et al., 2006), and a portion of rail failures also occur due to defective welds (Dick, 2001).

Traction/brake section: A traction/brake section is an area of rail where trains need to accelerate or decelerate. Due to the slipping action of wheels during starting and when brakes are applied to the moving trains, the metal of the top of rails burns, which can result in engine burns or worn rail. An et al. (2017) categorized traction/brake sections as an environmental factor and as more prone to breakage due to higher longitudinal stresses.

Inspection: Considering the ultimate consequences of rail flaws or defects resulting in broken rails, it is necessary to detect these flaws and take timely action to remedy them. Rail flaws can be detected either by visual inspection or by rail flaw detection equipment (e.g., non-destructive testing methods). The periodicity of rail flaw detection depends on the sectional gross tonnage of the rail, speed limit, and other factors. Possibly, prior to deteriorating into broken rails, some rail defects might be classified as undetectable at the time of rail flaw testing because defects might have been too small to be detected, or the surface condition of the rails might have presented additional "noise" that masks the defects. Increasing testing frequency or using more advanced ultrasonic/induction equipment may lower this possibility. Inspection effectiveness depends on the inspection technology and inspectors' experience (Kumar, 2006b). A higher inspection frequency would mean that inspectors catch crack occurrences and propagation earlier, which allows for earlier remediation. Therefore, a higher inspection frequency would reduce the occurrence of broken rails (but associated with higher inspection costs). Sourget and Riollet (2006) used a statistical tool to optimize rail defect inspection frequency.

Lubrication: The function of lubrication is to oil the running face of the outer rail in order to reduce friction. It has been noticed that lubrication considerably reduces wear by up to 50%. Thelen (1996) demonstrated that proper lubrication of rails can help reduce wheel-rail interface friction, and thus reduce RCF and rail failures. This study also suggested that the optimal lubrication applied to the rail should be based on the coefficient of friction on the segment.

2.3 Statistical Models for Broken Rail Prediction

Understanding the relationship between broken rail risk and contributing factors has been a core focus of broken rail risk management. A closed-form integrated engineering model, which reflects the multivariate nature of broken rail risk, is complex and difficult to generalize to specific operating scenarios. Therefore, many researchers have attempted to simultaneously analyze the effects of multiple factors in the deterioration process of rails using statistical (data analytic) models. Reviewing such data-driven models is the focus of this chapter. Table 2.2 provides an overview of the most significant research undertaken overseas on the modeling of broken rail-related issues.

Table 2. 2 Statistical Prediction Model

Statistical Model	References	Advantages	Limitations
Linear regression model	Aglan and Gan, 2001; Jovanovic, 2004	Able to provide a linear relationship between output variable and input variables; Easy to apply	Limited number of features were involved; Only estimating linear relationship between output variable and input variables; Sensitive to output outlier; Only suitable for the prediction of continuous dependent output
Logistic	Dick, 2001; Dick	Able to involve the	Difficult to develop the

regression model	et al., 2002; 2003; Schafer, 2008; Sourget and Riollet, 2006	interaction among input features; Able to not only predict the probability of broken rails on a segment but also estimate the number of broken rails on a longer track section	complex function between explanatory variables and the response variable; Ignoring the censored events; Not able to consider the time-dependent information
Hybrid ANN/Logistic model	Schafer, 2008; Schafer and Barkan, 2008a, 2008b	Able to account for all possible interactions among input features	Unable to present quantitative functional relationship between input variables and output variable, which limits the ability to be explained
Cox proportional hazard regression model	An et al., 2017	Assessment of the impact of features on the entire lifetime of rails; Able to consider the censored event	Requires prior assumption of the reference rail break risk rate at a given time. Ignores the change of the explanatory factors over time
Weibull model	Chattopadhyay and Kumar,	Able to consider the censored event;	Relies on the assumption that failure time follows a

	2009; Shyr, 1993; Shyr and Ben-Akiva, 1996	Able to account for the time-dependent information	known Weibull distribution
Markov Stochastic process and hazard model	Bai et al., 2015, 2017; Hokstad et al., 2005	Presents the probability matrix from one state to multiple states, rather than just focusing on the service failure	Finite rail condition states are considered; One-to-one correspondence between Markov transition probabilities and period interval was assumed
Fuzzy logic model	Veskovic et al., 2012	Resembles human decision making; Enables making decisions based on incomplete and imprecise information	Difficult to define the membership function; Only focuses on the analysis at a fixed time rather than the lifetime of rail; Finite condition states are considered
Reliability model	Zhao et al., 2006, 2007a	Does not rely on big data; Able to describe the impacts of certain activities (e.g., grinding, inspection) on the occurrence of	Relies on accurate information regarding the probabilities and hazard rate of fatigue defects; Unable to account for the impact of other track characteristics on broken

	broken rails	rails except for cumulative gross tonnages

The following paragraphs describe each widely used statistical approach for broken rail prediction.

2.3.1 Linear Regression

Aglan and Gan (2001) studied fatigue crack growth behavior of a premium rail steel using a modified crack layer theory. The rate of energy expended on damage formation and evolution within the active zone was evaluated. In their study, the crack growth was simulated in the laboratory with the determined maximum fatigue stress and stress ratio. Constant loading was forced on the track. The simulation data in the laboratory was collected and regressed to get three phases of crack growth: crack initiation, stable crack growth, and unstable crack growth. In the stable crack growth process, the crack growth and the cumulative traffic tonnage have a linear relationship. In their study, only traffic was involved in the simulation, without considering the impact of other potential factors on crack growth.

Overall, three limitations of linear regression analysis are as follows:

- It only considers the measured parameters related to track geometry, which does not directly reveal the impact of operational characteristics (train speed, annual traffic tonnages, etc.) and other rail characteristics (rail age, turnout, bridge, etc.) on track deterioration.

- It assumes a simple linear relationship between the independent variables and dependent variables, which cannot capture possibly more complex relationships. It is also unable to structure the correlation and interaction between independent variables.
- It is more suitable for the prediction of continuous output, rather than the prediction of discrete events like the probability of the occurrence of rail defects or broken rail defects.

2.3.2 Logistic Regression

The logistic regression model is a statistical approach where the dependent variable is categorical. It measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function. The logistic regression model has been used to describe service failure probabilities by combining multiple variables to identify the multivariate correlations (Dick, 2001; Dick et al., 2002; Dick et al., 2003, Schafer, 2008; Schafer and Barkan, 2008a, 2008b, Sourget and Riollet, 2006). Stepwise regression was used to determine the most relevant parameters or combinations of parameters. A probability threshold is determined to estimate whether one broken rail is predicted to occur on the location. Logistic regression can provide information to interpret the influence of each parameter on the dependent variable. However, the limitations of this approach include:

- It is difficult to develop the complex function between explanatory variables and the response variable.
- It ignores the role of time for the event to happen.

- It is unable to consider time-dependent information.

2.3.3 Artificial Neural Network

Artificial neural networks (ANN) have been used in various studies of event prediction and classification. ANNs are computational tools that have the ability to learn mathematical relationships between a series of input variables and their respective output values. Schafer and Barkan (2008a; 2008b) developed an artificial neural network model for classifying track segment locations as either failure or non-failure. They add hidden neurons one by one until the optimal network is obtained, in which the model accuracy and generalization can achieve an efficient tradeoff.

There are advantages and disadvantages using artificial neural network as a classification tool. In terms of advantages, ANN is able to capture the nonlinear relationship between input and output variables and possible interactions between inputs variables. It also allows for customizing the objective function and learning techniques in the neural network. The main disadvantage of the artificial neural network is that it is difficult to present a clear functional relationship between input and output variables, which limits its explanatory ability. To overcome some of the disadvantages of simple ANN, researchers proposed two types of hybrid model which combined artificial neural network and logistic regression (Schafer, 2008; Schafer and Barkan, 2008a; 2008b). The first hybrid model applied a logistic regression model to pre-select the most useful variables and then developed an ANN model based on these selected variables. The second hybrid model was developed using the logistic regression model to calculate the probability of broken rails and added the calculated probability as an additional input parameter into the ANN model.

The disadvantage of the hybrid model is that sometimes it might cause overfitting, which reduces its efficiency when applied to unseen data.

2.3.4 Survival Analysis

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of broken rails. In the existing researches, the time to occurrence of rail breaks (so-called survival time) can be measured in cumulative tonnages. As mentioned above, linear regression has been used to model continuous output variables as a function of a set of predictor variables. However, linear regression is not suitable for the modeling of survival time, due to its inefficiency at handling the censored observations which represent the information of survival time is incomplete. Suppose tracks were surveilled in a study for a period. A track which did not experience the occurrence of rail breaks is said to be censored.

Several survival models are available to analyze the relationship between predictor variables and the survival time, including parametric, nonparametric and semiparametric approaches. Parametric methods assume that the underlying distribution of the survival times follows known probability distributions, including parametric, nonparametric, and semiparametric approaches. Parametric methods assume that the underlying distribution of survival times may follow known probability distributions, such as exponential or Weibull distribution. Non-parametric models which do not represent a functional relationship between survival times and predictor variables have not yet been used in broken rail analysis.

2.3.4.1 Weibull Distribution

Exponential distribution assumes a constant hazard rate, which is not consistent with empirical studies. Most of the other distributions used for parametric models do not have closed-form expressions for the hazard rate in terms of predictor variables. As a result, one commonly used distribution is the Weibull distribution with a scale parameter and shape parameter. Weibull distribution is also a generalized form of the exponential distribution.

Chattopadhyay and Kumar (2009) collected field data over a period of time and developed a Weibull distribution model to estimate rail degradation accounting for operational conditions, curvature, track, and environmental conditions. However, the derived hazard function was only the function of cumulative tonnages. The collected field data was divided into different groups according to different combinations of other contributing variables. Corresponding failure models were developed related to each group. Due to the limitation of the collected field data, most of the groups had no rail failure records. In their research, only the broken rail events were observed, with non-broken rails being ignored. Therefore, the models regressed by the limited data might have a magnitude of variances and biases.

Shyr and Ben-Akiva (1996) also used Weibull distribution to structure the relationship between the probability of broken rails and the cumulative gross tonnages. In addition, the model was extended to multiple types of broken rails. It was assumed that the different types of broken rails were independent, and the overall hazard rate was derived as the sum of hazard rates from all potential types of broken rails. Furthermore, Shyr and

Ben-Akiva (1996) also included variables, such as maintenance, metallurgy, and operation activities, in the hazard functions, in addition to cumulative gross tonnages. These variables may change over time. Theoretically, the hazard rate was derived as a function of time-dependent features. However, in their research, to make the model simple, the values of features were assumed to change at most once a year. One limitation of their research is that it only focused on the progressive deterioration of track from one defect to another more severe defect, rather than the whole life cycle of track deterioration.

2.3.4.2 Cox Proportional Hazard Regression

The Cox proportional hazard regression model (hereafter referred to as Cox regression model) is used to present the rail break risk associated with the effects of various influencing factors. It can calculate the hazard rate at any particular time given specific explanatory variables. Here the time is measured by the cumulative tonnages passing on the rail. An et al. (2017) developed a grid-based (segment-based) model incorporating the Cox regression model to evaluate the effect of explanatory factors on broken rail risk. The coefficients of the explanatory variables can be determined by the maximum-likelihood estimation method, which maximizes the probability with which rail breaks on grid references occur simultaneously. The coefficients of the factors in the prediction model can be interpreted as the corresponding change of the broken rail risk with respect to variations of each factor (An et al, 2017). The strength of this reference is that it considered human factors (e.g., missing detection) and environmental factors (e.g., climate). Even though the Cox regression model can calculate the rail break risk given the cumulative tonnages of rails and the corresponding features, there are still some limitations of the Cox regression model. First, it requires prior assumption of the reference rail break rate at a given time,

which depends on the rail characteristics and necessitates additional information from experts, which is subject to simplification or human errors. Second, Cox regression model ignores the change of the explanatory factors over time.

2.3.5 Markov Stochastic Model

Hokstad and Langseth (2005) modeled the rail degradation process within the fixed inspection interval as a Markov chain. The final state of the Markov chain is rail failure. The rail failure is divided into two categories: sudden failure and gradual degradation to failure. Any previous degraded states might transition into sudden failure and only the last degraded state could degrade into the gradual failure. In Hokstad and Langseth (2005)'s research, the transition probabilities were estimated against the actual inspection and failure data. However, the probabilities are not related to any input variables (e.g., curvature, speed, traffic density, and others), which are constant regardless of the associated characteristics of the locations. Meanwhile, the efficiency of Hokstad and Langseth's research is subject to the efficiency of the inspection.

Bai et al. (2015, 2017) conducted research evaluating the deterioration of the rail. Understanding that the deterioration process is uncertain, a Markov stochastic model was used to describe the probabilities of the transition from one deterioration state to another. A linear relationship was developed between the Markov transition probabilities and the heterogeneity of the track section, such as gross tonnages and curvature. The parameters in the linear function were estimated via a maximum log-likelihood function. The Markov transition probabilities represent the probability from one degradation state to another after the constant inspection interval. However, there are still some disadvantages of the Markov

stochastic model in that research. First, a limited number of deterioration states were considered in Markov stochastic model. Theoretically, the track degradation process should be continuous, and track deteriorates gradually as the cumulative gross tonnage increases. Second, that research assumed a one-to-one correspondence between Markov transition probabilities and research period interval (e.g., half a year). The Markov transition probabilities matrix in their research were estimated based on the collected field data whose inspection period is constant. If the inspection period changes, the probabilities must be re-calculated using updated inspection data.

2.3.6 Fuzzy Logic Model

The fuzzy logic model is a good mathematical tool for modeling a process that is distinguished by subjectivity, uncertainty, ambiguity, and imprecision. Vesković et al. (2012) used the fuzzy logic model to predict the frequency of broken rails in terms of input variables. Four input variables are involved in the fuzzy logic model, including temperature, rail age, gross tonnage, and curve degree. The membership function for output variable, number of broken rails, involves three fuzzy sets - low, mid, and high. The fuzzy logic has the following advantages: 1) it enables making decisions based on incomplete and imprecise information, and 2) it does not rely on the previous states. However, there are still some disadvantages to use the fuzzy logic model:

- The fuzzy logic model works efficiently to predict the number of broken rails on track sections only when the length of track sections is large. As a result, the accurate prediction on a small track section is not available.

- The membership function of the fuzzy logic model should be defined based on expert knowledge.
- The fuzzy logic model only focuses on analysis at a fixed time. It cannot reflect the rail deterioration process during the entire lifetime of the rail.

2.3.7 Reliability Model

Zhao et al. (2006, 2007a) applied a reliability model to predict the number of derailments by modeling the development of rail defects leading to rail breaks as well as derailments. The risk of derailment is measured by the expected number of rail breaks multiplied by the severity of rail break. Four sub-models were integrated to predict the occurrence of rail defects and breaks, including a weld defect prediction model, fatigue model, impact of grinding model, and imperfect inspection model. They assumed that rail fatigue defects follow a Weibull distribution and imperfect inspection follows a non-homogeneous Poisson process. This research is able to calculate the probabilities of weld defects and fatigue defects, respectively. Accounting for the influence of the imperfect inspection and grinding maintenance, the probability of broken rails was obtained with pre-defined parameters. The difficulty in using the reliability model is the pre-definition of the probability required in the model. In addition, Zhao's reliability model failed to account for the impact of other features (track characteristics and defect histories) on the broken rails except for the gross tonnage.

2.4 Machine-Learning-Based Model

With the support of fast-growing volume and diversity of available data as well as the advanced big data analytics, the application of machine learning-based models has been

prevailing in the recent researches. Cha and Choi (2017) applied convolutional neural network to detect rail crack based on crack damages. This approach is immune to shooting conditions of test images. Support vector machine method was used for non-destructive detection of rail defects based on vibration signals which are resulted from rail defects (Sun et al., 2014). The research result proved the effectiveness of support vector machine method. In addition, extreme gradient boosting algorithm (XGBoost) was developed to address the impact of track geometry on development of rail defects (Mohammadi et al., 2019). This approach correctly predicted about 83% of rail defects based on tonnage and track geometry data. Tantithamthavorn et al. (2016) proposed automated parameter optimization of machine learning classification techniques for defect predictions. They concluded that parameter settings can indeed have a large impact on the performance of defect prediction models.

2.5 Derailment Severity Estimation Model

Simulation and statistical analysis are the two basic approaches used in previous studies to model train derailment severity. Simulation models predict the response of railroad vehicles to specific track and environmental conditions. These models are typically based on nonlinear wheel-rail interaction models. For example, Yang et al. (1972a, 1973b) developed a simulation model to determine the effect of ground friction, mating coupler moment, and brake retarding force on the number of cars derailed. They found that the position of the first car involved in the derailment (called point-of-derailment, or POD) and derailment speed could affect the number of cars derailed (Yang et al., 1972a, 1973b). In the late 1980s, Yang et al.'s model was extended by considering coupler failure and independent car motion (Coppens et al., 1988; Birk et al., 1990). The precision of

simulation models is subject to the accuracy of modeling train derailment dynamics. In addition to simulation models, train derailment severity can also be estimated based on historical data. Saccomanno and EI-Hage (1989, 1991) developed a truncated geometric model to estimate the mean number of cars derailed as a function of derailment speed, residual train length, and accident cause. Finally, all previous derailment severity models focused on analyzing the mean number of cars derailed. Liu et al. (2013b) developed a zero-truncated negative binomial (ZTNB) regression model to estimate the conditional mean of train derailment severity with historical derailments on U.S. Class I railroad mainlines from 2001 to 2010. Meanwhile, a quantile regression (QR) model is also developed to estimate derailment severity at different quantiles. However, these models may involve several assumptions (e.g., over-dispersed count variables) in the processed datasets. Martey and Attoh-Okine (2018) employed vine copula quantile regression model to predict conditional mean and quantiles of derailment severity and found that vine copula quantile regression model performed better at predicting at various quantile levels than the classical quantile regression approach.

CHAPTER 3

DATA DESCRIPTION AND PREPARATION

3.1 Introduction

A wide variety of field data collections were conducted to better understand the condition and behavior of various rail assets and they play a pivotal role in this research. The collected data comes from two sources: the FRA accident database and enterprise-level "big data" from one Class I freight railroad. The broken-rail derailment data comes from the FRA accident database, which records the time, location, severity, and contributing cause(s) of each train accident. Using this database, broken rail-caused freight train derailment data on the main tracks of the studied Class I railroad could be obtained for analyzing the relationship between broken rail and broken rail-caused derailments, as well as broken-rail derailment severity. The data provided by the railroad company includes: 1) traffic data; 2) rail testing and track geometry inspection data; 3) maintenance activity data; and 4) track layout data. Fifteen raw databases were received and listed in Table 3.1 below for the entire railway network covering the periods from 2011 to 2016.

The railroad network contains over 20,000 track miles including main lines (to include multiple tracks on the mains which are laid in parallel), siding, and yard tracks. The research in this dissertation focuses on the main line tracks which include single track, and multiple tracks. The details are listed in Table 3.2.

Table 3. 1 Summary of Provided Railroad Data

Dataset	Description
Rail Service Failure Data	Broken rail data from 2011 to 2016
Rail Defect Data	Detected rail defect data from 2011 to 2016
Track Geometry Exception Data	Detected track geometry exception data from 2011 to 2016
VTI Exception Data	Vehicle-track interaction exception data from 2012 to 2016
Monthly Tonnage Data	Gross monthly tonnage and car pass data from 2011 to 2016
Grinding Data	Grinding pass data from 2011 to 2016
Ballast Cleaning Data	Ballast cleaning data from 2011 to 2016
Track Type Data	Single track and multiple track data
Rail Laid Data	Rail laid year, new rail versus re-laid rail, and rail weight data
Track Chart	Track profile and maximum allowed speed
Curvature Data	Track curvature degree and length
Grade Data	Track grade data
Turnout Data	Location of turnouts
Signal Data	Location and type of rail traffic signal
Network GIS Data	Geographic information system data for the whole network

Table 3. 2 Track Mileage for Each Type of Main Line Track

Track Type	Track Miles
Single Track	13,513
Multiple Tracks	7,126

3.2 Database Description

As noted above, fifteen datasets are collected and provide raw supplementary feature information for a specific track location. Data is collected at a specific point on the rail or on smaller segments based on the type of measure. The details of each dataset are presented in the following sections.

3.2.1 Track File

This database includes the type of a specific segment. It specifies the starting and ending milepost locations by prefix and track number. The track file database is used as a reference database to overlay all other databases (Table 3.3). It contains more than 500 prefixes.

Table 3. 3 Track File Format

Prefix	Starting Milepost	Ending Milepost	Track Type

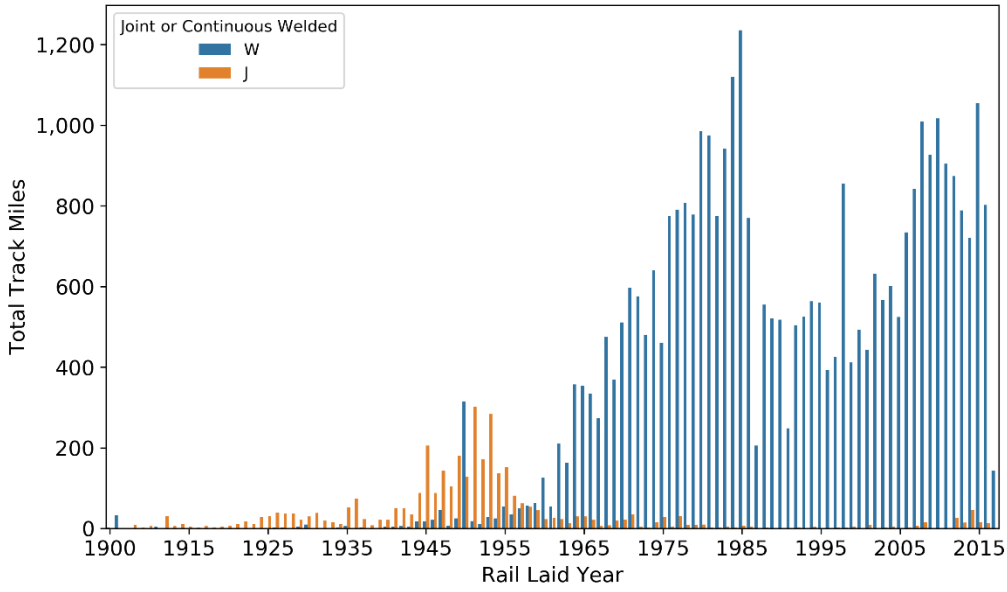
3.2.2 Rail Laid Database

This database contains rail weight, new rail versus re-laid rail, and joint versus continuous welded rails (CWR) (Table 3.4). Figure 3.1 illustrates the total rail miles in terms of rail laid year and rail type (jointed rail versus CWR). The figure shows that most welded rails were laid after the 1960s and most joint rails were laid before the 1960s on

this railroad. This research will focus on the CWR that accounts for around 90 percent of total track miles.

Table 3. 4 Rail Laid Data Format

Prefix	Starting Milepost	Ending Milepost	Track Type	Rail Side	Rail Weight	Rail Gang	New /Relaid	Joint/ Weld



Notes: W = welded rail; J = jointed rail

Figure 3. 1 Distribution of Rail Laid Year

3.2.3 Tonnage Database

In the tonnage data file, gross tonnage, foreign gross tonnage, hazmat gross tonnage, net tonnage, hazmat net tonnage, tonnage on each axle, and number of gross cars that have passed on each segment are recorded monthly. Every segment in the tonnage data file is distinguished by prefix, track type, starting milepost, and ending milepost. This research

uses the gross tonnage and number of gross cars (Table 3.5). Traffic tonnage is recorded on segments with varying lengths with an average length of 1.76 miles.

Table 3. 5 Tonnage Data Format

Prefix	Starting Milepost	Ending Milepost	Track Type	Monthly Gross Tonnage	Monthly Number of Cars	Year	Month

3.2.4 Grade Database

Grade data describes the vertical slope of ground measured in percent. The grade is recorded by dividing the entire network into smaller segments with an average length of 0.33 miles. The grade data format is illustrated in Table 3.6. The average grade of the railroad network weighted by length is approximately zero. Figure 3.2 illustrates the histogram of the percent of the grade on the railroad network.

Table 3. 6 Grade Data Format

Prefix	Starting Milepost	Ending Milepost	Percent

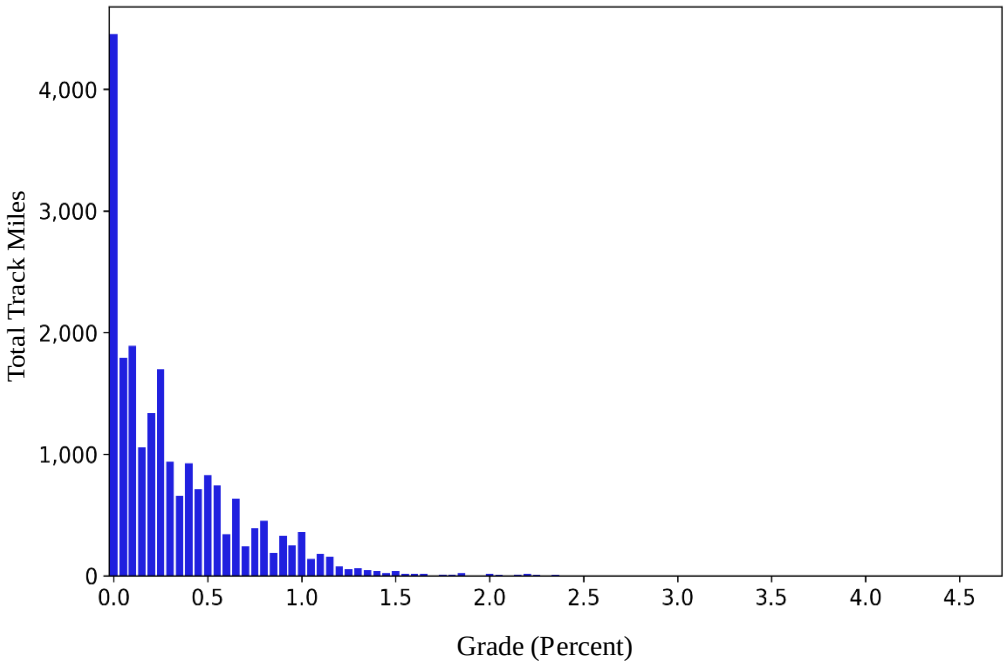


Figure 3. 2 Distribution of Grade (Percent)

3.2.5 Curvature Database

Curvature data includes the degree of curvature, length of curvature, direction of curvature, super elevation, offset, and spiral lengths. For the segments that are not included in this database, it is assumed that they are tangent tracks. There are approximately 5,800 curve-track miles (26% of the network track miles). The curve data format is illustrated in Table 3.7. Figure 3.3 shows the distribution of the curve degree on the railroad network.

Table 3. 7 Curvature Data Format

Prefix	Starting Milepost	Ending Milepost	Track Type	Curve Spiral	Curve Degrees	Curve Direction	Curve Super-elevation

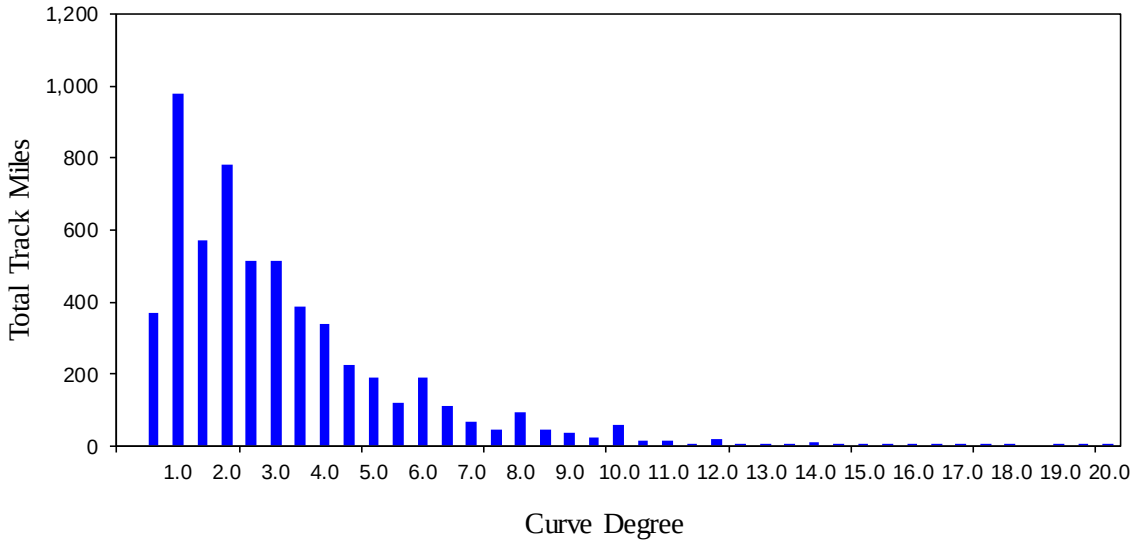


Figure 3. 3 Distribution of Data Curvature Degree (Curved Portion Only)

3.2.6 Track Chart

Track chart provides information on the track, including division, subdivision, track alignment, track profile, as well as maximum allowable train speed. The maximum freight speed on the network is 60 MPH. The weighted average speed on the network is 40 MPH. The distribution of the total segment length associated with speed categories is listed in Table 3.8.

Table 3. 8 Distribution of Speed Categories

Speed Category (MPH)	Total Track Miles	Percentage of Network
0 ~ 10	1,571.79	7.7%
10 ~ 25	4,237.83	20.7%
25 ~40	5,210.90	25.4%
40 ~60	9,482.31	46.2%

3.2.7 Turnout Database

Turnout data includes the turnout direction, turnout size and other information (Table 3.9). There are around 9,000 total turnouts in the network, with an average of 0.35 turnouts per track-mile. The distribution of track types with turnouts is illustrated in Table 3.10.

Table 3. 9 Turnout Data Format

Prefix	Milepost	Turnout Direction	Diverging Prefix	Turnout Size

Table 3. 10 Track Type Distribution of Turnouts

Track Type	Number of Turnouts
Single track	6,436
Multiple tracks	4,675

3.2.8 Signal Database

Signal data indicates whether a track is in a signalized territory (Table 3.11). There are approximately 14,500 track miles with signal, accounting for 67% of track miles of the railroad network.

Table 3. 11 Signal Data Format

Prefix	Starting Milepost	Ending Milepost	Signal Code[1]

[1] Descriptions of signal codes are presented in Appendix A.

3.2.9 Grinding Database

Rail grinding passes are used to remove surface defects and irregularities caused by rolling contact fatigue between wheels and the rail. In addition, rail grinding could reshape the rail-profile, resulting in better load distribution (Kalousek et al., 1989, Cuervo et al., 2015). The grinding passes for rails on the two sides of the track are recorded separately. Low rail passes and high rail passes are included in the dataset (Table 3.12). For tangent rail, the left rail is the low rail and the right rail is the high rail.

Table 3. 12 Grinding Data Format

Date	Subdivision	Prefix	Track Type	Starting Milepost	Ending Milepost	Low Rail Passes	High Rail Passes

Table 3. 13 Distribution of Grinding Frequencies with Respect to Year

Year	Grinding Frequency	Grinding-Rail-Miles	Total Grinding-Rail-Miles	Grinding Passes per Rail-Mile
2011	0	35,191	31,848.1	0.72
	1	12,935		
	2	3,475		
	2+	2,888		
2012	0	21,287	35,220.5	0.79
	1	16,297		
	2	4,216		

	2+	2,690		
2013	0	20,558	33,232.1	0.75
	1	19,949		
	2	2,348		
	2+	2,635		
2014	0	21,152	33,558.0	0.75
	1	16,354		
	2	5,008		
	2+	1,975		
2015	0	20,091	30,074.6	0.68
	1	21,085		
	2	1,755		
	2+	1,558		
2016	0	21,998	32,575.3	0.73
	1	15,438		
	2	5,245		
	2+	1,809		

3.2.10 Ballast Cleaning Database

Ballast cleaning replaces the “dirty” worn ballast with fresh ballast (Ashley, 2008). The locations of ballast cleaning are identified using prefix, track type, starting milepost and ending milepost (Table 3.14). The total mileage of ballast cleaning each year is listed in Table 3.15.

Table 3. 14 Ballast Cleaning Data Format

Year	Prefix	Track Type	Starting Milepost	Ending Milepost	Pass Miles

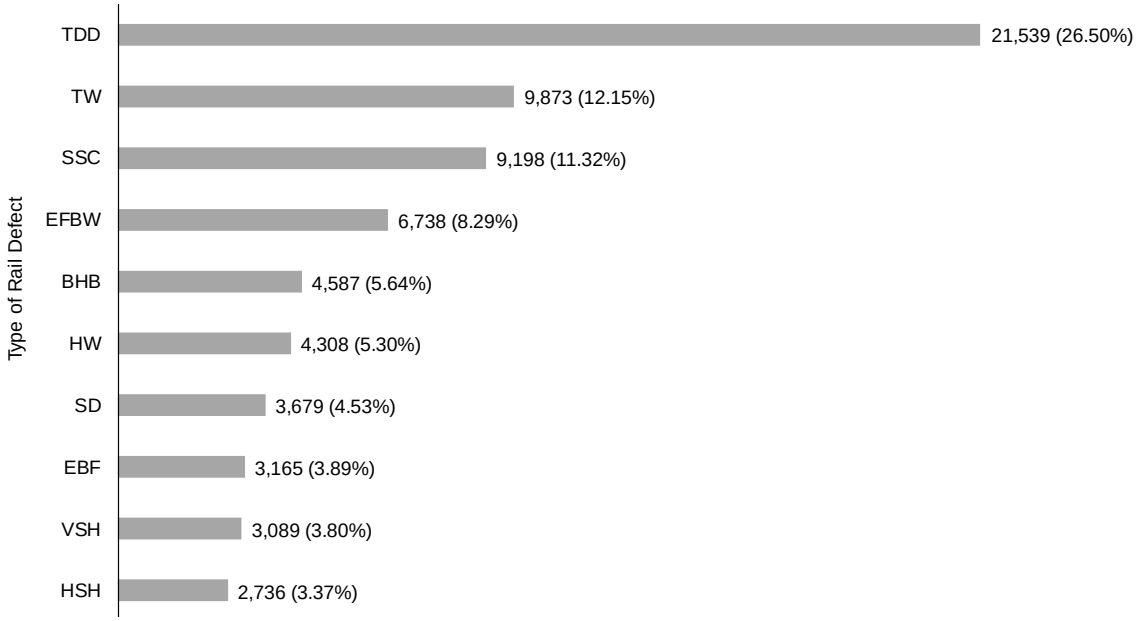
Table 3. 15 Total Track Mileage of Ballast Cleaning by Year

Year	Ballast Cleaning Frequency	Ballasted Track Miles	Total Track Miles with Ballast Cleaning
2011	1	900	1,149
	1+	116	
2012	1	1,609	1,864
	1+	122	
2013	1	1,335	1,763
	1+	193	
2014	1	1,735	2,393
	1+	285	
2015	1	1,862	2,299
	1+	213	
2016	1	932	1,166
	1+	99	

3.2.11 Rail Defect Database

Various types of rail defects during periods from 2011 to 2016 are recorded in the rail defect database. There are 25 different types of defects recorded. A necessary remediation action can be performed based on the type and severity of the detected defect. There are 31 different action types recorded in the database. In this research, the numbers

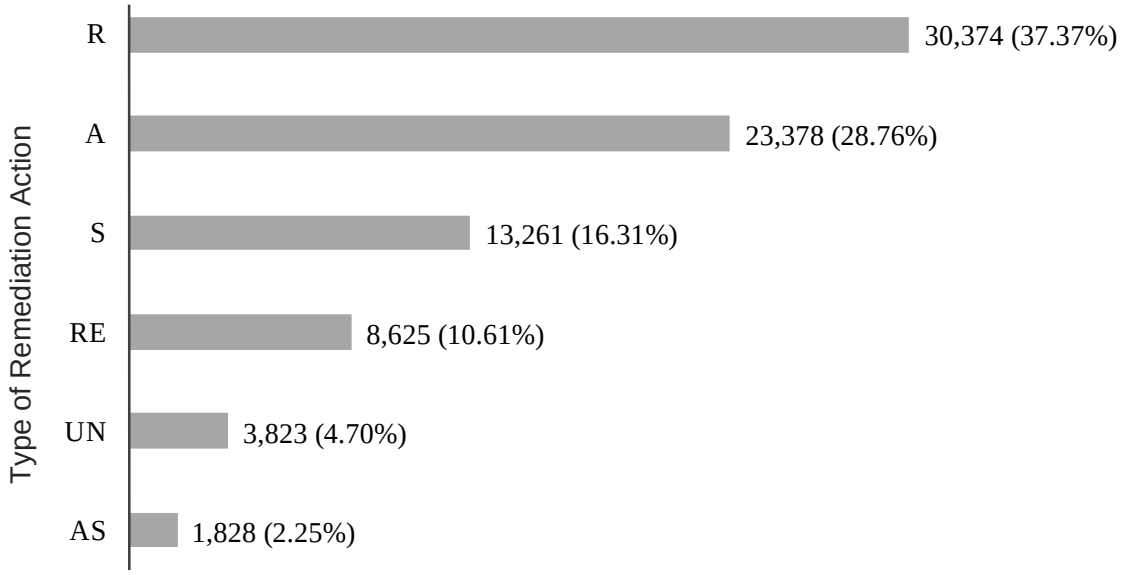
of each type of rail defects will be considered as input variables for predicting broken rail occurrence. The top 10 defect types account for around 85 percent of total defects (Figure 3.4). Figure 3.5 shows the distribution of remediation actions to treat defects.



Notes: TDD: detail fracture; TW: defective field weld; SSC: shelling/spalling/corrugation; EFBW: in-track electric flash butt weld; BHB: bolt hole crack; HW: head web; SD: shelly spots; EBF: engine burn fracture; VSH: vertical split head; HSH: horizontal split head.

Figure 3. 4 Top 10 Defect Types During 2011 and 2016[2]

[2] All the types of rail defects are listed in Appendix A.

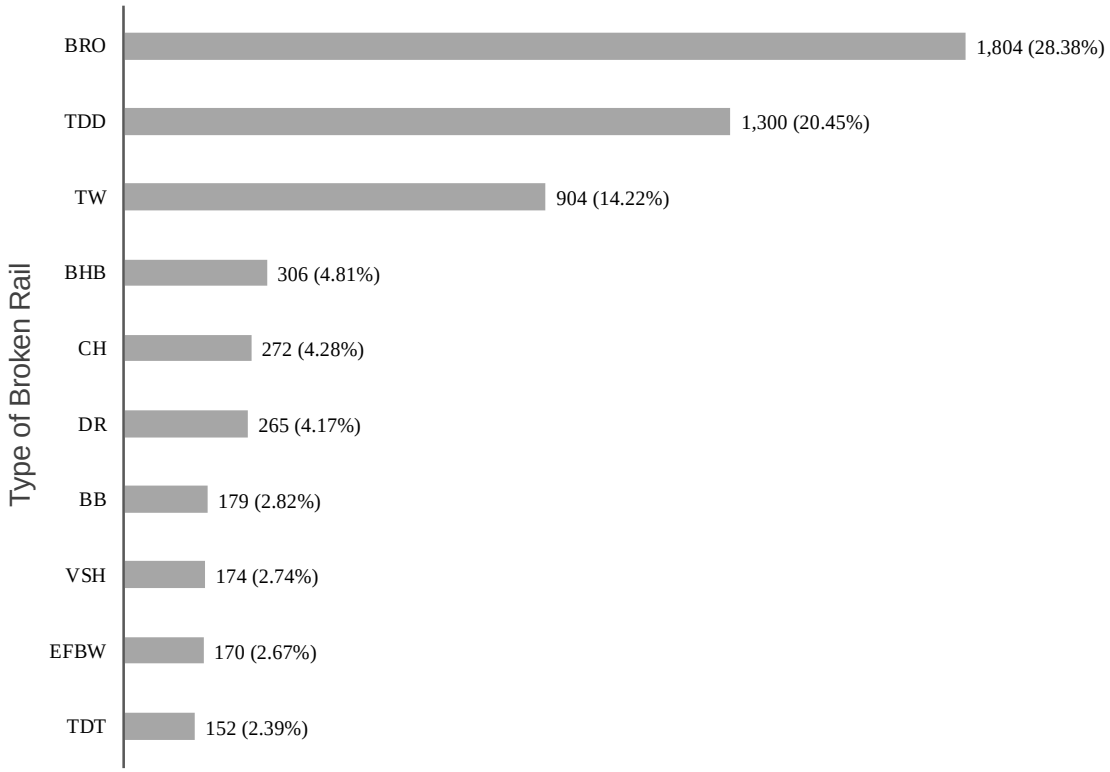


Notes: R: replace or remove rail section; A: apply joint/repair bars; S: slow down speed, RE: visually inspect or supervise movement; UN: unknown; AS: apply new speed.

Figure 3. 5 Distribution of Six Types of Remediation Action from 2011 to 2016

3.2.12 Broken Rail Database

The broken rail database contains 6,356 broken rails during the period from 2011 to 2016. Of the top 10 types of broken rails that account for around 87 percent of total broken rails, the distribution of each type is shown in Figure 3.6. The broken rail resulting from defect type BRO (broken rail outside joint bar limits) is dominant, which accounts for 28.3% of the total broken rails.



Notes: BRO: broken rail outside joint bar limits; TDD: detail fracture; TW: defective field weld; BHB: bolt hole crack; CH: crushed head; DR: damaged rail; BB: broken base; VSH: vertical split head; EFBW: in-track electric flash butt weld; TDT: transverse fissure

Figure 3. 6 Top 10 Types of Broken Rails from 2011 to 2016[3]

3.2.13 Track Geometry Exception Database

Track geometry was measured periodically and was corrected by taking necessary maintenance or repair actions if needed. There are 31 types of track geometry exceptions (track geometry defects) in the database provided by the railroad. Eight subgroups of track geometry exceptions, in which similar exception types are combined, are developed. The distribution of eight subgroups is listed in Figure 3.7.

[3] The pictures of top ten types of broken rails are listed in Appendix A.

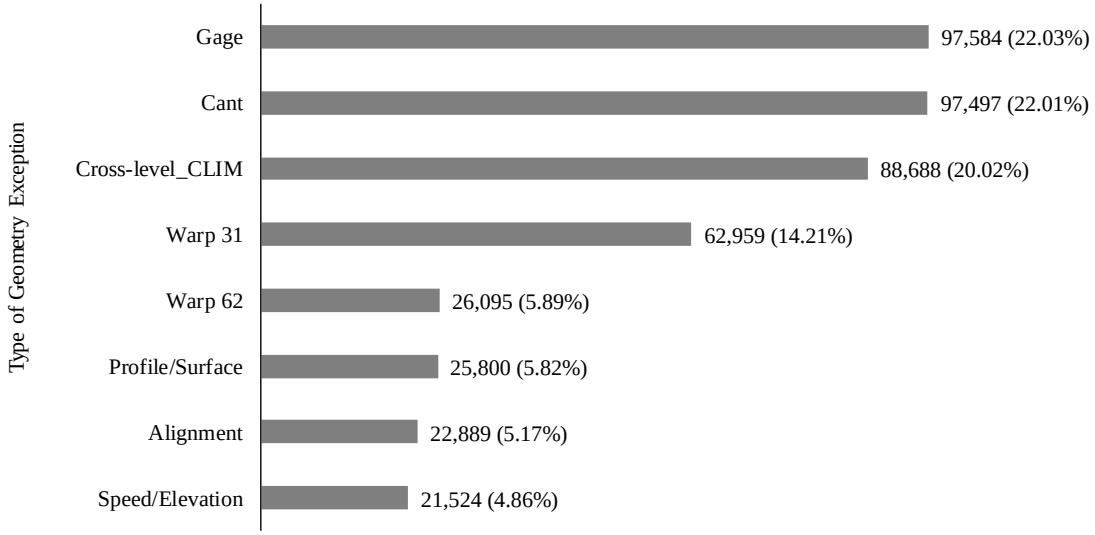


Figure 3. 7 Track Geometry Exception by Type from 2011 to 2016[4]

3.2.14 Vehicle-Track Interaction (VTI) Exception Database

The Vehicle Track Interaction (VTI) System is used to measure car body acceleration, truck frame accelerations, and axle accelerations, which can assist in early identification of vehicle dynamics that might lead to rapid degradation of track and equipment. When vehicle dynamics are beyond a threshold limit, necessary inspections and repairs are implemented. The VTI exception data contains the information about exception mileposts, GPS coordinates, speed, date, exception type, and follow-up actions for the period from 2012 to 2016. There are eight VTI exception types, and the distribution of each type is listed in Figure 3.8.

[4] The details of track geometry exceptions are described in Appendix A.

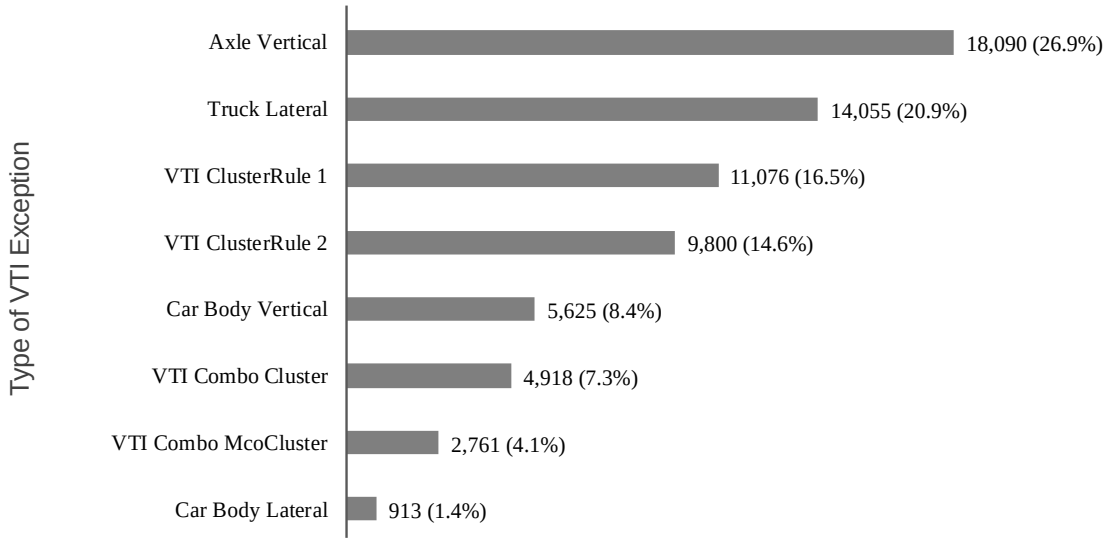


Figure 3. 8 Distribution of VTI Exception Types from 2012 to 2016

3.3 Data Preprocessing and Cleaning

The following sections describe the processing and cleaning of raw data, in order to build an integrated central database for developing and validating machine learning models.

3.3.1 Unify Data Column Names

To start, this research unified the formats of the column names and value types of corresponding columns in each database, especially for the location-related data columns.

- Prefix: an up-to-3-letter coding system working as route identifiers.
- Track Type: differentiate between single track and multiple tracks.
- Start MP: Starting milepost of one segment, if available.
- End MP: Ending milepost of one segment, if available.
- Milepost: If available, used to identify points on the track.
- Side: Including right side (R) and left side (L) to distinguish different sides of the track.

3.3.2 Detection of Data Duplicate

One of the common issues in data analysis is the duplicated data record. There are two common types of data duplications: (a) two data records (each row in the data file represents a data record) are exactly the same and (b) more than one record is associated with the same observation, but the values in the rows are not identical, which is so-called partial duplication. To determine the duplicates, selecting the unique key is the first step for handling duplicate records. Selection of unique key varies with the databases. For the databases which are time-independent (meaning that this information is not time-stamped), such as curve degree and signal, a set of location information is used to determine the duplicates. For the databases which are time-dependent, such as the rail defect database and service failure database, time information can be used to determine the duplicates. Meanwhile, using the set of location information alone is likely to be not sufficient to identify data duplicates because of possible recurrence of rail defects or service failures at the same location. Figures 3.9 to 3.12 show some examples of data duplicates in certain databases. In the following tables, the actual location information is masked to preserve privacy.

Prefix	MilePost	TrackType	Curve_Degrees	Curve_Elevation	Curve_Direction	Offset	Spiral_1	Curve_Length_PARTIAL	Spiral_2
ABC	143.6	SG	10.17	2.5	L	2597	310	220	130
ABC	143.6	SG	7	2	L	NaN	NaN	80	130

Figure 3. 9 Example of Partial Duplication in Curve Degree Database

Prefix	Start MP	End MP	Signal_Code
ABC	801.5	801.51	YL-S
ABC	801.5	801.51	YL-S

Figure 3. 10 Example of Exact Duplication in Signal Database

Prefix	Start MP	End MP	Signal_Code	Signal
ABC	323.6	323.61	CP	1
ABC	323.6	323.61	YL	0
ABC	323.61	323.62	CP	1
ABC	323.61	323.62	YL	0

Figure 3. 11 Example of Partial Duplication in Signal Database

Prefix	TrackType	Start MP	End MP	Side	Defect_Types	Date_Found	Defect_Size
ABC	SG	175.2	175.21	L	SDZ	7/26/2013	20
ABC	SG	175.2	175.21	L	SDZ	7/26/2013	20

Figure 3. 12 Example of Exact Duplication in Rail Defect Database

Different strategies for handling data duplications are listed below. Table 3.16 shows the selection of unique keys and proper strategies for databases. For the databases which are not listed in Table 3.16, it has been verified that no duplicates exist.

- Record Elimination: For exact duplications, there are two options for removing duplicates. One is dropping all duplicates and the other is to drop one of the duplicates.
- Worst Case Scenario Selection: For a partial duplication, select the worst-case-scenario value. For instance, over the junction of two consecutive curves, it is possible that two different curve degrees were recorded. In this case, assign the maximum curve degree to the junction (the connection point of two different curves).

Table 3. 16 Strategies for Duplication

Database	Unique Key to Identify Data Duplicate	Deduplication Strategy
Curve	Prefix, track type, milepost, side	Greater curve degree
Signal	Prefix, milepost, signal code	Drop either one

Rail Defect	Prefix, track type, milepost, side, defect type, date found, defect size	Drop either one
Service Failure	Prefix, track type, milepost, side, date found, failure type	Drop either one

3.3.3 Information Combination for Right Rail and Left Rail

Some databases differentiate between the left and right rail of the same track. For instance, the rail defect database can specify the side of the track where the rail defect occurred. Also, the rail laid database can specify the rail laid date for each side of the rail. However, some other databases cannot differentiate track sides, such as track geometry exception database and turnout database. This research combines the data from two sides of a track. It is possible that two sides of the track have different characteristics. When combining the information from the two sides of the track, there are five possible values for each attribute. They are “Select either one”, “Sum”, “Mean”, “Minimum”, and “Maximum”. The principle of selecting preferred value for the track is to set the track at the “worse condition”. For example, in terms of rail age, when combining right rail and left rail, the older rail age between right rail and left is selected, while for rail weight, the smaller rail weight is selected. This approach assigns more conservative attribute data to each segment. The details are listed in the Table B.1 in Appendix B.

3.4 Data Integration

To develop the comprehensive database, all of the collected data from all sources except geographical information system (GIS) data would be trackable using a reference database (which is the track file used in this research). The reference database contains the

location information (route identifier or prefix, starting milepost, ending milepost, and track type) without information on any features affecting broken rail occurrence. The information from each database which would be mapped into the comprehensive database is listed in Table 3.17. Figure 3.13 also presents the multi-source data fusion process.

Table 3. 17 Information from Each Database Involved in the Integrated Database (Partial List)

Database	Information
Service Failure	Failure found date, failure type, curvature or tangent, curve degree, rail weight, freight speed, annual traffic density, remediation action, remediation date
Rail Defect	Defect found date, defect type, remediation action
Geometry Exception	Geometry defect type, geometry defect date, track class reduced due to geometry exception, geometry exception priority, exception remediation action
VTI Exception	VTI type, VTI occurrence date, VTI priority, VTI critical
Tonnage	gross tonnage, number of car passes,
Grinding	grinding passes, grinding location
Ballast Cleaning	Ballast cleaning date, ballast cleaning location
Rail Laid	Rail weight, rail laid year, rail quality (new rail or re-laid rail), joint rail or continuous welded rail
Track chart	Maximum allowed freight speed
Curve Degree	Curve degree, super-elevation, curve direction, offset, spiral
Grade	Grade (percent)

Turnout	Turnout direction, turnout size
Signal	Signal code

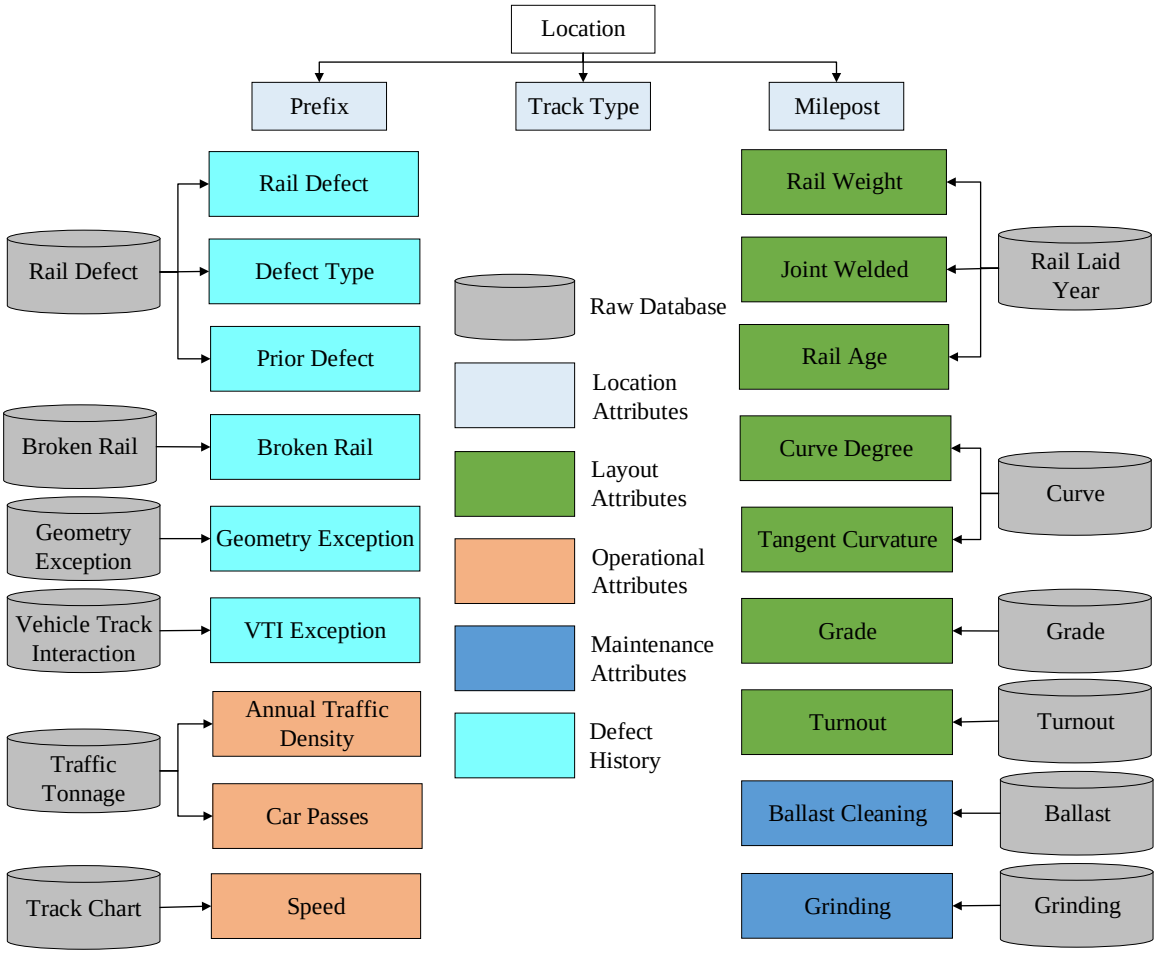


Figure 3. 13 Data Integration

The minimum segment length available for most of the collected databases was 0.1 mile (528 ft). There are over 206,000 track segments, each 0.1 mile in length, representing an over 20,600 track-mile network. All supplementary attributes from other databases would be mapped into the reference database based on the location index (Figure 3.14). This process is known as data integration. The location index contains information including prefix, track type, start MP, and End MP. In the reference database, each

supplementary feature for one location represents information series which cover the period from 2011 to 2016.

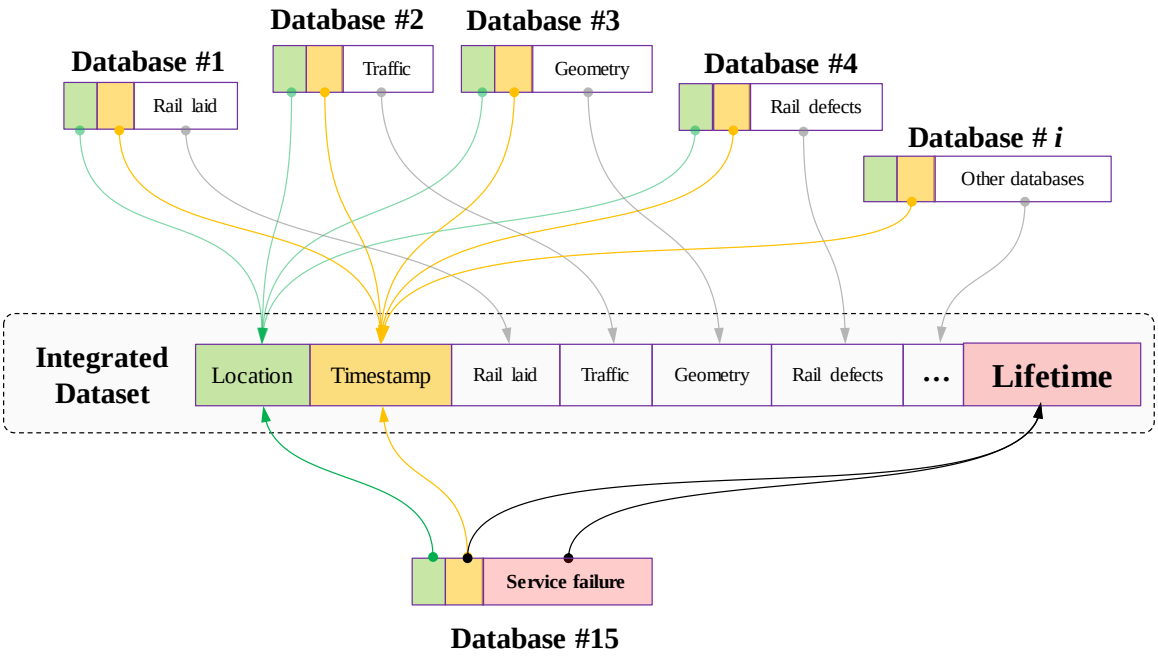


Figure 3. 14 Data Mapping to Reference Location

3.4.1 Handling Information Contradiction

Contradiction is a conflict between two or more different non-null values that are all used to describe the same property of the same entity. Contradiction is caused by different sources providing different values for the same attribute of the same entity. For example, tonnage data and rail defect data both provided the traffic information and may have different tonnage values for the same location. Data conflicts, in the form of contradictions, can be resolved by selecting the preference source based on the data source that is assumed to be more “reliable”. For example, both the curvature database and service failure database contain location-specific curvature degree information. If there is information conflict on the degree of curvature, the information from the curvature

database is used based on the assumption that this is a more “reliable” database for this data. The comprehensive database only retains the value of the preferred source. Table 3.18 shows the preferred data source for the attributes that have potential contradiction issues.

Table 3. 18 Preferred Database for Each Attribute

Attribute	Database Containing the Attribute	Preferred Database
Curve degree	Service failure, rail defect, VTI exception, curve degree	Curve degree
Rail weight	Service failure, rail defect, rail laid	Rail laid
Freight speed	Service failure, rail defect, track chart	Track chart
Annual traffic	Service failure, rail defect, monthly tonnage	Monthly Tonnage

3.4.2 Handling Missing Values

Handling missing data is one important task when overlaying information from different data sources to a reference dataset. Different solutions are available depending on the cause of the data missing. For example, one reason for missing data in the integrated database is that there was no occurrence of events at the specific location, for instance, grinding, rail defect, and service failures, etc. Fill the blank cells for this type of missing data with zeros because they represent no observations of events of interest. The other reason for missing data is that there is a missing value in the source data. For this type of missing data, Select the preferred value to fill it. Take the speed information in the integrated dataset as an example. Approximately 0.1 percent of the track network has missing speed information. The track segments with missing speed information would be

filled with the mean speed of the whole railway network. Table 3.19 lists the preferred values for the missing values of each attribute.

Table 3. 19 Preferred Values of Missing Information

Preferred Value	Attribute
Mean value	Rail laid year, speed, grade, rail weight, monthly tonnage, number of car passes, grinding, ballast cleaning
Zero	Curve degree, curve elevation, spiral, turnout, turnout size, rail defect, service failure, track geometry exception, VTI exception, measure of VTI exception
Worse case	Signal, rail quality (new rail versus re-laid rail)

3.4.3 Feature Construction

In the integrated database, two types of attributes (single-value attribute and stream attribute) were mapped (Figure 3.15). A single-value attribute is defined as a time-independent attribute, such as rail laid year, curve degree, grade, etc. A stream attribute (aka time series data) is defined as a set of the time-dependent data during a period. For most stream attributes, the period covers from 2011 to 2016, except for the attribute of vehicle-track interaction exception, which covers from 2012 to 2016. Twenty timestamps are defined with a unique time interval of three months from January 1st, 2012 to extract shorter-period data streams. In order to achieve that, a time window would be introduced. A time window is the period between a start and end time. A set of data would be extracted through the time window moving across continuous streaming data.

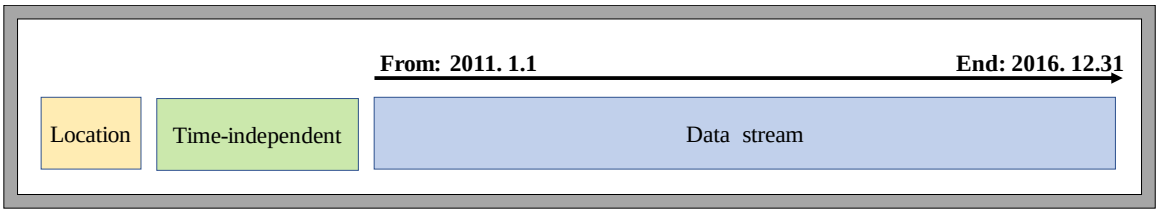
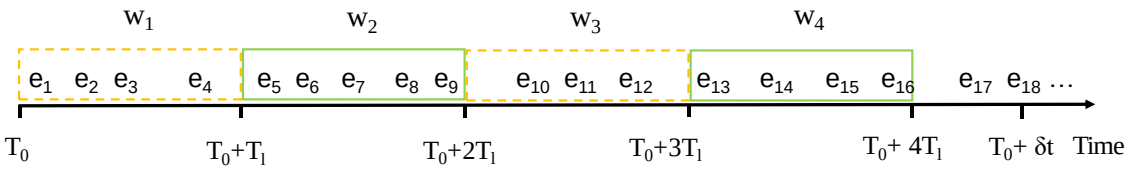


Figure 3. 15 Structure of the Integrated Database

Tumbling windows are one common type of time windows, which move across continuous streaming data, splitting the larger data stream into finite sets of small data streams. Finite windows are helpful for the aggregation of a data stream into one attribute with a single value. In this research, tumbling window is applied to split the data stream into finite sets.

In a tumbling window (Figure 3.16), events are grouped in a single window based on time of occurrence. An event is within only one window. A time-based tumbling window has a length of T_l. The first window (w_1) contains events that arrive at the time T_0 and $T_0 + T_l$. The second window (w_2) contains events (e.g., occurrence of rail defects in this research) that arrived between the time $T_0 + T_l$ and $T_0 + 2T_l$. The length of the tumbling

window is T_l and none of the windows overlap and each tumbling window represents a distinct time segment.



Note: e_i represents the i-th event

Figure 3. 16 Example of Tumbling Window

In this dissertation, the tumbling window is used to split the larger stream data into sets of small data streams (Figure 3.17 and Figure 3.18). The length of the tumbling window is set as half a year. Two features are extracted by two consecutive tumbling windows as shown in Figure 3.17 and Figure 3.18. Three timestamps are assigned to location "Loc1" (Figure 3.17). For the three timestamps, the time-independent features are unchanged for "Loc1". Taking rail defect as an example, the counts of rail defects are grouped by the tumbling window. For timestamp "2013.1.1", two tumbling windows are generated: Window 1 from 2012.7.1 to 2012.12.31 and Window 2 from 2012.1.1 to 2012.6.30. One feature about rail defect is the total number of rail defects that occurred in Window 1, which is from 2012.7.1 to 2012.12.31, and is denoted as "Defect_fh". Another feature about rail defect is the count number of rail defects that occurred in Window 2, which is from 2012.1.1 to 2012.6.30, and is denoted as "Defect_sh". If there was service failure which occurred after timestamp 2013.1.1, the lifetime is calculated by the days between the timestamp and the date of the nearest (in terms of time of occurrence) service failure. The event index is set to 1, which represents that service failure was observed after the timestamp. If there was no service failure after timestamp 2013.1.1 (Figure 3.18), the

lifetime would be calculated by the days between the timestamp and the end time of information stream “2016.12.31”. The event index is set to 0 which represents that service failure was not observed after that specified timestamp.

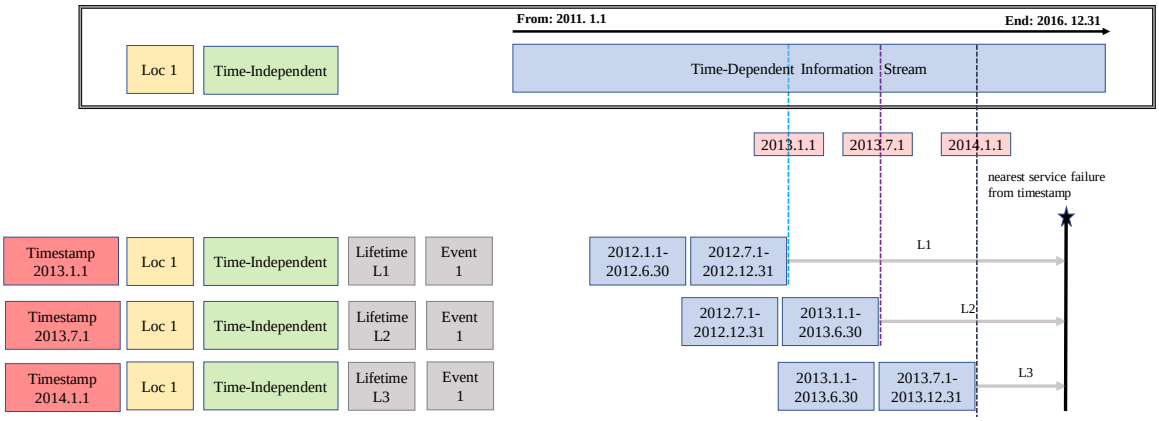


Figure 3. 17 Feature Construction with Nearest Service Failure in the Study Period

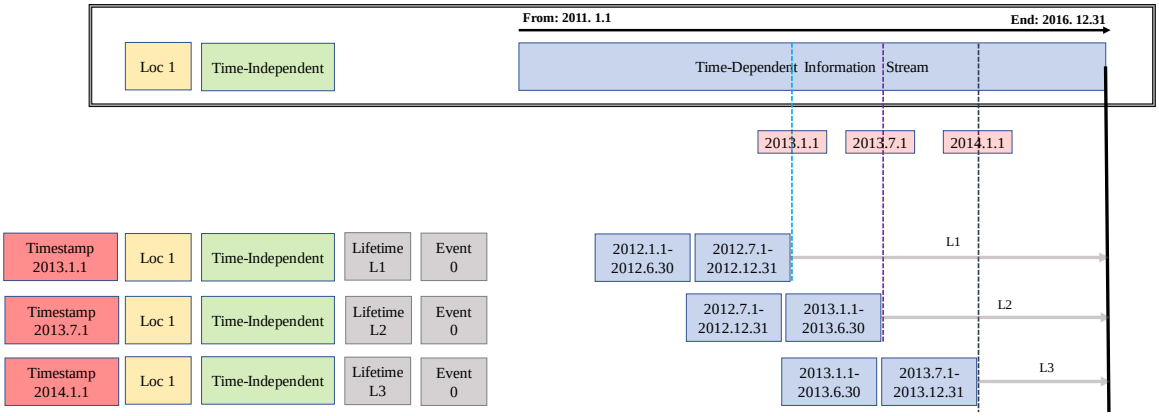


Figure 3. 18 Feature Construction without Nearest Service Failure in the Study Period

3.5 Exploratory Data Analysis

Exploratory data analysis (EDA) is conducted to develop a preliminary understanding of the relationship between most of the variables and broken rail rate, which is defined as the number of broken rails normalized by some metric of traffic exposure.

Because many other variables are correlated with traffic tonnage, broken rail frequency is normalized by ton-miles in order to isolate the effect of non-tonnage-related factors. The result of exploratory data analysis is summarized in Table 3.20. The following sections will detail the EDA finding for each variable based on the data in more recent years in the database (i.e. 2013 to 2016).

Table 3. 20 Summary of Exploratory Data Analysis Results

Variable	Relationship with Broken Rail Rate (per Billion Ton-Miles)
Rail age (years)	Broken rail rate first increases and then decreases with increasing rail age. The turning point for rail age is at 40 years. The non-monotonous relationship may be due to the correlation between rail age and annual traffic density.
Rail weight (lbs/yard)	Broken rail rate decreases monotonously with increased rail weight.
Curve degree	A higher rate is associated with a higher curve degree.
Grade (percent)	Broken rail rate increases with increasing grade.
Maximum allowed speed (MPH)	Higher broken rail rate is associated with lower maximum allowable speed on track.
Rail quality	Re-laid rail has a higher broken rail rate than new rail.
Traffic density (MGT)	A higher broken rail rate is associated with a lower annual traffic density.
Prior track geometry exceptions	Broken rail rate increases in the presence of prior track geometry exception defects.

Prior VTI exceptions	Broken rail rate increases in the presence of prior VTI exceptions.
Grinding	Broken rail risk initially decreases and then increases with increasing grinding passes. The turning point is at one rail grinding pass per year. The non-monotonous relationship may be due to the different purposes associated with different grinding operations.
Ballast cleaning	Broken rail rate decreases with ballast cleaning.

Rail Age

The broken rail rates are calculated for each category of the rail age (Table 3.21). The rates are determined by dividing the total number of broken rails that has occurred in a certain category of rail age by the total ton-miles in that category. With increasing rail ages, the broken rail rate per billion ton-miles first increased and then decreased. The turning point of the rail age is at 40 years. In other words, rail aged around 40 years (e.g., 30-39 years, 40-49 years) has the greatest number of broken rails per billion ton-miles. The possible reason is that rail age might have correlations with other variables, for example traffic tonnage and maintenance operations, which bring a compound effect together with rail age on broken rail rate. Figure 3.19 shows the correlation between rail age and annual traffic tonnage, in which annual traffic tonnage is decreasing as rail age increases. Based on this assumption, the numbers of broken rails per mile are calculated for categories of the products of annual traffic densities (MGT) and rail ages (years). It is shown in Table 3.22 that the increasing product of annual traffic tonnage and rail age increases the number of broken rails per track-mile.

Table 3. 21 Broken Rail Rate (per Billion Ton-Miles) by Rail Age,

All Tracks on Mainlines, 2013 to 2016

Rail age (years)	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
1-9	515	380.50	1.35
10-19	591	333.06	1.77
20-29	555	250.90	2.21
30-39	940	355.36	2.65
40-49	533	203.22	2.62
50-59	128	52.50	2.44
60 and above	16	8.84	1.81

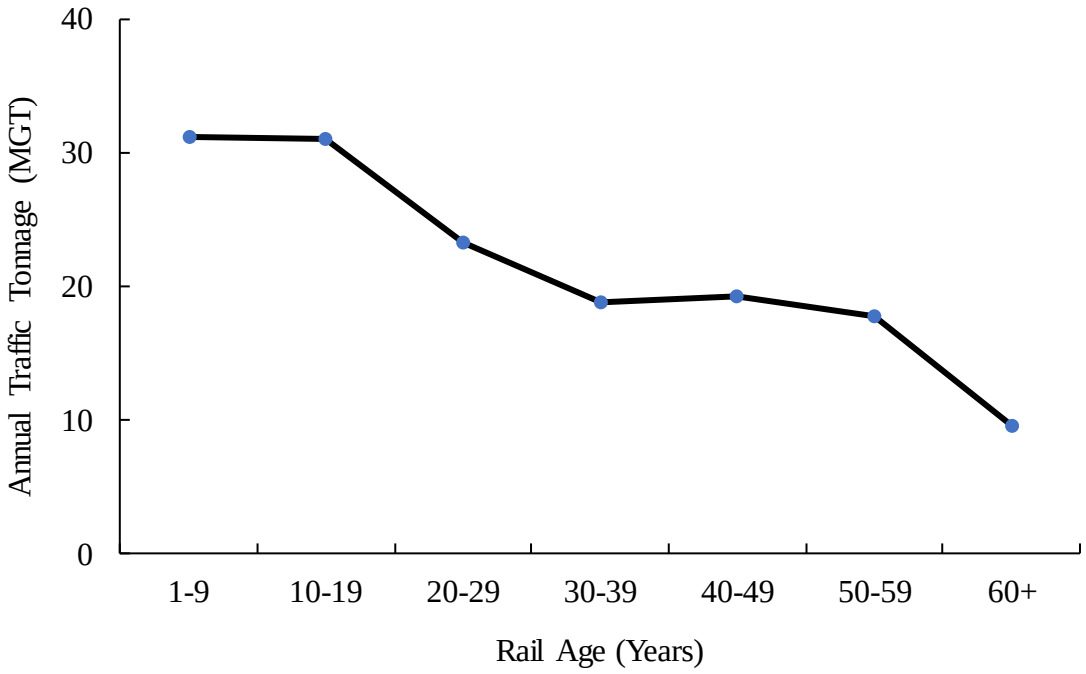


Figure 3. 19 Mean Annual Traffic Tonnage (MGT) in Terms of Rail Age Categories

Table 3. 22 Broken Rail Rate (per Track-Mile) by Product of Annual Traffic Tonnage and Rail Age, All Tracks on Mainlines, 2013 to 2016

Traffic × age (MGT× years)	Number of broken rails	Track-miles	Number of broken rails per track-mile
0-300	825	30,956	0.027
300-600	598	13,655	0.044
600-900	579	9,418	0.061
900-1,200	510	6,319	0.081
1,200-1,500	295	3,647	0.081
1,500 and above	470	4,501	0.104

Rail Weight

The broken rail rate in terms of the rail weight is presented in Table 3.23. It shows that, all else being equal, a heavier rail with a larger rail weight is associated with a lower broken rail rate, measured by number of broken rails per billion ton-miles. This conclusion is consistent with previous studies (e.g., Dick et al., 2002, 2003). Stress in rail is related to rail section and weight. Smaller, lighter rail sections experience more stress under a given load and may be more likely to experience broken rails.

Table 3. 23 Broken Rail Rate (per Billion Ton-Miles) by Rail Weight, All Tracks on Mainlines, 2013 to 2016

Rail weight (lbs/yard)	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
115 and below	288	72.57	3.97

115-122	452	156.83	2.88
122-132	1,022	384.29	2.66
132-136	1,490	830.20	1.79
136 and above	356	235.24	1.51

Curve Degree

Curvature increases rail wear and causes additional shelling and defects that might increase the probability of broken rails (Shyr and Ben-Akiva, 1996). Table 3.24 presents broken rail rates by curve degrees. Tangent tracks had around 70 percent of broken rails, but the broken rail rate per billion ton-miles is smaller than that on curvature tracks due to much larger traffic exposure (in terms of ton-miles) on tangent tracks. In terms of tracks with curves, the sharper curves involve higher broken rail rates. Previous research also shows that curved rails are more likely to have rail failures, all other conditions being equal (An et al., 2017; Dick, 2001).

Table 3. 24 Broken Rail Rate (per Billion Ton-Miles) by Curve Degree, All Tracks on Mainlines, 2013 to 2016

Curve degree	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
Tangent	2,501	1,217.87	2.05
0-4	837	372.45	2.25
4-8	222	78.56	2.83
8 or more	48	10.25	4.68

Grade

The effect of grade is illustrated in Table 3.25, in which the broken rail rate for each grade category (0 - 0.5 percent 0.5-1.0 percent, and over 1.0 percent) is presented. It indicates that increasing grade has slightly greater broken rail rate. The highest broken rail rate is on the tracks with the steepest slope (over grade percent 1.0). Steep grade might increase longitudinal stress due to the amount of tractive effort and braking forces, thereby increase broken rail probability (Dick, 2001; An et al., 2017).

Table 3. 25 Broken Rail Rate (per Billion Ton-Miles) by Grade,

All Tracks on Mainlines, 2013 to 2016

Grade (percent)	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
0-0.5	2,778	1,296.31	2.14
0.5-1.0	668	309.35	2.16
1.0 +	162	73.47	2.21

Rail Grinding

Rail grinding can remove defects and surface irregularities from the head of the rail, which lowers the probability of broken rails due to fractures originating in rail head. As mentioned previously, there are preventive grinding and corrective grinding. Preventive grinding is normally applied periodically to remove surface irregularities, and corrective grinding with multiple passes each time is usually performed to remove serious surface defects.

As Table 3.26 shows, broken rail rate without any grinding (0 grinding pass) is higher than that with preventive grinding passes (this research presumes one grinding pass indicates preventative grinding). This may imply that preventive grinding can reduce broken rail probability compared with the case of no grinding. However, the broken rail rate associated with more than one grinding pass is higher than that associated with just one grinding pass. The multiple grinding passes, which might be scheduled as corrective grinding passes, are associated with higher broken rail rates. This is analogous to the "chicken-and-egg" problem. There are more defects, and therefore corrective grinding is used. Because there is no identification of the type of grinding (preventive versus corrective) in the database, the assumption and observation mentioned above need further scrutiny.

Table 3. 26 Broken Rail Rate (per Billion Ton-Miles) by Grinding Passes, All Tracks on Mainlines, 2013 to 2016

Grinding passes per year	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
0	835	294.32	2.84
1	1,836	998.06	1.84
2+	937	386.74	2.42

Ballast Cleaning

Ballast cleaning aims to replace small worn ballasts with new ballasts. Table 3.27 shows that the broken rail rate without ballast cleaning is slightly higher than that with ballast cleaning. This potentially illustrates that proper ballast cleaning can improve

drainage and track support, which would thus reduce the probability of broken rails (Kumar 2006a, 2006b).

Table 3. 27 Broken Rail Rate (per Billion Ton-Miles) by Ballast Cleaning, All Tracks on Mainlines, 2013 to 2016

Ballast cleaning	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
No	3,151	1,454.47	2.17
Yes	457	224.67	2.03

Maximum Allowable Track Speed

To study the relationship between track speed and broken rail rate, broken rail rates were calculated for each category of track speeds (Table 3.28). The distribution indicates that broken rails on Class 4 or above (speed above 40 mph) account for over half of the total number of broken rails but the broken rail rate (number of broken rails per billion ton-miles), is the lowest. Instead, the highest broken rate is associated with maximum track speed from 0 to 25 mph that corresponds to FRA track Class 1 or Class 2. The possible reason is that maximum allowable track speed is also correlated to other track characteristics, engineering and maintenance standards. Higher track class, associated with higher rail quality, would bear higher usage (higher traffic density), which requires more frequent maintenance operations accordingly.

Table 3. 28 Broken Rail Rate (per Billion Ton-Miles) by Track Speed, All Tracks on Mainlines, 2013 to 2016

Track speed (MPH)	FRA track class	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
0-25	Class 1 & 2	430	132.48	3.25
25-40	Class 3	1,075	348.92	3.08
40-60	Class 4	2,103	1,197.73	1.76

Rail Quality

The broken rail rates with respect to rail quality (new rail versus re-laid rail) are listed in Table 3.29. In terms of the number of broken rails, new rails involve four times of that for re-laid rails. However, after normalizing broken rail frequency by traffic exposure in ton-miles, the broken rail rate of re-laid track is higher than that of new rails.

Table 3. 29 Broken Rail Rate (per Billion Ton-Miles) by Rail Quality, All Tracks on Mainlines, 2013 to 2016

Rail quality	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
New rail	2,484	1,299.83	1.91
Re-laid rail	644	196.68	3.27

Annual Traffic Density

The annual traffic density is measured in gross million tonnages (MGT). Table 3.30 lists the broken rail rates by annual traffic density categories. There is an approximately

monotonic trend showing that a higher annual traffic density category is associated with a lower broken rail rate. Rail tracks with a higher traffic density category (> 20 MGT) have a smaller number of broken rails per billion ton-miles, which is around half of that on tracks with a lower traffic density (< 20 MGT). The possible reason is that annual traffic density was probably correlated with other factors, such as rail age or FRA track class. In Figure 3.19, it is shown that higher annual traffic density is associated with smaller rail age. A track with higher annual traffic density is more likely to have higher FRA track class and correspondingly have more or better track inspection and maintenance.

Table 3. 30 Broken Rail Rate (per Billion Ton-Miles) by Annual Traffic Density (MGT), All Tracks on Mainlines, 2013 to 2016

Annual traffic density (MGT)	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
0-20	947	276.42	3.43
20-60	2,153	1,100.65	1.96
60 and above	508	302.06	1.68

Track Geometry Exception

The distribution of broken rail rate by track geometry exception is in Table 3.31. Around 94 percent of broken rails occurred at locations which did not experience track geometry exceptions, which covered 98 percent of the traffic volume in ton-miles. In contrast, around 6 percent of broken rails occurred at locations that experienced track geometry exceptions, which accounted for only 2 percent of traffic volume in ton-miles. In other words, the broken rail rate at locations with track geometry exceptions is

approximately three times as high as that at locations without track geometry exceptions. A similar conclusion was also found in the literature (e.g., Zarembski et al., 2016).

Table 3. 31 Broken Rail Rate (per Billion Ton-Miles) by Presence of Track Geometry Exceptions, All Tracks on Mainlines, 2013 to 2016

Track geometry exception	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
No	3,403	1,644.92	2.07
Yes	205	34.21	5.99

Vehicle-Track Interaction Exception

Table 3.32 presents the number of broken rails, traffic exposures, and broken rail rate by vehicle-track interaction (VTI) exceptions and non VTI exceptions. Around 2.8 percent of broken rails occurred on tracks with at least one VTI exception, while these locations only had 0.3 percent of traffic volumes (in ton miles). The broken rail rate with occurrence of vehicle-track interaction exceptions is six times as that without occurrence of vehicle-track interaction exceptions. However, the results herein may be subject to statistical uncertainty given the much smaller sample size and traffic volume for the trackage with VTI exceptions.

Table 3. 32 Broken Rail Rate (per Billion Ton-Miles) by Presence of Vehicle-Track Interaction Exceptions, All Tracks on Mainlines, 2013 to 2016

VTI	Number of broken rails	Billion ton-miles	Number of broken rails per billion ton-miles
No	3,507	1,670.84	2.10
Yes	101	8.29	12.18

Correlation between Input Variables

Correlation between input variables is measured by correlation coefficient to measure how strong a linear relationship between two variables is. The correlation coefficient is determined by dividing the covariance by the product of the two variables' standard deviations.

$$\rho_{X_i X_j} = \frac{cov[X_i, X_j]}{\sigma_{X_i}\sigma_{X_j}} = \frac{E[(X_i - E[X_i])(X_j - E[X_j])]}{\sigma_{X_i}\sigma_{X_j}} \quad (3\text{-}1)$$

where

$\rho_{X_i X_j}$ = correlation coefficient

$cov[X_i, X_j]$ = Covariance of variables X_i and X_j

$E(X)$ = expected value (mean) of variable X

σ_{X_i} = standard deviation of X_i

σ_{X_j} = standard deviation of X_j

X_i, X_j = two measured values

In general, the value of the correlation coefficient can vary between -1 and 1. “-1” indicates a perfectly negative linear correlation. “1” indicates a perfectly positive linear correlation. 0 means that there is no linear correlation between the two variables. Figure 3.20 shows the correlation matrix between the variables. Below are the major findings observed from the correlation matrix:

- The most remarkable correlation occurs between maximum allowed track speed and annual traffic density. There is a positive relationship (correlation coefficient is 0.51) between these two variables, which means higher annual traffic density is associated with higher maximum allowed track speed.

- Annual traffic density is also correlated with rail quality (new rail versus re-laid rail). New rail is associated with higher annual traffic density (correlation coefficient is 0.46) while re-laid rail is associated with lower annual traffic density (correlation coefficient is -0.46).

- Curve degree has a negative correlation with the maximum allowed track speed (correlation coefficient is -0.35). This represents that the tracks with higher curve degrees are associated with lower maximum allowed track speeds.

- Rail age and annual traffic density have a negative correlation (correlation coefficient is -0.26), which means the older rail is associated with lower annual traffic density, which is also shown in Figure 3.19.

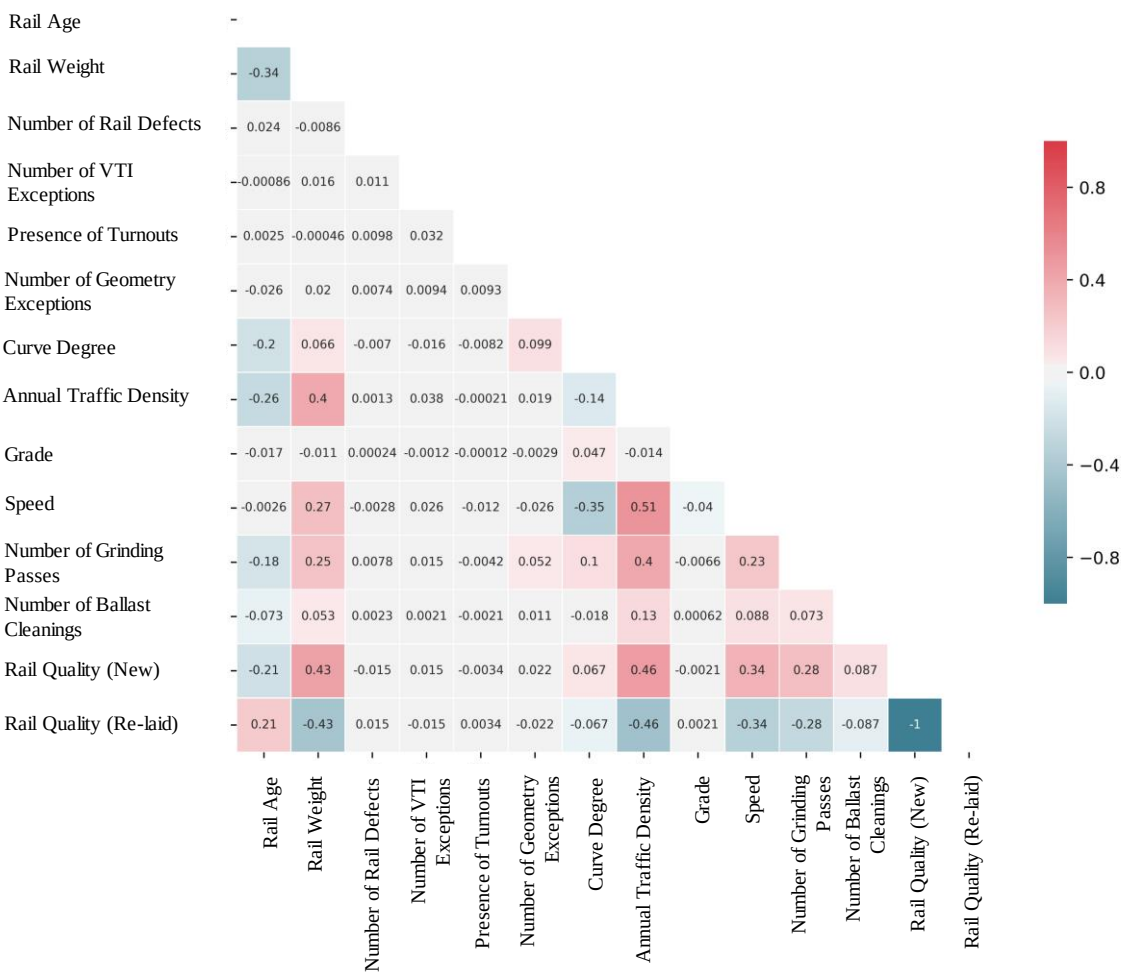


Figure 3. 20 Correlation between Each Two Input Variables

CHAPTER 4

TRACK SEGMENTATION

The track segmentation process is an important step for broken rail prediction. However, to our knowledge, little prior research is available to provide guidance for segmenting railway networks in order to predict potential broken rail locations.

4.1 Fixed-Length versus Feature-Based Segmentation

There are two types of strategies for the segmentation process: fixed-length segmentation and feature-based segmentation. fixed-length segmentation divides the whole network into segments with a fixed length. For feature-based segmentation, the whole network can be divided into segments with varying lengths. If fixed-length segmentation is applied and the small adjacent segments are combined, these combined segments may have different characteristics of certain influencing factors (e.g., traffic tonnage, rail weight) affecting broke rail occurrence. This combination would introduce potentially large variance into the database and further affect the prediction performance. For feature-based segmentation, segmentation features are used to measure the uniformity of adjacent segments. It is feasible to group and combine these adjacent segments under the condition that these adjacent segments embody similar features. Otherwise, these adjacent segments are isolated. Feature-based segmentation can reduce the variances in the new segments.

All features involved in the segmentation process can be divided into three categories: (1) track-layout-related features, (2) inspection-related features and (3) maintenance-related features, as illustrated in Table 4.1. The track-layout-related features are the essential information of rail and track, such as rail age, curve, grade, rail weight, traffic tonnages etc. The track-layout-related features would keep consistent on a relatively longer track milepost in general. The inspection-related features refer to the information obtained according to the measurement or inspection records, such as track geometry exceptions, detected rail defects, and VTI exceptions. These features may change frequently with time and locations. For example, the rail defect information can only be recorded when there is an inspection and the equipment or worker finds the defect(s). Also, it is possible the more inspections, the more defects might be found. This can lead to uncertainty for broken rail prediction. The maintenance-related features include grinding, ballast cleaning, rail repair or maintenance etc. Different types of maintenance actions may have different influences on rail integrity.

4.2 Track Segmentation Strategy

As mentioned above, there are two types of segmentation strategies: fixed-length segmentation and feature-based segmentation. Furthermore, there are two methods for feature-based segmentation: static-feature-based segmentation and dynamic-feature-based segmentation (Table 4.1). The details will be introduced as follows.

Table 4. 1 Track Segmentation Strategy

Segmentation strategies	Fixed-length segmentation	Feature-based segmentation	
		Static-feature-based segmentation	Dynamic-feature-based segmentation
Considered features	None	Track-layout-related features (including traffic tonnage)	Track-layout-related features, inspection-related features, maintenance-related features
Rules	The length of the newly emerged segment is fixed	If the difference between two adjacent 0.1-mile segments in feature values is above a given threshold, there will be two different new segments, otherwise, these two 0.1-mile segments are merged into one segment	The “best” segment length is found by minimizing a pre-defined loss function

During the segmentation process, the whole set of 0.1-mile network segments are divided into different groups. Each group should keep the uniformity on each segment. Aggregation functions are applied to assign the updated values to the new segment. The aggregation functions are given in Table 4.2. Specifically, the average values of nearby 0.1-mile segments for features such as the traffic density and speed are assigned to the new

segment, and the summation value is used for features such as rail defects, geometry defects and VTI exceptions.

Table 4. 2 Feature Aggregation Function in Segmentation Process (Partial List)

Features	Operation
Traffic density	Mean
Rail weight	Minimum
Rail age	Maximum
Rail defect	Sum
Broken rail number	Sum
Grinding passes	Mean
Ballast cleaning	Mean
Geometry defects	Sum
Speed	Mean
Curve in degree	Maximum
Grade in percent	Maximum
VTI exceptions	Sum

4.2.1 Fixed-Length Segmentation

The fixed-length segmentation is the segmentation strategy that uses the pre-defined fixed length to merge consecutive 0.1-mile segments compulsively, without accounting for the variance of the features on these segments. This forced segmentation strategy can be understood as a moving average filtering along the rail line. In the example shown in Figure 4.1, there are a total of fifteen (15) 0.1-mile segments. The values of two

features, rail age and annual traffic density, are described by two lines. In the fixed-length segmentation, a pre-determined fixed length is set to 0.3 miles. Therefore, three consecutive 0.1-mile segments are combined. For example, segment A-1 is composed of the original 0.1-mile segments 1 to 3. The rail ages of these three 0.1-mile segments are not identical, being 20, 20, and 24 years, respectively. The rail age assigned to the new segment A-1 should be the mean value of these three values (i.e. 21.3 years).

- **Advantages**: Fixed-length segmentation is the most straightforward (easiest) approach for track segmentation and the segmentation algorithm is the fastest.
- **Drawbacks**: The inherent difference of certain features on the same segment may be significant, which would affect the modeling accuracy.

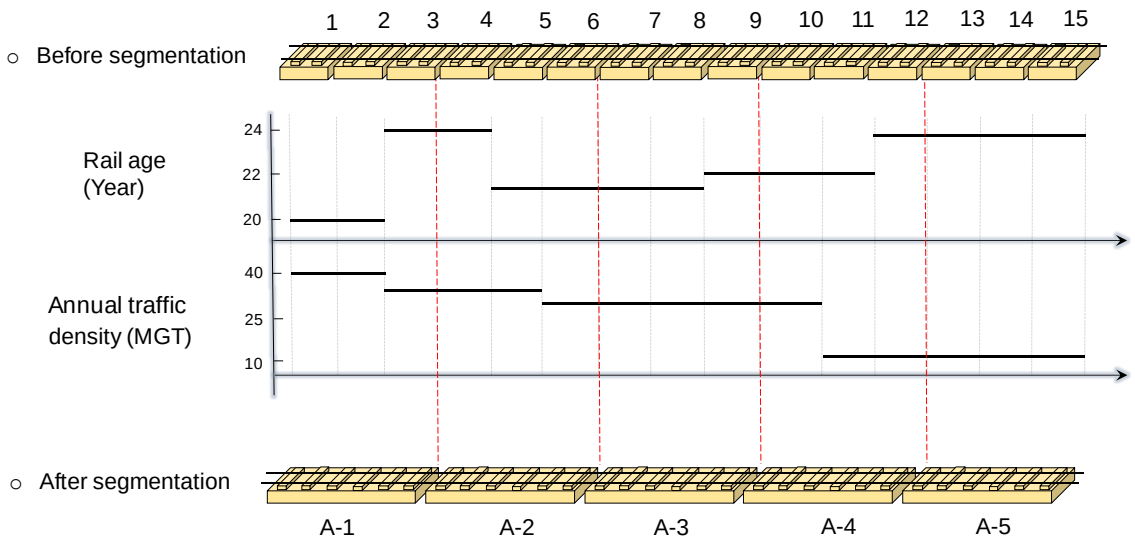


Figure 4. 1 Schematic Illustration of Fixed-Length Segmentation

4.2.2 Feature-Based Segmentation

Feature-based segmentation aims to combine uniform segments together. The uniformity is defined by the internal variance (i.e. variance among the 0.1-mile segments)

on the new segment. The uniformity is measured by the information loss, which is calculated by summing the weighted variances on involved features. The formula shown below is used to calculate the information loss. In feature-based segmentation, the information loss $\mathrm{Loss}(A)$ should be minimized (ideally zero) when determining the length of a newly merged segment.

$$\mathrm{Loss}(A) = \sum_{i\in[1,n]} w_i \cdot \mathrm{std}(A_i) \tag{4-1}$$

where

A: the feature matrix

n: number of involved features

A_i: the i^{th} column of *A*

w_i: the weight associated with the *i-th* feature

$\mathrm{std}(A_i)$: the standard deviation of the *i-th* column of *A*

The loss function can be interpreted as follows. Given multiple features, the information loss can be calculated by the weighted summation of the standard deviation of each feature. The information loss represents the internal variance of records of the involved features. The smaller the value of the loss functions, the "better" the segmentation result can be, due to minimizing the internal variances of selected features on the same segmentation. Ideally, within the same segment, selected feature values should be identical or close, while on different segments, at least one feature significantly differs.

Static-Feature-based Segmentation

In preparation for static-feature-based segmentation, segmentation features should be selected to determine the uniformity of the adjacent 0.1-mile segments. The static-

feature-based segmentation only uses the track-layout-related (static) features to measure the information loss when combining consecutive segments to a new longer segment. In static-feature-based segmentation, the information loss $\mathrm{Loss}(A)$ should be minimized to zero when determining the length of newly merged segment. The weights associated with each involved segmentation features are identical. Therefore, static-feature-based segmentation is an adaptive segmentation scheme in which a segment is assigned when at least one involved feature changes. Figure 4.2 shows an illustrative static-feature-based segmentation example. The selected segmentation features might be continuous or categorical. For categorical features, the uniformity is defined by whether the features among 0.1-mile segments are identical. For continuous features, a tolerance threshold should be used to define the uniformity. If the difference of continuous feature values of adjacent segments is smaller than the defined tolerance, the uniformity still exists. In this research, for feature-based segmentation, 10% of the standard deviation of differences of continuous features of the two consecutive 0.1-mile segments is used as the tolerance. In the example as shown in Figure 4.2, two features, rail age and annual traffic density, are both continuous variables. In order to simplify the illustration of the segmentation process, it is assumed that the differences of each value for each feature are beyond the tolerance. In the example, fifteen 0.1-mile segments are combined into seven new, longer segments. A new segment is assigned when any involved feature changes.

- **Advantages**: Static-feature-based segmentation is easy to understand, and the algorithm is easy to design. The internal difference of static rail feature information is also minimized.

- **Drawbacks**: When considering more features, the final merged segments can be more scattered with a large number of segmentations. The difference of features between the adjacent 0.1-mile segments, such as maintenance and defect inspection history, are difficult to be considered in static-feature-based segmentation because they are point-specialized events (non-static).

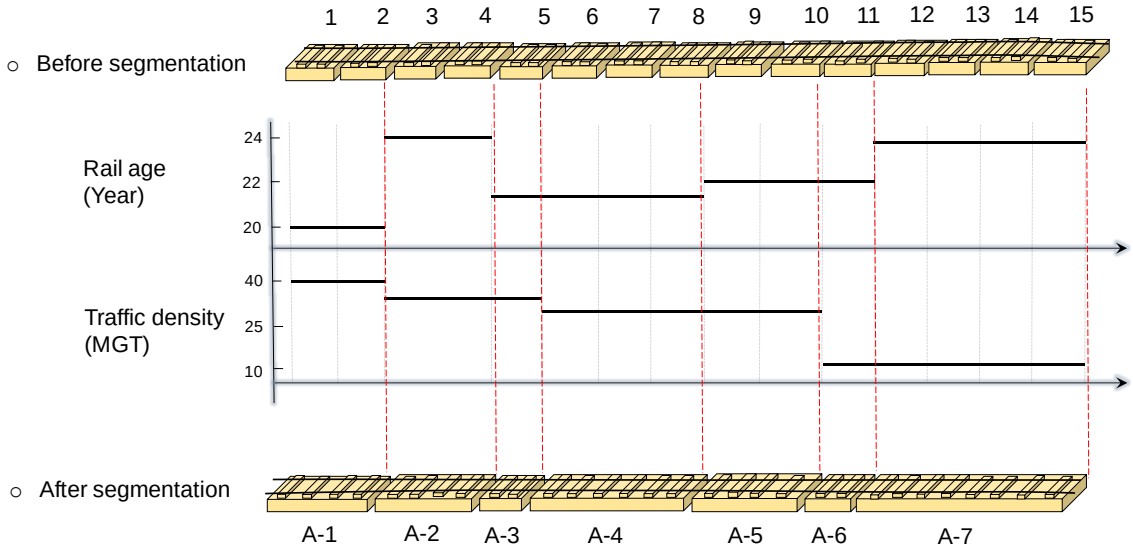


Figure 4. 2 Static-Feature-Based Segmentation

Dynamic-Feature-Based Segmentation

Differing from fixed-length segmentation and static-feature-based segmentation, dynamic-feature-based segmentation is the segmentation strategy that uses an optimization model to minimize a pre-defined loss function to find the "best" segment length around a local milepost. All features are used to calculate the information loss function to evaluate the internal difference of the merged segments. The optimization model is shown as

$$L = \underset{m}{\operatorname{argmin}}\, Loss(A^m) \tag{4-2}$$

$$\text{Loss}(A^m) = \sum_{i\in[1,n]} w_i \cdot \text{std}(A_i^m) \tag{4-3}$$

where

A^m: feature matrix with m rows (the number of 0.1-mile segments is m)

n: number of involved features

A_i^m: the i-*th* column of A^m (the i-*th* feature)

w_i: the weight associated with the *i-th* feature

$\text{std}(A_i^m)$: the standard deviation of the *i-th* column of A^m

With a fixed beginning milepost, find the best A^m that is associated with the minimum information loss. A^m indicates a feature matrix associated with m consecutive 0.1-mile segments. The optimization model can be interpreted as: finding the best segment length to minimize the loss function, from all possible segment combinations. One example is illustrated in Figure 4.3. To solve the optimization model, iteration algorithm is used to optimize the segmentation and get the approximately optimal solution. Besides, the loss function is also employed to find the best segment length. For the example shown in Figure 4.3, two features are involved for dynamic-feature-based segmentation, which are rail age and annual traffic density. The weights associated with the two features in the information loss function are assumed to be the same. To illustrate this type of segmentation, the minimum length of combined segment is set to 0.3 miles. It is shown that the minimum information loss is obtained at the original segment 8. Then the other segments are combined to develop another new segment.

- **Advantages**: Dynamic-feature-based segmentation takes all features (both time-independent or time-dependent) into consideration. The influence of the diversity of involved features can be controlled by changing the weights in the loss function. Dynamic-feature-based segmentation can also avoid the combined segments being too

short. Therefore, this type of segmentation strategy might be more appropriate for network-scale broken rail prediction.

- **Drawbacks**: The computation is time-consuming compared with fixed-length segmentation and static-feature-based segmentation. The development of algorithm is more complex.

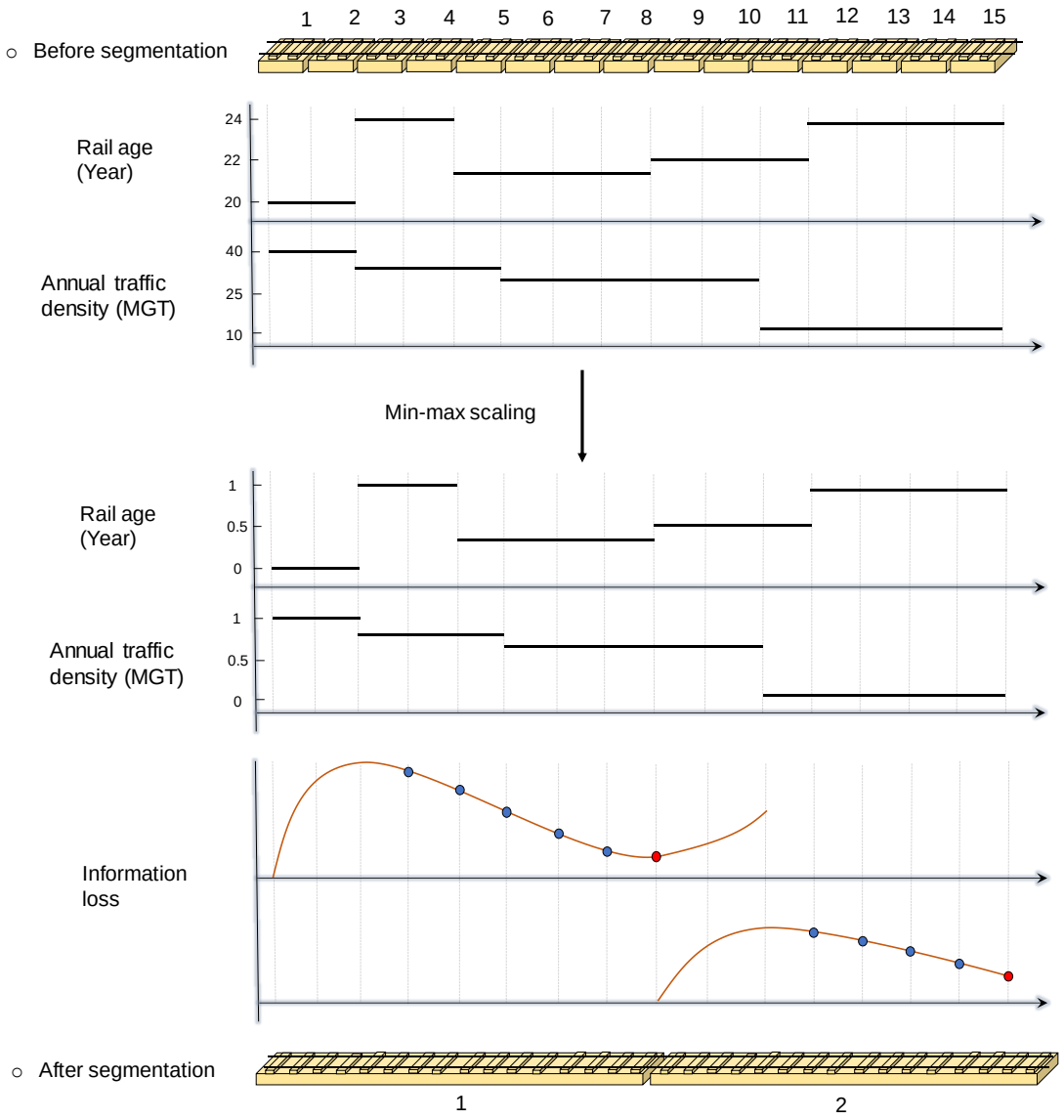


Figure 4. 3 Dynamic-Feature-Based Segmentation

4.3 Comparison of Track Segmentation Strategies

To compare the performance of different segmentation strategies, numerical experiments are conducted. In this section, the performance of three fixed-length segmentation setups, eight dynamic-feature-based segmentation setups, and one feature-based segmentation setup are compared. The area under the receiver operating characteristics (ROC) curve is used as the metric. ROC is a graph showing the performance of a classification model at all classification thresholds. The area under the curve (AUC) measures the entire two-dimensional area underneath the entire ROC curve. It is one of the most widely used evaluation metrics for checking any classification model's performance with two main advantages. Firstly, AUC is scale-invariant and measures how well predictions are ranked, rather than their absolute values. Secondly, it is classification-threshold-invariant and measures the quality of the model's predictions irrespective of what classification threshold is chosen. In general, the higher the AUC is, the better the model performs.

To compare the performance of different segmentation strategies, Naïve Bayes classifier is used as a reference model to evaluate the performance of a segmentation strategy. Naïve Bayes classifier is used to select the optimal segmentation strategy because of its fast computation speed. The segmented data selected by the Naïve Bayes method will later be applied in other machine learning algorithms.

The comparison result is shown in Table 4.3. U-0.2, U-0.5, and U-1.0 represent the fixed-length segmentation with fixed segment length of 0.2 miles, 0.5 miles, and 1.0 mile, respectively. For the dynamic-feature-based segmentation, D-1 to D-8 represent eight

alternative setups, in which varying feature weights in the loss function are assigned, respectively. In dynamic-feature-based segmentation, the involved features are categorized into four groups. Features in Group 1 are related to the number of car passes. Group 2 contains features which are associated with traffic density. Group 3 contains features which are related to the track layouts and rail characteristics, such as curve degree, rail age, rail weight etc. Features in Group 4 are associated with defect history and maintenance history, such as prior defect history and grinding passes. The feature weights assigned to each group in each dynamic-feature-based segmentation setups are in Table 4.4.

Table 4. 3 Comparison of Different Segmentation Strategies

	Fixed-Length Segmentation			Static-Feature-Based Segmentation	Dynamic-Feature-Based Segmentation							
	U-0.2	U-0.5	U-1.0		D-1	D-2	D-3	D-4	D-5	D-6	D-7	D-8
Average Segment Length	0.200	0.500	1.000	0.300	0.621	0.282	0.377	0.360	0.327	0.197	0.220	0.341
AUC	0.705	0.704	0.700	0.813	**0.832**	0.777	0.821	0.793	0.796	0.825	0.827	0.804

Table 4. 4 Feature Weights in Dynamic-Feature-Based Segmentation

	Group 1	Group 2	Group 3	Group 4
D-1	100	10	1	1
D-2	1	1	1	1
D-3	0	1	1	0
D-4	1	0	0	0
D-5	1	1	0	0
D-6	10	5	1	1
D-7	10	10	5	1
D-8	20	20	1	1

As shown in Table 4.3, the dynamic-feature-based segmentation with the D-1 setup performs the best using the AUC as the metric. For the D-1 setup, features about number of car passes have the largest weight. Features about track and rail characteristics as well as features about defect history and maintenance history have the least weights in the loss function. The new segmented dataset contains approximately 664,000 segments including twenty timestamps. There are 37,162 segments experiencing at least one broken rail from 2012 to 2016, accounting for about 5.6% of the whole dataset. By comparison, in the original 0.1-mile dataset, there are 47,221 segments (1.1%) with broken rails among 4,143,600 segments.

4.4 Chapter Summary

There are three segmentation strategies to segment the railroad network: fixed-length segmentation, static-feature-based segmentation, and dynamic-feature-based

segmentation. Fixed-length segmentation does not account for feature variance on the same segment. Static-feature-based segmentation only accounts for the track-layout-related features (time-independent or static features) and has two limitations. Firstly, when considering additional features for network delineation, the segments can be more scattered and shorter. Secondly, based on the exploratory data analysis, the presence of geometry exceptions and rail defects could increase the broken rail rate. However, in the static feature-based segmentation, defect-related features are not considered. Among these three segmentation strategies, the dynamic-feature-based segmentation scheme using a loss function is shown to have the highest performance (using AUC as the metric) using the Naïve Bayes algorithm. The possible reason is that dynamic-feature-based segmentation minimizes the feature variances of the merged segments. In the following chapter 5, a broken rail prediction model is proposed in which the input and output data is generated using dynamic-feature-based segmentation strategy.

CHAPTER 5

DEVELOPMENT AND VALIDATION OF BROKEN RAIL PREDICTION MODEL

The objective of this chapter is to predict broken rail probability using machine learning algorithms empowered by fast-growing railroad "big data", related to network-level track characteristics, maintenance activities, traffic and operations, as well as track and rail inspection histories. To develop an efficient, high-accuracy prediction model, a novel, customized Soft Tile Coding based Neural Network model (STC-NN) is proposed to predict the probability of broken rail occurrence by time and location. This chapter will introduce feature creation, feature transformation, feature selection, model development, and validation.

5.1 Nomenclatures, Variables, and Operators

Table 5. 1 Nomenclatures, Variables, and Operators

Terminology	Explanation
STC-NN	Soft-Tile-Coding-Based Neural Network
NN	Neural Network
MCP	Multi-Classification Problem
BCP	Binary Classification Problem

TPTR	Total Predictable Time Range, describing the upper time limit of the STC-NN model
FIR	Feeding Imbalance Ratio
IR	Imbalance Ratio
TPR	True Positive Rate
FPR	False Positive Rate
ROC	Receiver Operating Characteristics
AUC	Area Under ROC Curve

Variable	Denotation
t	A variable representing a timestamp
T	Lifetime from starting observation time to occurrence time of broken rails or ending observation time for a segment
m	The number of tiling for soft-tile-coding
n	The number of tiles in a tiling
d_j	The initial offset of the *j*-th tiling
ΔT	The length of the time range of each tile
$F(T\|m,n)$	Tile-encoded vector of a lifetime T with parameter m and n
$S(T\|m,n)$	Soft-tile-encoded vector of a lifetime T with parameter m and n
θ	The weights of a neural network
g	An input feature set of one rail segment
$p(g\|\theta)$	The output soft-tile-encoded vector of the STC-NN model with parameters θ, given input feature set g
G	$\{g_1, g_2, \dots, g_N\}$ is a batch of input feature set

T	$\{T_1, T_2, \dots, T_N\}$ is a batch of input lifetime corresponding to G
p_{ij}	The output probability of the *j*-th tile in the *i*-th tiling
$r_{ij}(T)$	The effective coverage ratio of the *j*-th tile in the *i*-th tiling
p_{ij}^*	The probability density of the *j*-th tile in the *i*-th tiling
$t_{ij}(T)$	$[\![[i\Delta T + d_j, (i+1)\Delta T + d_j) \cap [0, T]]\!]$ is the length of intersection between time range of the *j*-th tile in the *i*-th tiling and the range $t \in [0, T]$
$\mathcal{L}(g, T\|\theta, m, n)$	The loss function of STC-NN model
α	The learning rate of training algorithm of STC-NN model
T_0	A lifetime threshold used to cut out a lifetime into binary value
P_0	A probability threshold used to cut out a cumulative probability into binary value
$L_r(T_i\|T_0)$	The binary label generated from a lifetime, given T_0 as the threshold
$L_p(T\|P_0)$	The binary label generated from $P(t < T)$, given P_0 as the threshold

Operator	Denotation
$P(t < T)$	The cumulative probability of broken rail within $t \in [0, T)$
(a, b)	A mapping from vector a to vector b
$[a, b], [a, b), (a, b]$	A range from a to b
$\{\cdot\}$	A set with discrete elements
$[\![\cdot]\!]$	An operator to obtain the length of a set with continuous values

5.2 Feature Engineering

Feature engineering is an important process of model development which aims to improve model accuracy or efficiency. Among feature engineering in this dissertation, feature creation, feature transformation and feature selection are included. Feature creation focuses on deriving new features from the original features, while feature transformation is used to scale the range of features and normalize the length-related features (e.g., number of rail defects) by segment length. Feature selection identifies the set of features that accounts for most variances in the model output.

5.2.1 Feature Creation

Original Features

The original features in the integrated database include:

- Rail age (year), which is the number of years since the rail was first laid
- Rail weight (lbs/yard)
- New rail versus re-laid rail
- Curve degree
- Curve length (mile)
- Spiral (feet)
- Super-elevation (feet)
- Grade (percent)
- Maximum allowed operational speed (MPH)
- Signaled versus non-signaled
- Number of turnouts

- Number ballast cleaning
- Number of grinding passes
- Number of car passes
- Annual gross tonnages
- Number of broken rails
- Number of rail defects (by type)
- Number of track geometry exceptions (by type)
- Number of vehicle-track interaction exceptions (by type)

Cross-Term Features

In addition to considering the features individually, this research also considers the interaction items (i.e. creating cross-term features). In this study, cross-term features can be products, divisions, sums, or the differences between two or more features. In addition to finding the product of rail age and traffic tonnages, the products of rail age and curve degree, curve degree and traffic tonnage, rail age and track speed, and others are also created. The division between traffic tonnage and rail weight is calculated. In terms of the sums of some features, the aim is to combine sparse classes or sparse categories. Sparse classes (in categorical features) are those that have very few total observations, which might be problematic for certain machine learning algorithms, causing models to be overfitted. Taking rail defect types as an example, there are more than ten different types of rail defect recorded in the rail defect database. However, several rail defect types rarely occur, which belong to sparse classes. To avoid sparsity, this research groups similar classes together to form larger classes (with more observations). Finally, this research can group the remaining sparse classes into a single “other” class. There is no formal rule for

how many classes that each feature needs. The decision also depends on the size of the dataset and the total number of other features in the database. Later, for feature selection, all possible cross-term features originating from raw features in the database are calibrated, and the optimal combination of features are selected to improve the model performance. The creation of cross-term features is done based on the data structure and domain knowledge. The selection of cross-term features is conducted based on model performance.

5.2.2 Feature Transformation

Min-Max Normalization

The range of values of some features in the database varies widely. For instance, the value magnitudes for traffic tonnage and curve degree can be very different. For some machine learning algorithms, objective functions may not work properly without normalization. Feature normalization makes each feature contribute proportionately to the objective function. Moreover, feature normalization can speed up the convergences for gradient descent that are used in various machine algorithm trainings. Min-max normalization is calculated using the following formula:

$$x_{new} = \frac{x - min\,(x)}{max(x) - min\,(x)} \tag{5-1}$$

where x is an original value, and x_{new} is the normalized value for the feature.

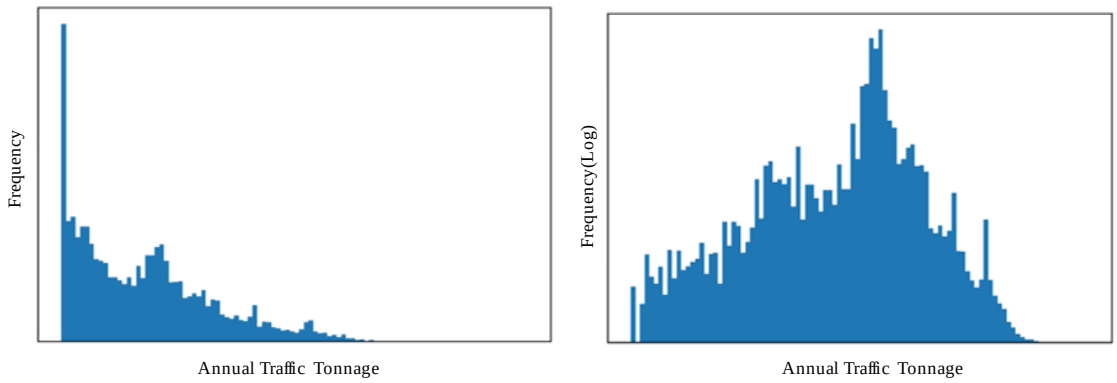
Categorization of Continuous Features

There are two types of features: categorical (e.g., signaled versus non-signaled) and continuous (e.g., traffic density). The categorization of all continuous features are also done,

besides retaining the original continuous type. For instance, track speed is in the range of 0 to 60 mph, which can be categorized in accordance with FRA track class, in the range of [0,10], [10,25], [25,40], [40-60], which designates track classes from 1 to 4, respectively.

Feature Distribution Transformation

The distributions of continuous features values are calculated. It is found that some features are distributed skewed towards one direction. Transformation functions are applied to transform the feature distribution into a normal distribution, in order to improve the performance of the prediction algorithm (Trawiński et al., 2012). For example, Figure 5.1 plots the distributions of traffic tonnages before and after feature transformation. The distribution of raw traffic tonnages is distributed skewed towards smaller values. However, traffic tonnages are distributed approximately normally after logarithmic transformation. Normalization of feature distribution through feature transformation could benefit the convergence efficiency of the model, especially for machine learning models.



(a) Before (b) After

Figure 5. 1 Distribution of Annual Traffic Tonnage Before and After Feature Transformation

Feature Scaling by Segment Length

After network segmentation based on input features, the segment lengths could vary. It is also possible that the values for some features over the segments are proportional to segment lengths. To avoid repeated consideration of the impact of segment length, feature scaling by segment length is applied to the related features, such as the total number of rail defects and track geometry exceptions over the segments. In this way, the density of some feature values by segment length is calculated. However, there are some segments with very small segment lengths. The density of the features for these short segments cannot represent the correct characteristics due to the randomness of occurrence. A length threshold of 1 mile is used for the feature scaling by segment length. No scaling is conducted if a segment is shorter than 1 mile.

5.2.3 Feature Selection

One key concern in the original model dataset is high-dimensionality which would increase the complexity and the overfitting probability of the model if all features are involved in the model. dimensionality reduction is a process of reducing the number of features under consideration by obtaining a set of principal variables. Principal Component Analysis (PCA) is the process to create new terms to obtain the principal variables. However, in the engineering field, the effect of original engineering variables is preferred. Therefore, feature selection is applied in the dissertation for dimensionality reduction.

Feature selection is the process in which a subset of features is automatically or manually selected from the set of original ones to optimize the model performance using defined criteria (Cai et al., 2018). With feature selection, features contributing the most to

the model performance are selected. Irrelevant features are discarded in the final model. Feature selection can also reduce the dimension of model feature matrix and speed up the model training. There are several approaches to feature selection, such as forward selection, backward selection, genetic algorithms and so on. In this dissertation, we developed a comprehensive algorithm which combines forward selection, backward selection and simulated annealing together. One of the most prevalent criteria for feature selection is the area under the operating characteristics curve (aka. AUC).

In this dissertation, a machine learning algorithm called LightGBM (Light Gradient Boosting Machine) is used for feature selection considering its fast-computational speed as well as acceptable model performance based on the AUC. The details about LightGBM can be found in (Ke et al., 2017). In feature selection, there are thousands of possible combinations of features. It is impossible to iterate all possible combinations of features to search for the optimal subset of features. A designed method is applied for feature selection which is described in detail in Figure 5.2. In this optimization-based feature selection method, the forward searching, backward searching and simulated annealing techniques are used in steps:

Step 1. In forward searching, select one feature each time to be added into the combination in order to maximally improve AUC, until the AUC is not improved further.

Step 2. Use backward searching to select one feature to be removed from the combination of features obtained from step 1, in order to maximally improve AUC, until AUC is not improved further.

Step 3. After step 2, make multiple loops between step 1 and step 2 until the AUC is not improved further.

Step 4. Because forward searching and backward searching select the features greedily, it is possible to result in a local optimal combination of features. The simulated annealing (SA) algorithm makes the local optima stand out amidst the combination of features. In this step, record the current combination of features with local optima and the corresponding AUC.

Step 5. First, create the cross-term features based on the combination of features obtained from step 4. After creating the cross-term features, repeat steps 1 to 4 until obtaining the optimal combination of current features. Due to the computational complexity of step 5, cross-term development is only conducted one time. This research does not consider the interaction between the new created cross-term features and other features. In the process, we can use an indicator N to represent whether creation of cross-term features has been conducted or not. If N is equal to “False”, then create cross-term features and repeat steps 1 to 4. If N is equal to “True”, then the optimal combination of features has been obtained and the process is complete.

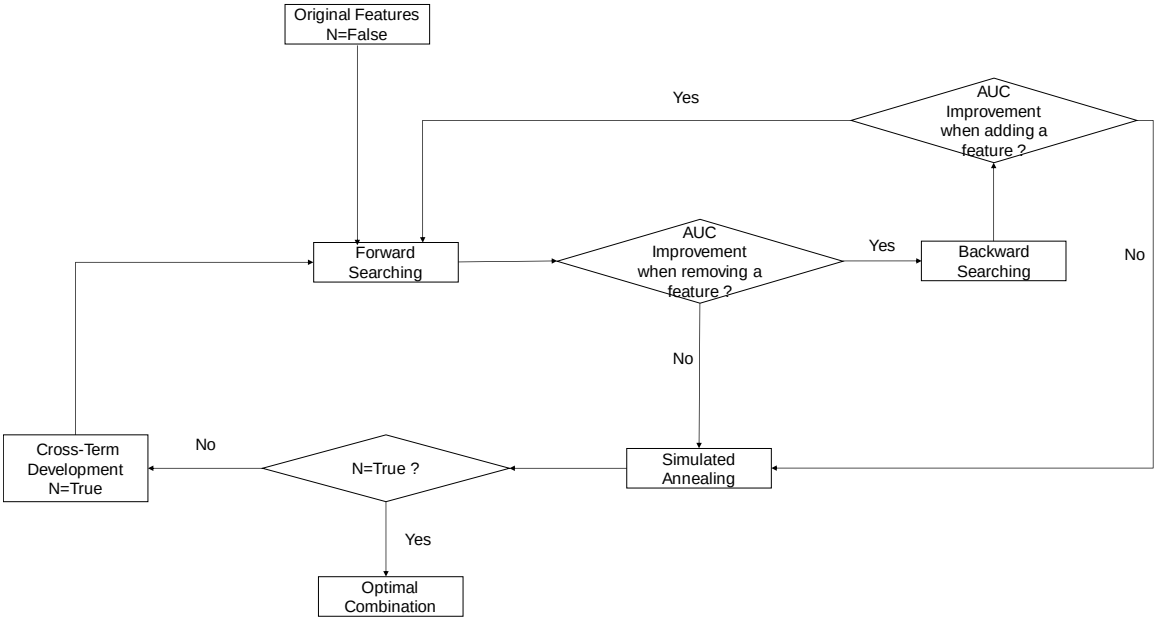


Figure 5. 2 Optimization-Based Feature Selection Process

Originally, the number of variables involved in the model (including dummy variables) is about 200. After feature selection, the top 10 variables are selected. Figure 5.3 lists the 10 features chosen from approximately 200 features.

- Segment Length: The length of the segment (mile)
- Traffic_Weight: The division between annual traffic density and rail weight (annual traffic density divided by rail weight)
- Car_Pass_fh: The number of car passes in the prior first half year
- Rail_Age: The year between the research year and the rail laid year
- Defect_hf: The number of detected defects in the prior first half year
- Curve Degrees: The curve degree
- Turnout: The presence of turnout
- Service_Failures_fh: The number of detected broken rails in the prior first half year

- Speed×Segment Length: The product of the maximum allowed operational speed and the segment length
- Age_Curve: The product of rail age and curve degree

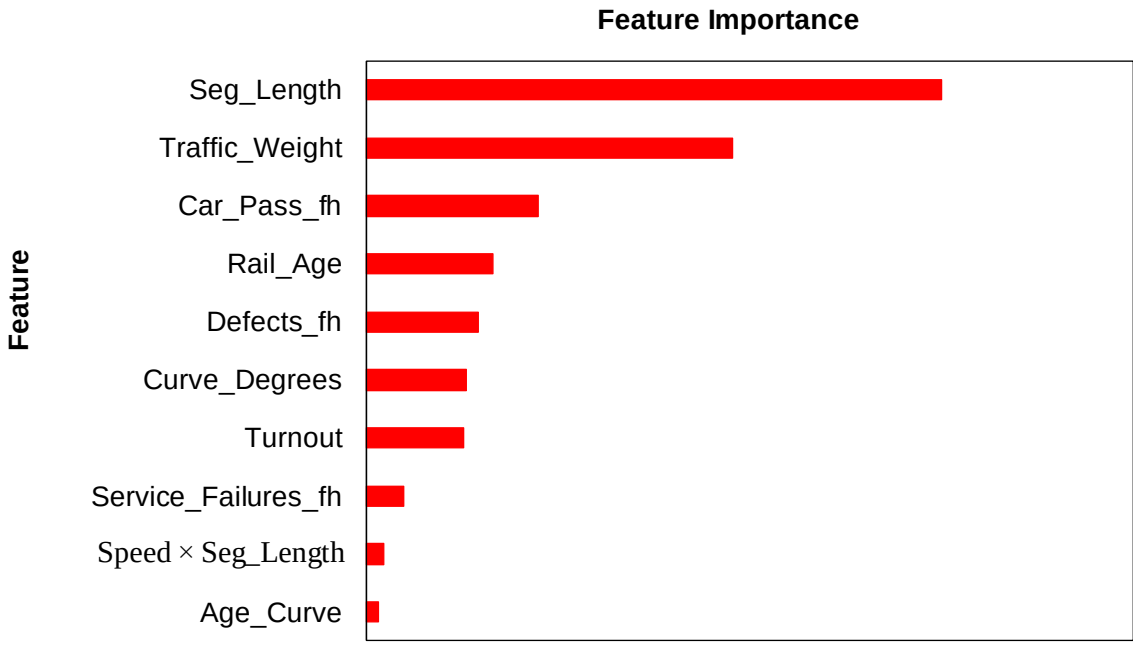


Figure 5. 3 Selected Top 10 Important Features using LightGBM Algorithm

In Figure 5.3, segment length shows the highest importance rate, and the ratio between annual traffic density and traffic weight is the second most important. Table 5.2 shows the impacts of the important features on the broken rail probability. A comparison of the distribution of the important features among different tracks was conducted. Two distributions of the important features are calculated, one for the top 100 track segments with the highest predicted broken rail probabilities, the other for the entire network of the studied railroad.

From Table 5.2, it is found that the top 100 track segments (with highest estimated broken rail probabilities) have larger average lengths. The distributions of traffic/weight

for the railway network and the top 100 track segments appear to be different, which reveals that track segments with larger traffic/weight are prone to having higher broken rail probabilities. The statistical distributions of the number of car passes and rail age also illustrate that higher broken rail probability is associated with higher rail age and more car passes on the track.

Table 5. 2 Selected Features on Top 100 Segments versus the Whole Network

	Segment Mileage		Traffic (MGT)/Rail Weight (lbs/yard)		Number of car passes		Rail Age (years)	
	Network	Top 100 Segments	Network	Top 100 Segments	Network	Top 100 Segments	Network	Top 100 Segments
Mean	0.20	3.24	0.16	0.32	247,435	465,958	25	36
25%	0.04	1.44	0.04	0.18	85,097	277,319	11	32
50%	0.10	2.62	0.14	0.32	225,740	474,450	25	38
75%	0.21	4.15	0.14	0.42	356,337	641,610	36	44

5.3 Overview of Soft-Tile-Coding-Based Neural Network

The relationship between contributing factors and broken rail risk has been investigated using a variety of models in previous studies, such as logistic regression, survival analysis (e.g., Weibull model, Cox model), and the Markov stochastic model etc. As stated in the literature review section, there are advantages and limitations for each model. To address the challenges of predicting broken rail occurrence by location and time, a Soft-Tile-Coding-Based Neural Network (STC-NN) is proposed in this dissertation. As illustrated in Figure 5.4, the model framework contains five parts: (a) Dataset preparation; (b) Input features; (c) Encoder: soft-tile-coding of outcome labels; (d) Model architecture; and (e) Decoder: probability transformation.

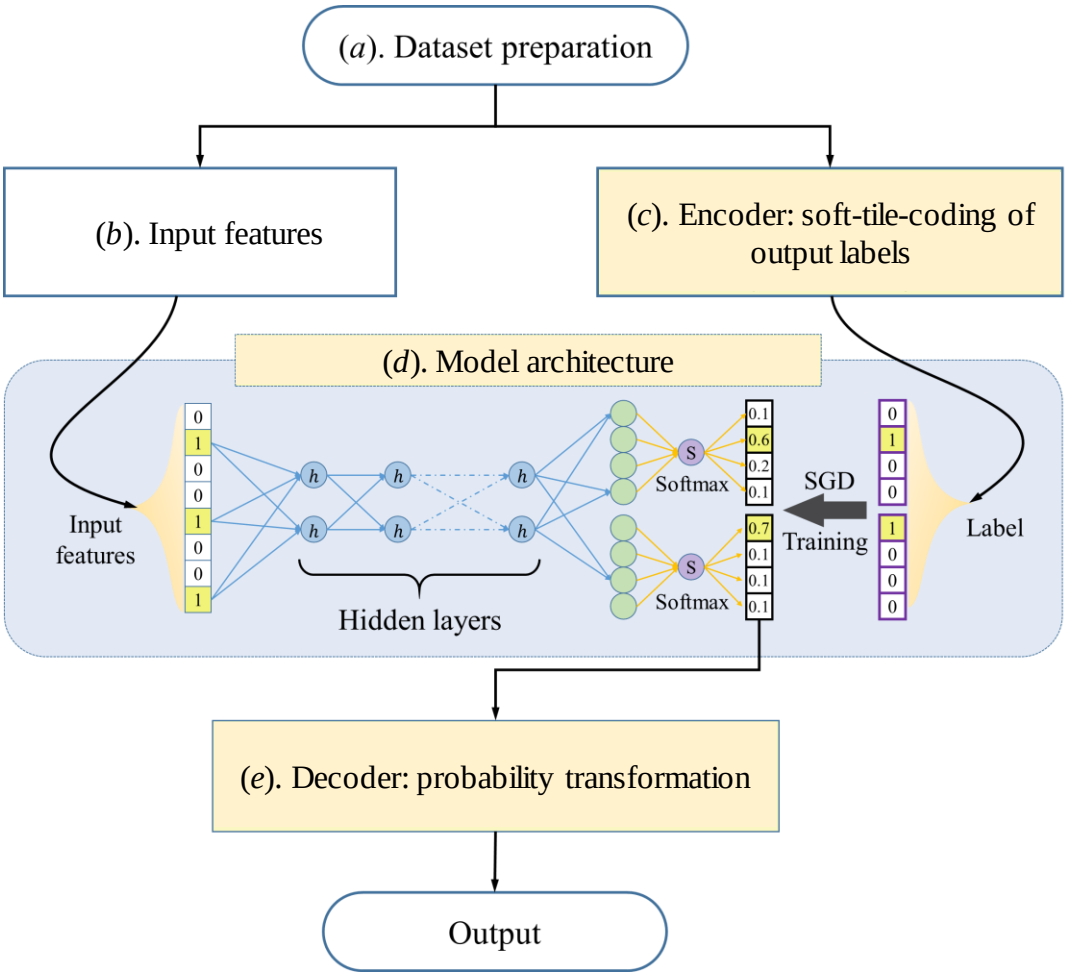


Figure 5. 4 Schematic Illustration of STC-NN Algorithm Framework

In part (a), dataset preparation, an integrated dataset should be developed which contains input features and outcome variables. The outcome variables are continuous lifetimes, which may have a large range. In survival analysis research, the lifetime could be the exact lifetime or censored lifetime. The exact lifetime is defined as the duration time from the starting observation time to the occurrence time of the event of interest (the found date of broken rail in this research), while censored lifetime is the duration from the starting time to the ending observation time if no event occurs. Input features might be categorical or continuous variables. For categorical features, one-hot encoding is applied to transform categorical features into a binary vector, in which only one element is 1 and the summation

of the vector is equal to 1 (Buckman et al., 2018). To improve computational efficiency and model convergence for continuous features, min-max scaling is used to rescale the continuous features in the range from zero to one. Scaling the values of different features on the same magnitude efficiently avoids neuron saturation[5] when randomly initializing the neural network.

In original datasets, the outcome variables are continuous lifetime values. A special soft-tile-coding method is developed to transform the continuous outcome into a soft binary vector. Similar to a binary vector, the summation of a soft binary vector is equal to one. The difference is that the soft binary vector not only consists of the values of 0 and 1, but also of some decimal values such as $1/n$ $(n = 2, 3, \dots)$. This type of soft binary vector is referred to as a soft-tile-encoded vector in this dissertation.

After the encoding process of input features and outcome variables, a customized Neural Network is proposed with a SoftMax layer to learn the mapping between the input features and the encoded output labels. Specifically, the output of the SoftMax layer corresponds to the encoded output label using the soft-tile-coding technique. The customized Neural Network with its output related to a soft-tile-encoded vector is named as the STC-NN model.

A decoder process is developed for the soft-tile-coding. The decoding process is a method that transforms a soft-tile-encoded vector into its probability along its original continuous lifetime. Instead of obtaining one output (like a common tile-coding-based

[5] Without scaling features, the coefficients of the features with larger magnitude will be smaller. The coefficients of features with smaller magnitude will be larger.

neural network), the STC-NN algorithm can obtain a probability distribution of broken rail occurrence within a specified study period.

5.4 Encoder: Soft-Tile-Coding

5.4.1 Tile-Coding

Tile-coding is a general tool used for function approximation (Sherstov and Stone, 2005). In this study, the continuous lifetime is partitioned into multiple tiles which are treated multiple categories, and each category relates to a unique time range. One partition of the lifetime is called one tiling. Generally, multiple overlapping tiles are used to describe one specific range of the lifetime. There is a finite number of tiles in a tiling. In each tiling, all tiles have the same length of time range, except for the last tile.

For a tile-coding with m tilings and each with n tiles, for each time period T on the lifetime horizon, the encoded binary feature is denoted as $F(T|m,n)$, and the element $F_{ij}(T)$ is described as:

$$F_{ij}(T) = \begin{cases} 1, & T \in \left[i\Delta T - d_j, (i+1)\Delta T - d_j\right) \\ 0, & \text{otherwise} \end{cases}; \quad i = 1,2,\dots,n; j = 1,2,\dots,m \tag{5-2}$$

where ΔT is the length of the time range of each tile, and d_j is the initial offset of each tiling.

In general, the tile-coded vector is defined as follows:

Definition 1: $F(T|m,n) = \left\{F_{ij}(T)\middle|\, i = 1,2,\dots,n; j = 1,2,\dots,m\right\}$ is called a tile-encoded vector with parameter m and n if it satisfies the conditions (a) $F_{ij}(T) \in \{0,1\}$ and (b) $\sum_i F_{ij}(T) = 1$.

Figure 5.5 illustrates two examples for tile-coding of two lifetime values at time (a) and (b) with three tilings ($m = 3$) which contain four tiles ($n = 4$). It is found that time (a) is located in the tile-1 for tiling-1, and in the tile-2 for both tiling-2 and tiling-3. The encoded vector of time (a) is given by $(1,0,0,0 \mid 0,1,0,0 \mid 0,1,0,0)^T$. Similarly, for time (b) The encoded vector is calculated as $(0,0,1,0 \mid 0,1,0,1 \mid 0,0,0,1)^T$.

Normally, a specific lifetime value could be encoded into a binary vector using tile-coding if an event occurs. However, in some situations, no events occur during the observation time and the event of interest is assumed to occur in the future. In this case, the censored lifetime is obtained, and the exact lifetime is unavailable. The usual tile-coding function cannot be used to encode this censored data. To address this issue, the soft-tile-coding approach is proposed.

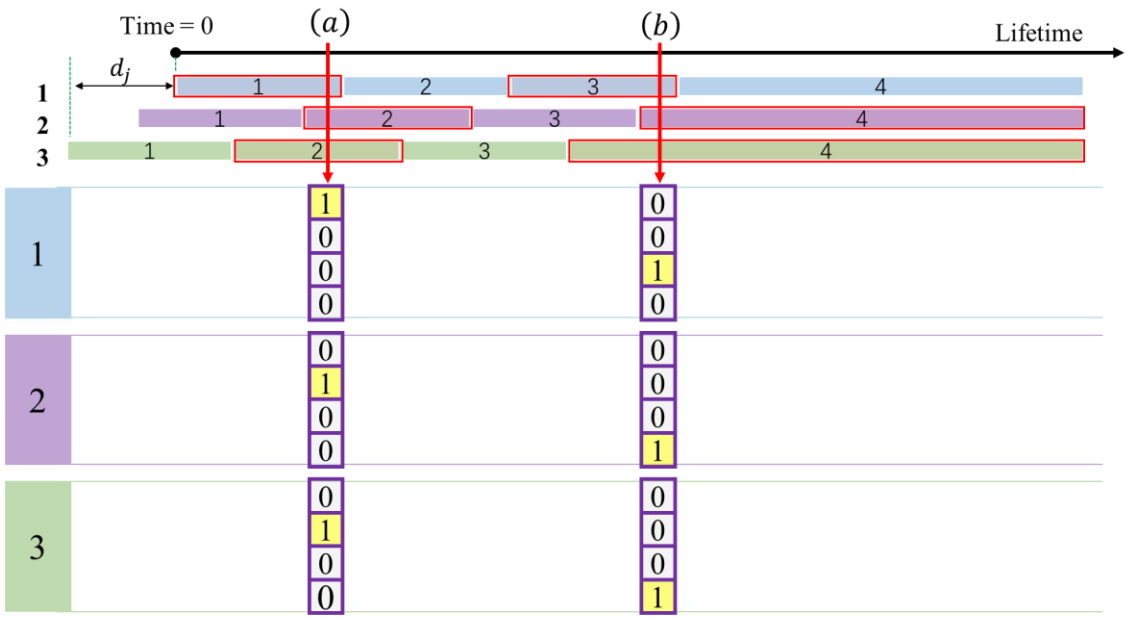


Figure 5. 5 Illustrative Example of Tile-Coding

Note: (a) and (b) indicate two different lifetimes.

5.4.2 Soft-Tile-Coding

The soft-tile-coding function is applied to transform the continuous lifetime range into a soft-binary vector, which is a vector whose value is in range [0, 1]. When the event of interest is not observed before the end of observation, the lifetime value is censored, and exact lifetime is not observed. Although the exact lifetime is not observed for the event, it is true that the event of interest did not occur within the observation time period. Equivalently, the event will happen in the future, beginning at the current ending observation time. By using soft-tile-coding, the proposed model could be improved and achieve better prediction performance. The mathematical process is as follows:

For a soft-tile-coding with m tilings and n tiles in each tiling, given a time range $T \in [T_0, \infty)$ on the timeline, the encoded binary feature is denoted as $S(T|m,n)$, and the element $S_{ij}(T)$ is described as:

$$S_{ij}(T) = \begin{cases} 1/(n-k_j+1), & i \geq k_j \\ 0 \ , \ \text{otherwise} \end{cases}; \ \ i = 1,2,\dots,n; j = 1,2,\dots,m \tag{5-3}$$

where

$$k_j = \underset{j}{\text{argmax}}\, F_j(T_0) \tag{5-4}$$

and $F_j(T_0)$ is the encoded binary feature vector of the jth tiling using tile-coding.

In general, define the soft-tile-encoded vector as follows:

Definition 2: $S(T|m,n) = \{S_{ij}(T) \big| \ i = 1,2,\dots,n; j = 1,2,\dots,m\}$ is called a soft-tile-encoded vector with parameter m and n if it satisfies the conditions (a) $S_{ij}(T) \in [0,1]$ and (b) $\sum_i S_{ij}(T) = 1$.

One example of soft-tile-coding with three tilings ($m = 3$), each of which contain four tiles ($n = 4$), is illustrated in Figure 5.6. It is found that the time T is located in the tile-3, tile-3, and tile-4 for tiling-1, tiling-2, and tiling-3, respectively. The soft-tile-encoded vector is given as $(0,0,0.5,0.5 \mid 0,0,0.5,0.5 \mid 0,0,0,1)^T$. In comparison, the tile-encoded vector is $(0,0,1,0 \mid 0,0,1,0 \mid 0,0,0,1)^T$.

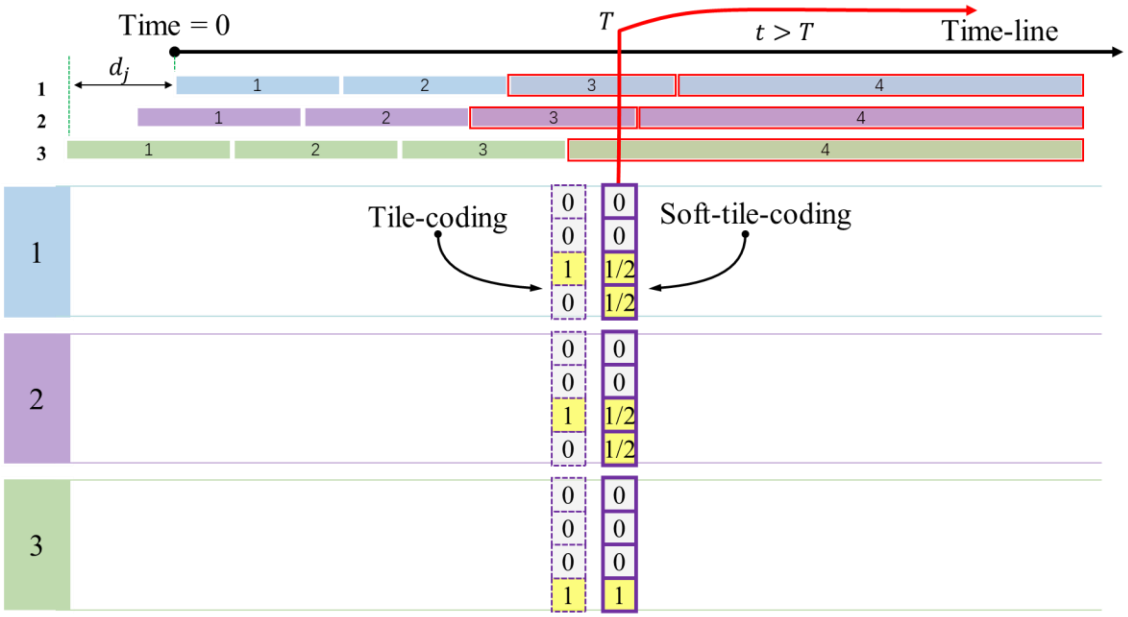


Figure 5. 6 Illustrative Example of Soft-Tile-Coding

5.5 Architecture of STC-NN Model

5.5.1 Forward Architecture of STC-NN Model

As presented in Figure 5.7, the forward architecture of STC-NN model is mainly based on a Neural Network. There are multiple processes to get the output probability of event occurrence over time from the input features. There are three main parts of the model: (1) a neural network, (2) a SoftMax layer with multiple soft-max functions, and (3) a decoder: probability transformation. The input of the model is transformed into a vector with values in range [0, 1]. The input vector is denoted as $g = \{g_i \in [0,1] | i = 1,2, \dots M\}$. The hidden layers are densely connected with a nonlinear activation function specified by the hyperbolic tangent, $\tanh(\cdot)$.

There are $m \times n$ output neurons of the neural network, which connect to a SoftMax layer with m softmax functions. Each SoftMax function is bound with n neurons. The

mapping from the input g to the output of the softmax layer can be written as $p(g|\theta)$, where θ is the parameter of the NN. According to Definition 2, $p(g|\theta)$ is a soft-tile-encoded vector with parameter m and n.

The soft-tile-encoded vector $p(g|\theta)$ is only an intermediate result, and it can be transformed into probability distribution by a decoder.

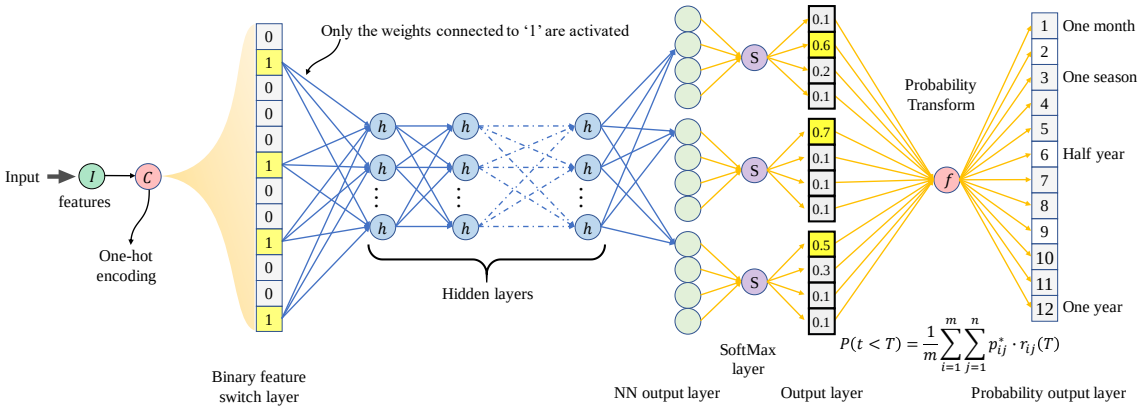


Figure 5. 7 Forward Architecture of STC-NN Model for Prediction

5.5.2 Backward Architecture of STC-NN Model

The backward architecture of the STC-NN model for training is presented in Figure 5.8. Given a feature set as the input, a soft-tile-encoded vector is calculated through the SoftMax layer. Instead of going further for probability transformation, in the training process the model loss is calculated with the soft-tile-encoded vector as the final output and a loss function can be defined as Equation (5-5):

$$\mathcal{L}(g, T|\theta, m, n) = \frac{1}{2}\|p(g|\theta) - F(T|m, n)\|^2 \tag{5-5}$$

where, $p(g|\theta)$ is the output of the STC-NN model, given input g with parameters θ. $F(T|m, n)$ is a tile-encoded vector if the feature set g relates to an observed lifetime T;

otherwise, $F(T|m,n) = S(T|m,n)$, which is a soft-tile-encoded vector if the feature set g relates to an unknown lifetime during the observation period with length T.

Given a training dataset with batch size of N, denoted as $\{G = \{g_1, g_2, \dots, g_N\}, T = \{T_1, T_2, \dots, T_N\}\}$, the overall loss function can be written as:

$$\mathcal{L}(G,T|\theta,m,n) = \frac{1}{2}\sum_{i=1}^{N}\|p(g_i|\theta) - F(T_i|m,n)\|^2 \quad (5\text{-}6)$$

The training process is given as an optimization problem which aims to find the optimal parameters θ^*, such that the loss function $\mathcal{L}(G,T|\theta,m,n)$ is minimized. The optimal parameters θ^* is written as Equation (5-7).

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}(G,T|\theta,m,n) \quad (5\text{-}7)$$

The optimal solution of θ^* can be estimated using the stochastic gradient descent (SGD) algorithm, which is achieved by randomly picking one record $\{g_i, T_i\}$ from the dataset, and following the updated process using Equation (5-8):

$$\theta \leftarrow \theta - \alpha \cdot \frac{\partial p(g_i|\theta)}{\partial \theta} \cdot \big(p(g_i|\theta) - F(T_i|m,n)\big);\ i = 1,2,\dots,N \quad (5\text{-}8)$$

where α is the learning rate and $\partial p(g_i|\theta)/\,\partial\theta$ is the gradient (first-order partial derivative) of the output soft-tile-encoded vector to parameter θ. It should be noted that the calculation of the gradients $\partial p(g_i|\theta)/\,\partial\theta$ is based on the chain rule from the output layer backward to the input layer, which is known as the error back propagation. Practically, this research uses a mini-batch gradient descent algorithm instead of a pure SGD algorithm to balance the computation time and convergence rate.

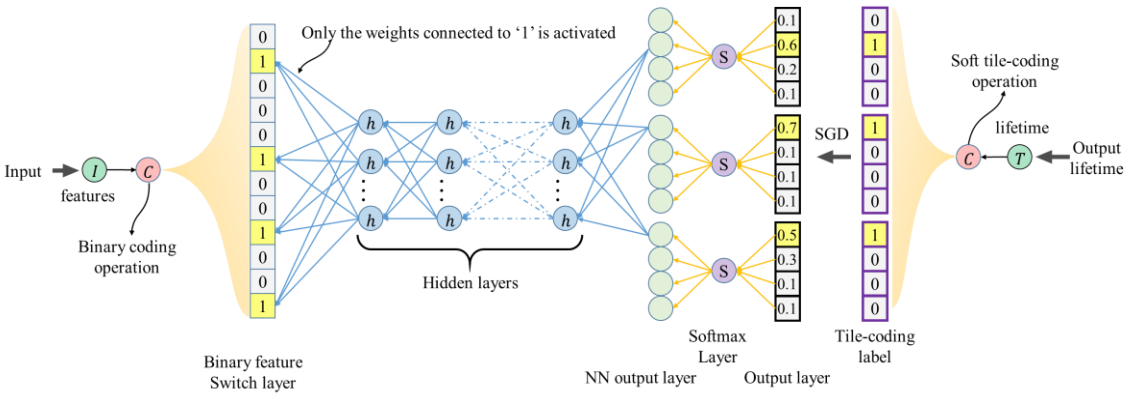


Figure 5. 8 Backward Architecture of STC-NN Model for Training Process

5.5.3 Training Algorithm of STC-NN Model

Different from the training algorithms commonly used for typical neural networks, the training algorithm of STC-NN is customized to deal with the skewed distribution in the database. For a rare event, the dataset recording can be highly imbalanced (i.e. more non-observed events than the observed events of interest due to their rarity). In the model dataset, the number of records with the observation of broken rails is about 4.34% of the whole dataset. According to Definition 3, the IR of the broken rail dataset is about 22:1, and it is quite challenging to learn from our dataset.

Definition 3: Imbalance Ratio (IR) is defined as the ratio of the number of records without event occurrence to the number of records with events.

To enhance the performance of the neural network model, instead of feeding the data randomly, a constraint is proposed for fed model data (training data) in the training process. The definition of Feeding Imbalance Ratio (FIR) is described below.

Definition 4: Feeding Imbalance Ratio (FIR) is defined as the IR of each mini-batch of data to be fed into the model during the training process.

For example, if $FIR = 1$, it means that each fed mini-batch of data contains half records with events and the other half records without events. When $FIR = 22$, the ratio between records with non-event and records with events in the dataset fed into the model is the same as the original dataset. Note that FIR is an important parameter for training the STC-NN model. If the FIR is too large, the dataset fed into the model is highly imbalanced, and it is difficult to learn the feature combination related to the event occurrence. However, if the FIR is too small, the features related to the event are well learned by the model. However, it will lead to a problem of over-estimated probability of the event occurrence. The pseudo code of the training algorithm is presented as follows:

Table 5. 3 Training Algorithm for Probability Prediction by STC-NN Model

Input:

FIR, $batch\_size$, n_epoch, m, n, α

Training dataset: (G, T);

The numbers of layers and neurons of neural network;

Initialize:

Initialize a neural network $p(* \mid \theta)$;

Split the (G, T) into $(G, T)^+$ and $(G, T)^-$ according to broken rail occurrence;

Main:

For _ in range (n_epoch), do

$(G, T)^+ = (G, T)^+.\text{shuffle}()$

$(G, T)^- = (G, T)^-.\text{shuffle}()$

For _ in range ($\text{round}(size((G, T)^+)/batch\_size)$), do

$(G,T)_i^+ = (G,T)^+.\text{next\_batch}(batch\_size)$

$(G,T)_i^- = (G,T)^-.\text{next\_batch}(FIR * batch\_size)$

$F_i^+ = tile\_coding(T_i^+)$

$S_i^- = soft\_tile\_coding(T_i^-)$

$(G,F)_i = \text{shuffle}\big(\text{concat}(G_i^+, G_i^-), \text{concat}(F_i^+, S_i^-)\big)$

Update the parameter θ of $p(*|\theta)$ given mini-batch $(G,F)_i$.

End For

End For

Output: The neural network $p(*|\theta)$.

Note: all superscript + and – indicate records with and without broken rails, respectively.

5.6 Decoder: Probability Transformation

The decoder of soft-tile-coding is used to transform a soft-tile-encoded vector into a probability distribution with respect to lifetime. Given the input of a feature set g, soft-tile-encoded output $p(g|\theta) = \{p_{ij}|i = 1, \dots n; j = 1, \dots m\}$ can be obtained through the forward computation of the STC-NN model. Decoder-like operation is used to transform $p(g|\theta)$ into probability of broken rails associated with a specific prediction period. The decoder of soft-tile-coding is defined as follows:

Definition 5: Soft-tile-coding decoder. Given a lifetime value $T \in [0, \infty)$, and a soft-tile-encoded vector $p = \{p_{ij}|i = 1, \dots n; j = 1, \dots m\}$, the occurrence probability $P(t < T)$ is estimated as:

$$P(t < T) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}^* \cdot r_{ij}(T) \tag{5-9}$$

where, *m* and *n* are the number of tilings and tiles respectively; p_{ij}^{*} and $r_{ij}(T)$ are the probability density and effective coverage ratio of the *j*-th tile in the *i*-th tiling, respectively. The value of p_{ij}^{*} can be calculated using p_{ij} divided by the length of time range of the corresponding tile. Note that there is no meaning for time $t < 0$, so the length of the first tile of each tiling should be reduced according to the initial offset d_j, and p_{ij}^{*} is calculated as follows.

$$p_{ij}^{*} = \begin{cases} p_{ij}/\Delta T & , \ i > 1 \\ p_{ij}/(\Delta T - d_j) & , \ i = 1 \end{cases} \tag{5-10}$$

The effective coverage ratio $r_{ij}(T)$ can be calculated according to Equation (5-11):

$$r_{ij}(T) = \begin{cases} t_{ij}(T)/\Delta T & , \ i > 1 \\ t_{ij}(T)/(\Delta T - d_j) & , \ i = 1 \end{cases} \tag{5-11}$$

where, $t_{ij}(T) = [\![[i\Delta T + d_j, (i+1)\Delta T + d_j) \cap [0, T]]\!]$ is the length of intersection between time range of the *j*-th tile in the *i*-th tiling and the range $t \in [0, T]$. The operator $[\![\cdot]\!]$ is used to obtain the length of time range.

According to Definitions 2 and 5, it is easy to verify that $P(t = 0) = 0$ and $P(t < T \mid T \to \infty) = 1$. And $P(t < T)$ can be interpreted as the accumulative probability of event occurrence within the lifetime T. An example of the soft-tile-coding decoder is given in Figure 5.9. The vector p is the output of the STC-NN model and the red rectangles on the tiles are $t_{ij}(T)$.

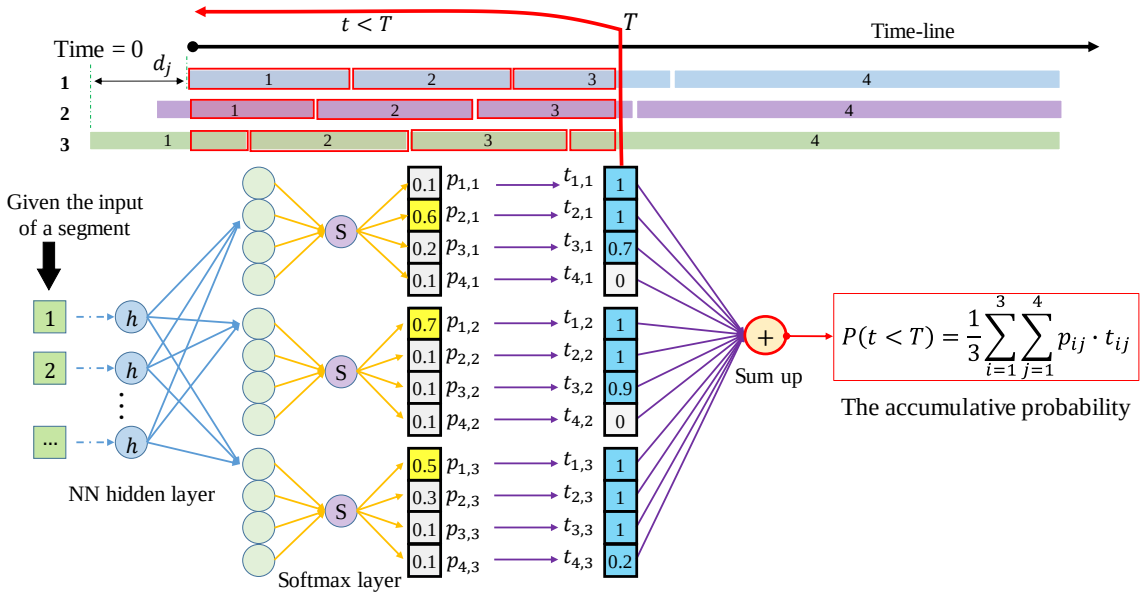


Figure 5. 9 Process of Probability Transformation

There is an upper time limit when the essential parameter n and ΔT are determined. Definition 6 is used to specify the total predictable time range of the STC-NN model, as follows.

Definition 6: Total Predictable Time Range (TPTR) is defined as the time period between defined starting observation time and ending observation time. The TPTR of the STC-NN model is defined as $TPTR = (n-1)\Delta T$, where n is the number of tiles in each tiling and ΔT is the length of each tile. In this research, n tiles in each tiling cover the lifetime range between starting observation time and maximum failure time among all the research data. Normally, the failure has not been observed till the ending observation time would involve censored lifetime data. Therefore, the maximum failure time among all the data should be infinite. The first n-1 tiles are set with a fixed and finite time length of ΔT which covers the observation period. The last tile covers the time period $t > (n-1)\Delta T$ which is beyond the observation. No additional information about the failure time is

provided by the last tile for the prediction. Therefore, the effective total predictable time range (TPTR) equals $(n-1)\Delta T$.

5.7 Model Development

After the dataset is prepared, the whole dataset is split into the training dataset and test dataset according to different timestamps. As a standard setup, the data from 2012 to 2014 are used for training, while the data from 2015 and 2016 are used as a test dataset to blindly validate the model.

The STC-NN model is developed and trained with the training dataset. In this case study, the default parameters of the STC-NN model are presented in Table 5.4. There are 50 tilings, and 13 tiles in each tiling. The length of each tile ΔT is 90 days, which means the TPTR of the STC-NN model is 3 years. Furthermore, the parameters of the training process are presented in Table 5.4. Note that the learning rate is set to be 0.1 initially, and then decreases by 0.001 for each epoch of training.

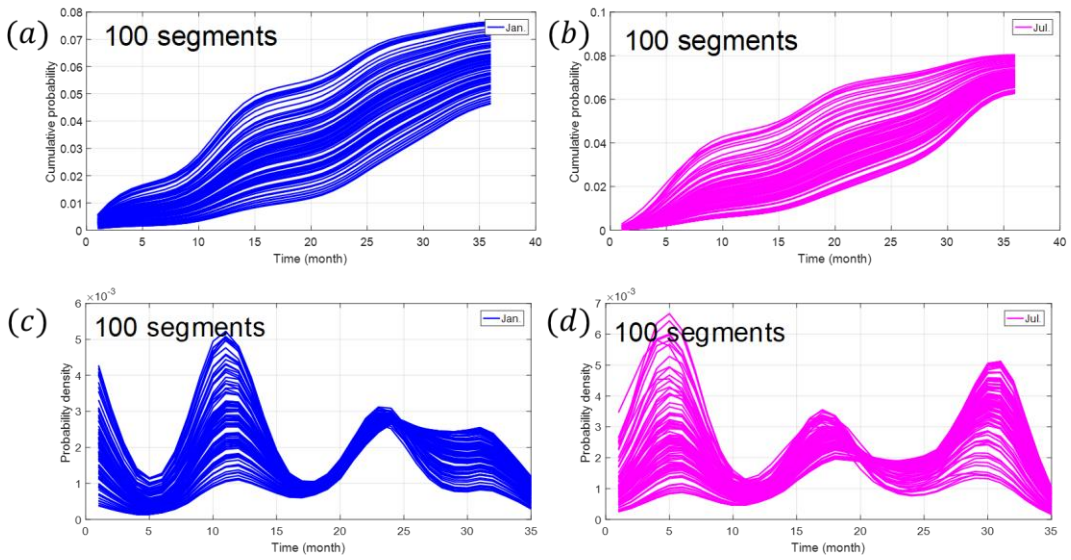
Table 5. 4 Parameter Setup of STC-NN Model

Parameter	Setup
m	50
n	13
ΔT	90 days
d_j	Randomly generated from a uniform distribution between $[0, \Delta T)$
FIR	1
$batch\_size$	128

n_epoch	20
α	0.1, decreasing by 0.001 for each epoch of training.
Hidden layers of NN	2 layers, each with 200 neurons.

5.7.1 Cumulative Probability and Probability Density

In order to illustrate the output of the STC-NN model, 100 segments are randomly selected from the test dataset. The predicted results of these 100 segments are described in Figure 5.10. The left two sub-figures (a) and (c) show the cumulative probability and probability density respectively with timestamp (starting observation time) January 1, and the right two, (b) and (d), show these with the timestamp July 1. The overall length of the time axis is 36 months which equals to the total predictable time range. As shown in Figure 5.10.a and 5.10.b, the slope of the cumulative probability curve varies in terms of time axis. The time-dependent slope of cumulative probability is measured as the probability density in terms of time axis which are plotted as Figure 5.10.c and Figure 5.10.d. The probability density is a wave-shaped curve which represents the fluctuation periodically. In Figure 5.10.c and Figure 5.10.d, the peaks of the probability density curve occur regularly with a time circle which is proved to be one year. The probability density represents the hazard rate or broken rail rate with respect to the time axis. Figure 5.10.c and 5.10.d state that the broken rail risk varies in one year and the highest broken rail probability is associated with the winter season in one year. With the timestamp being same, the probability density curves of different segments have the same shape. The values of the probability density given a time moment are different which is due to the variant characteristics associated with different segments. More details would be explained in the following Section 5.7.2.



Note: Jan indicates January 1st; Jul indicates July 1st; a. cumulative probability with timestamp January 1st; b. cumulative probability with timestamp July 1st; c. probability density with timestamp January 1st; d. probability density with timestamp July 1st

Figure 5. 10 Cumulative Probability and Probability Density for STC-NN Model

5.7.2 Illustrative Comparison between Two Typical Track Segments

In this section, two typical segments from the test dataset are selected to discuss details of the cumulative probability and probability density. Some main features for the two selected segments are listed in Table 5.5. There are totally over one hundred features (raw features and their transformations or combinations). However, only the most important features are shown in Table 5.5. The table shows that Segment A is 0.3 miles in length with 135 lbs/yard rail and it has been in service for 18.7 years, while Segment B is 0.5 miles in length with 122 lbs/yard rail and its age is 37 years. As for the broken rail occurrence, compared to Segment A where no broken rail was observed, there is a broken rail found at Segment B in 341 days with the starting observation date of January 1, 2015. It means that a broken rail occurred at Segment B around the end of year 2015.

Table 5. 5 Comparison of Two Segments from the Test Dataset

Features	Segment A	Segment B
Division	D1	D1
Prefix (Anonymized)	AAA	BBB
Track type	Single track	Single track
Starting observation date	January 1, 2015	January 1, 2015
Rail weight (lbs/yard)	135	122
Rail age (years)	18.7	37
Curve or not	With curve	With curve
Annual traffic density	25.12 MGT	23.57 MGT
Segment Length (miles)	0.3	0.5
Broken rail occurrence	None found in two years (2015 and 2016)	Found in 341 days

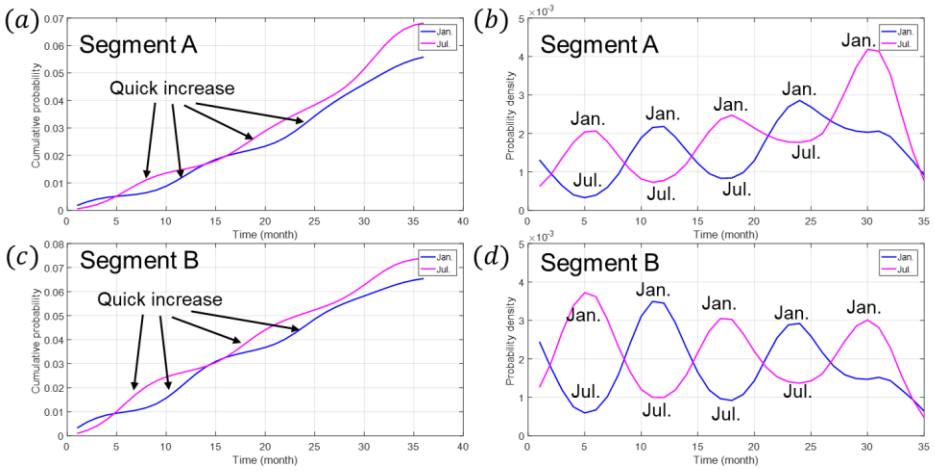
Using the trained STC-NN model, the broken rail probabilities of these two segments are predicted and the results are presented in Figure 5.11. The top two figures show the cumulative probability and probability density of Segment A, while the bottom two show the cumulative probability and probability density of Segment B. The blue and pink curves represent the starting timestamps of January 1st and July 1st, respectively. The following observations are made:

(1) The overall cumulative probability of Segment A is smaller than that of Segment B.

(2) The slope of cumulative probability curve varies by time, which result in a curve-shaped probability density.

(3) The peak of the probability density curve occurs regularly with a time circle, which is approximately one year.

(4) Probability density of broken rail occurrence is higher in the colder season (e.g., December, January, and February) compared to the warmer season (e.g., June, July, August). Similar conclusions can be found in previous studies (e.g., Liu et al., 2013a).



Notes: Pink lines represent the prediction with January 1st as the starting observation time (timestamp). Blue lines represent the prediction with July 1st as the starting observation time (timestamp).

Figure 5. 11 Illustrative Comparison Between Two Typical Segments in Terms of Broken Rail Probability Prediction

5.7.3 Sensitivity Analysis of STC-NN Model

Some assumptions and parameters are generated during the development of the STC-NN. A sensitivity analysis is necessary to test the reasonability of the model setting.

5.7.3.1 Training Step Analysis

Training step in neural network is an important parameter that could affect the model performance on both the training data and test data. In the sensitivity analysis of training step, the range of test training step is changed from 50 to 500. Figure 5.12 plots the AUC for one season and one year with respect to training steps. More details about AUC is explained in the following Section 5.8. The AUC values for one season and one year increase as the training step increases for the training data, while the AUC values for test data decrease as the training step increases.

The possible reason is that more training step increases the complexity of the model, and further increases the performance of the classifier on the training data. However, the complexity of the model affects its generalization. The more complex the model is, the less generalized it might be. Less generalizability of the model could result in an overfitting problem, leading to decreased model performance for the testing data.

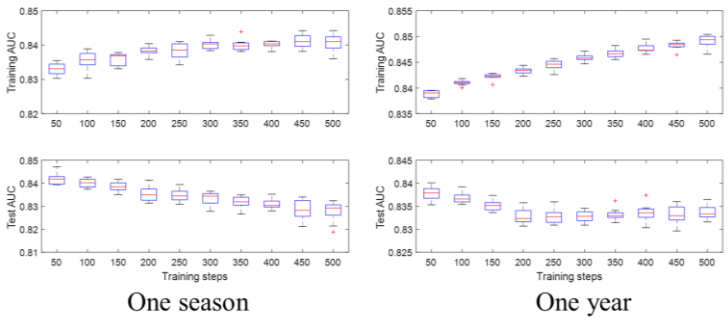


Figure 5. 12 AUC Values with Respect to Number of Training Steps

5.7.3.2 Sensitivity Analysis of Model Parameters

This section discusses some main factors that contribute to model performance. Generally, almost all the presented parameters have significant influence on the

performance of the STC-NN model. The model parameters can be divided into three groups according to their functions: (1) soft-tile-coding of the output label: number of tilings m, number of tiles in each tiling n, length of each tile ΔT, the initial offset of each tiling d_j; (2) the FIR used in the training algorithm; and (3) the nonlinear function approximation using neural network: the training step n_epoch, learning rate α, the batch size $batch\_size$ and the number of hidden layers and neurons.

Since the core part of the STC-NN model is a neural network with multiple layers, the influence of n_epoch, α, $batch\_size$ and the numbers of hidden layers and neurons can be tuned similarly as in commonly used neural networks. For illustrative convenience, this research only focuses on the influence of the parameters of soft-tile-coding and the FIR during the training process.

For soft-tile-coding, the number of tilings m should be large enough so that the decoded probability can be smooth. Otherwise, the probability density will become stair-stepping. Especially, when $m = 1$, the STC-NN model degenerates into the model for the Multi-Classification Problem (MCP). The ΔT and n together influence the TPTR. Firstly, TPTR is determined according to the maximal lifetime observed from the training dataset. Secondly, a proper value of ΔT is set. With the pre-defined ΔT, the number of tiles is calculated to keep TPTR unchanged. In an extreme condition, if $\Delta T = \text{TPTR}$, $n = 2$ and $m = 1$, the STC-NN model degenerates into a model for the Binary Classification Problem (BCP).

To analyze the influence of FIR on the performance of the STC-NN model, a replication experiment is carried out, where the training algorithm is executed 10 times to

evaluate the AUC of each FIR in $\{1, 2, 3, 4, 5, 7, 10, 15, 22\}$. The results are presented using box-plot, as shown in Figure 5.13, where the red notch is the median value, and the upper and lower limit of the blue box show the 25% and 75% percentile, respectively. Figures (a), (b) and (c) in Figure 5.13 are related to one-month, one-season and one-year time prediction period, respectively. It shows that the AUCs decrease and the variance of AUCs gets larger given the larger FIR values, indicating that the prediction accuracy becomes lower. The result becomes more unstable when the mini-batch of data fed into the dataset is more imbalanced. When the value of FIR equals 22, which is the exact IR of the training dataset, most of the AUCs are less than 0.8, and some even become less than 0.7 within the one-year prediction period. The large variance indicates that the performance is unstable, and the results may be hard to repeat. In contrast, if FIR is set to be 1, the AUCs outperform all those with FIR > 1 and the variance is very small as well, indicating that the result is more stable and repeatable.

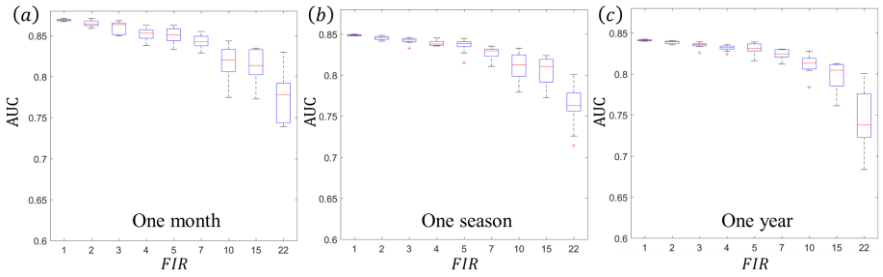


Figure 5. 13 AUC Values with Respect to FIR in the STC-NN Model

5.8 Model Validation

5.8.1 Comparison with Alternative Models

Prior studies have argued that threshold-dependent performance metrics (i.e., precision and recall) are problematic for two reasons: (1) they depend on an arbitrarily-

selected threshold (Rahman and Devanbu, 2013); and (2) they are sensitive to imbalanced data (He and Garcia, 2009). Instead, the Area Under the Receiver Operating Characteristics Curve (AUC) is used to measure the discrimination power of our STC-NN model as suggested by recent research (Lessmann et al., 2008).

AUC is computed by measuring the area under the curve that plots the true positive rate (TPR) against the false positive rate (FPR), while varying the threshold that is used to determine whether a record is classified as positive or negative. Values of AUC range from 0 (worst performance), to 0.5 (random guessing performance), and to 1 (best performance). The true positive rate and the false positive rate can be calculated based on the confusion matrix. The confusion matrix (Table 5.6) is used to present the results of a classification algorithm, TP (True Positives) records the actual positives that are correctly classified. Similarly, FP (False Positive) records the actual negatives that are incorrectly classified as positives. In this research, the "positive" represents the occurrence of broken rail, and the "negative" means no broken rail on the rail segment within the prediction period. The four values, TP, FN, FP, TN, are used to calculate true positive rate (TPR) and false positive rate (FPR) as shown in Equation (5-12) and Equation (5-13) respectively.

Table 5. 6 Confusion Matrix for Classification Validation

	Predicted (Classified)	
Actual	Positive	Negative
Positive	TP	FN
Negative	FP	TN

$$True\ Positive\ Rate\ = \frac{TP}{TP+FN} \tag{5-12}$$

$$False\ Positive\ Rate\ = \frac{FP}{FP+TN} \tag{5-13}$$

As shown in Table 5.7, the proposed STC-NN model appears to outperform several commonly used alternative algorithms such as Random Forests, XGBoost, Logistic Regression, traditional Neural Network and Cox Proportional Hazard Regression model, using AUC as the performance metric. The AUC for one-month prediction in the STC-NN model is around 0.86, while the largest AUC for one month of alternative algorithms is around 0.76. The AUC for one-year prediction in the STC-NN model is around 0.84, while the largest AUC for one year of alternative algorithms is around 0.83. The computational time represents the running efficiency of the model, which also indicates the model's superiority. The STC-NN model is implemented on the platform of TensorFlow-GPU, which ran on NVIDIA GTX860m of a Lenovo Y50 laptop. For the STC-NN model and Cox Proportional Hazard Regression model, the models only need to be trained one time to obtain the prediction results for one month, three months, and one year. However, for the other models, three different models should be trained to get the results for these three different prediction periods. As shown in Figure 5.14, the STC-NN model performs more efficiently than alternative algorithms.

Table 5. 7 Model Comparison

Model	AUC		
	One month	Three months	One year
Random Forests	0.71	0.78	0.79

Logistic Regression	0.62	0.69	0.71
XGBoost	0.76	0.79	0.83
Neural Network	0.78	0.80	0.81
Cox Proportional Hazard Regression	0.75	0.78	0.81
STC-NN (Our Developed Algorithm)	0.86	0.85	0.84

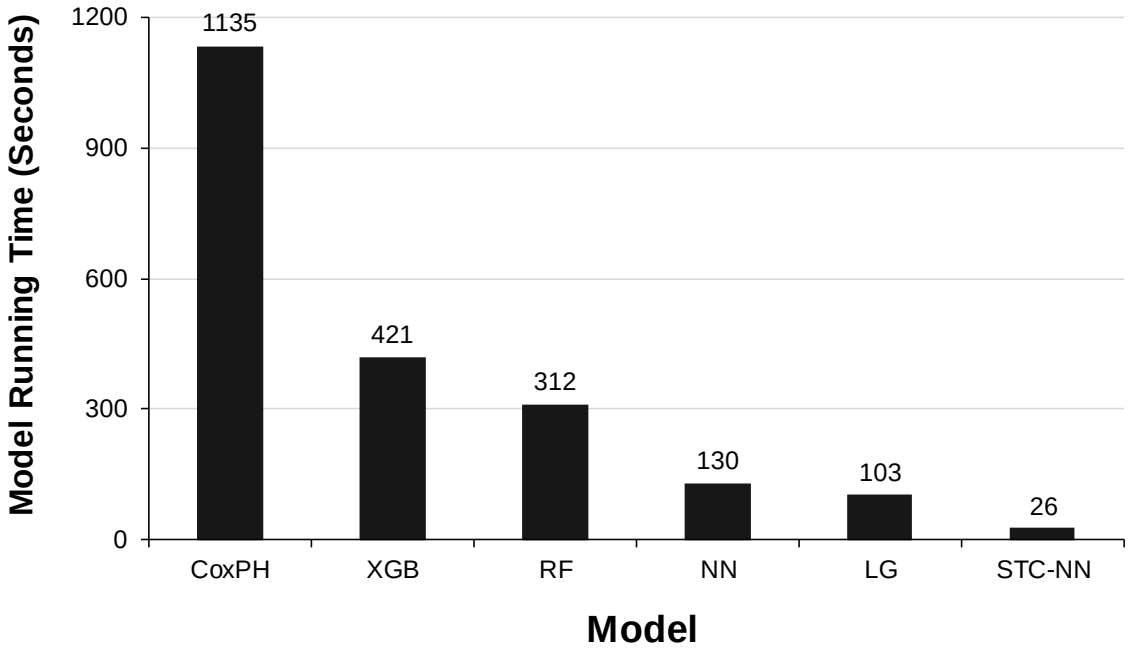


Figure 5. 14 Comparison of Computation Time for One-Month Prediction by Alternative Models

5.8.2 Model Performance with Respect to Prediction Period

For a given observation time T_0, the reference label $L_r(T_i|T_0)$ is defined as follows:

$$L_r(T_i|T_0) = \begin{cases} 1, & T_i < T_0 \\ 0, & otherwise \end{cases}; \quad i = 1,2 \tag{5-14}$$

where T_i is the lifetime of the i-th segment from the test dataset. Equation (5-14) can be interpreted as a binary operator that labels T_i as 1 if T_i is less than T_0, otherwise, labelling it as 0.

Given the same observation time T_0, the cumulative probability at time T_0 is calculated as its predicted probability. When given a specific threshold $P_0 \in [0, 1]$, the predicted probability is converted into a binary vector as shown in Equation (5-15).

$$L_p(T_0|P_0) = \begin{cases} 1, & P(t < T_0) > P_0 \\ 0, & otherwise \end{cases} \tag{5-15}$$

The output of the STC-NN model is the probability with respect to different prediction periods. Given a specific T_0, the corresponding AUC is calculated. Therefore, a series of time-dependent AUCs are obtained. Figure 5.15 plots the AUCs in term of different prediction period. Figure 5.16 shows the true positive rate against false positive rate with one month as the prediction period. The major findings from Figure 5.15 and Figure 5.16 are described below:

(1) As shown in Figure 5.15, the AUCs within three years are larger than 0.83, and the largest AUC is found in the first month, with AUC=0.86. Figure 5.16 shows the ROC curve with the prediction period as one month.

(2) As shown in Figure 5.15, the AUCs show a decreasing trend over time, indicating the model perform better for shorter term prediction.

(3) The time scope is 36 months (three years), which is related to the setup of the STC-NN model. Figure 5.15 shows that the AUCs stay almost unchanged after two years. This is because in test dataset the observation period only covers two years.

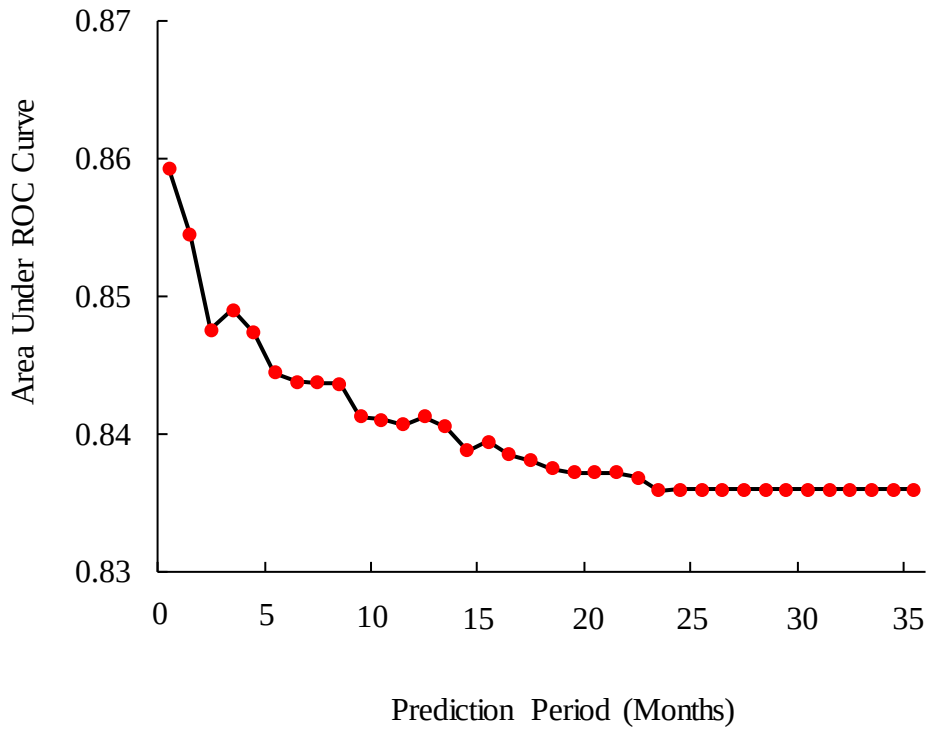


Figure 5. 15 Time-Dependent AUC Performance

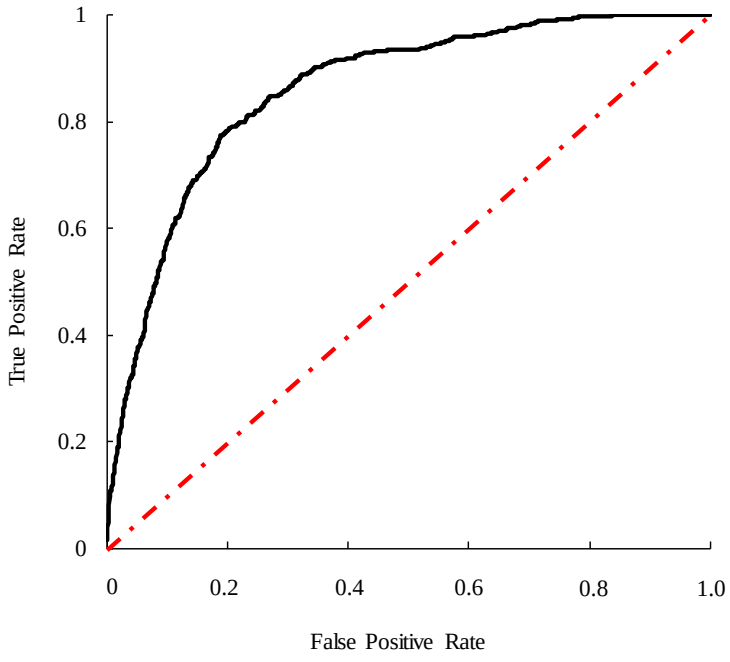


Figure 5. 16 Receiver Operating Characteristics Curve with Prediction Period as One Month

Figure 5.17 shows a comparison of the cumulative probability over time between the segments with (blue color line) and without (red color line) broken rails, respectively. The four sub-figures from (a) to (d) show the cumulative probabilities of broken rails at half-year, one-year, two-years and 2.5-years, respectively. For a short-term period, such as one-half year (Figure 5.17 a), the red curve (without observed broken rails) and blue curve (with observed broken rails) are clearly separated. As the prediction period gets longer (Figure 5.17 b, c, d), the cumulative probability curves overlap for the blue and red, making it difficult to separate the two curves, which leads to the decreasing trend of AUCs over time as shown in Figure 5.15. For long term prediction, the input feature set changes during the ‘long term’ as time-dependent factors such as traffic, rail age, geometry defects and some other maintenance are highly time-variant. This is a possible explanation for why the STC-NN model has better performance in the short term than in the long term.

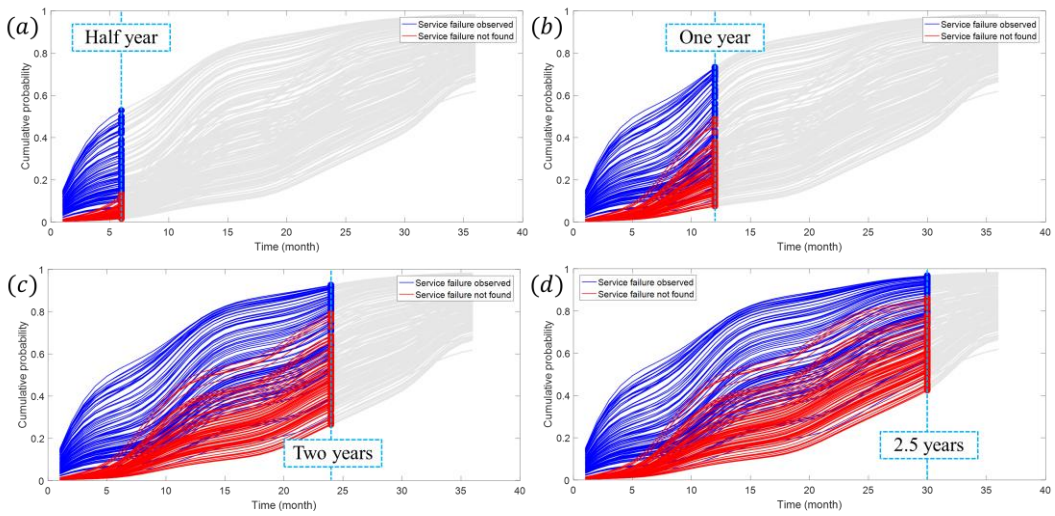


Figure 5. 17 Comparison of the Cumulative Probability by Prediction Periods Between the Segments with and without Broken Rails

5.8.3 Comparison between Empirical and Predicted Number of Broken Rails

To illustrate the model performance, this research also compares the empirical number of broken rails and predicted number of broken rails in one year on the network level. As Figure 5.18 shows, the total empirical numbers of broken rails in 2015 and 2016 are 823 and 844 respectively. The predicted number of broken rails for 2015 and 2016 are 768 and 773 correspondingly. The errors for 2015 and 2016 are 6.7 percent and 8.4 percent, respectively.

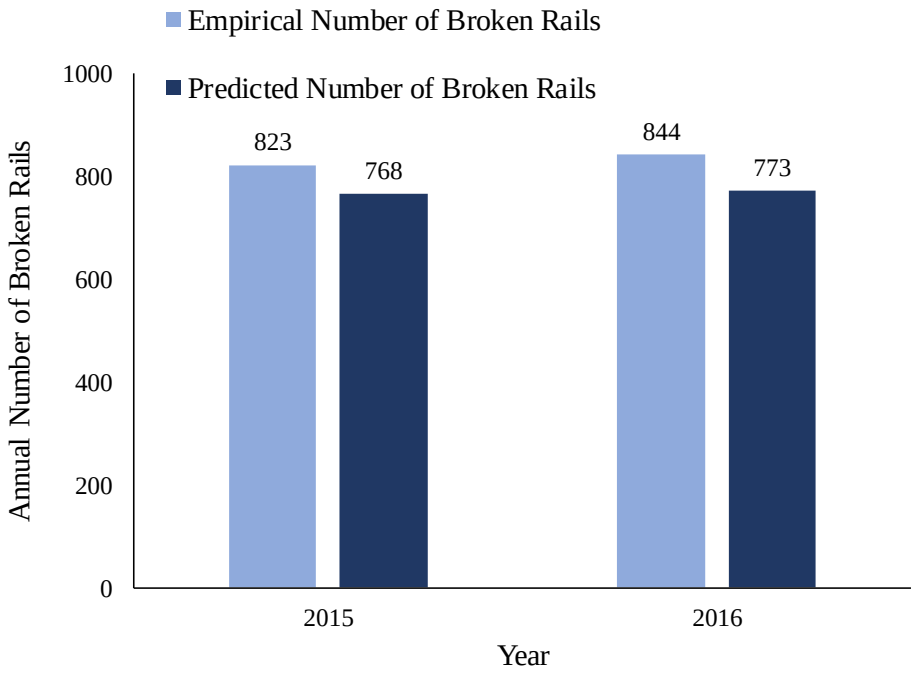


Figure 5. 18 Empirical and Predicted Numbers of Broken Rails on Network Level

5.9 Model Application

5.9.1 Network Screening to Identify Locations with High Broken Rail Probabilities

The prediction model can be used to screen the network and identify locations which may be more prone to broken rail occurrence. The results can be displayed via a curve in Figure 5.19. The x-axis represents the percentage of network scanned, while the y-axis is the percent of correctly "captured" broken rails, if scanning such scale of subnetwork. For example, if the broken rail prediction model (e.g., STC-NN as described above) is used to predict the probability of broken rails in one month, it can "find" over 71% of broken rails in one month (the percentage is weighted by segment length) by focusing on 30% of network mileage (as shown in Table 5.8). Without a model to identity broken-rail-prone locations, a naïve rule (which assumes that broken rail occurrence is

random on the network) might be screening 71% of network mileage to find the same percentage of broken rails.

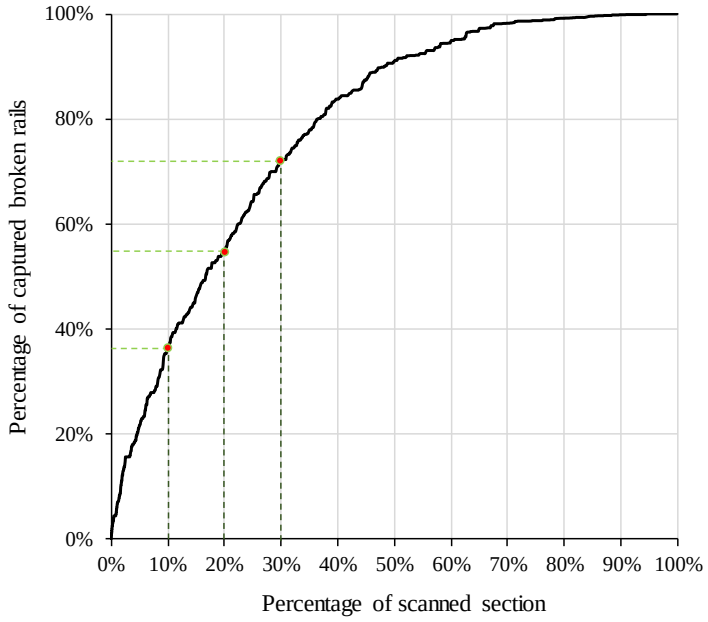


Figure 5. 19 Risk-Based Network Screening for Broken Rail Identification with Prediction Period as One Month

Table 5. 8 Percentage of Network Screening versus Percentage of "Captured" Broken Rails Weighted by Segment Length with Prediction Period as One Month

Percentage of Network Screening	Percentage of "Captured" Broken Rails (Percentage is Weighted by Segment Length)
10%	36.5%
15%	46.2%
20%	54.9%
25%	64.3%

30%	71.8%
35%	77.6%
40%	83.8%

5.9.2 GIS Visualization

The proposed broken rail prediction model can be applied to identify a shortlist of segments that may have higher broken rail probabilities. This information may be useful for the railroad to prioritize the track inspection and maintenance activities. In addition, the analytical results can be visualized on a Geometric Information System (GIS) platform. Figure 5.20 visualizes the predicted broken rail probability based on the categories of the probabilities (e.g., extremely low, low, medium, high, extremely high). These five groups are defined based on Jenks natural breaks classification method with predicted probabilities.

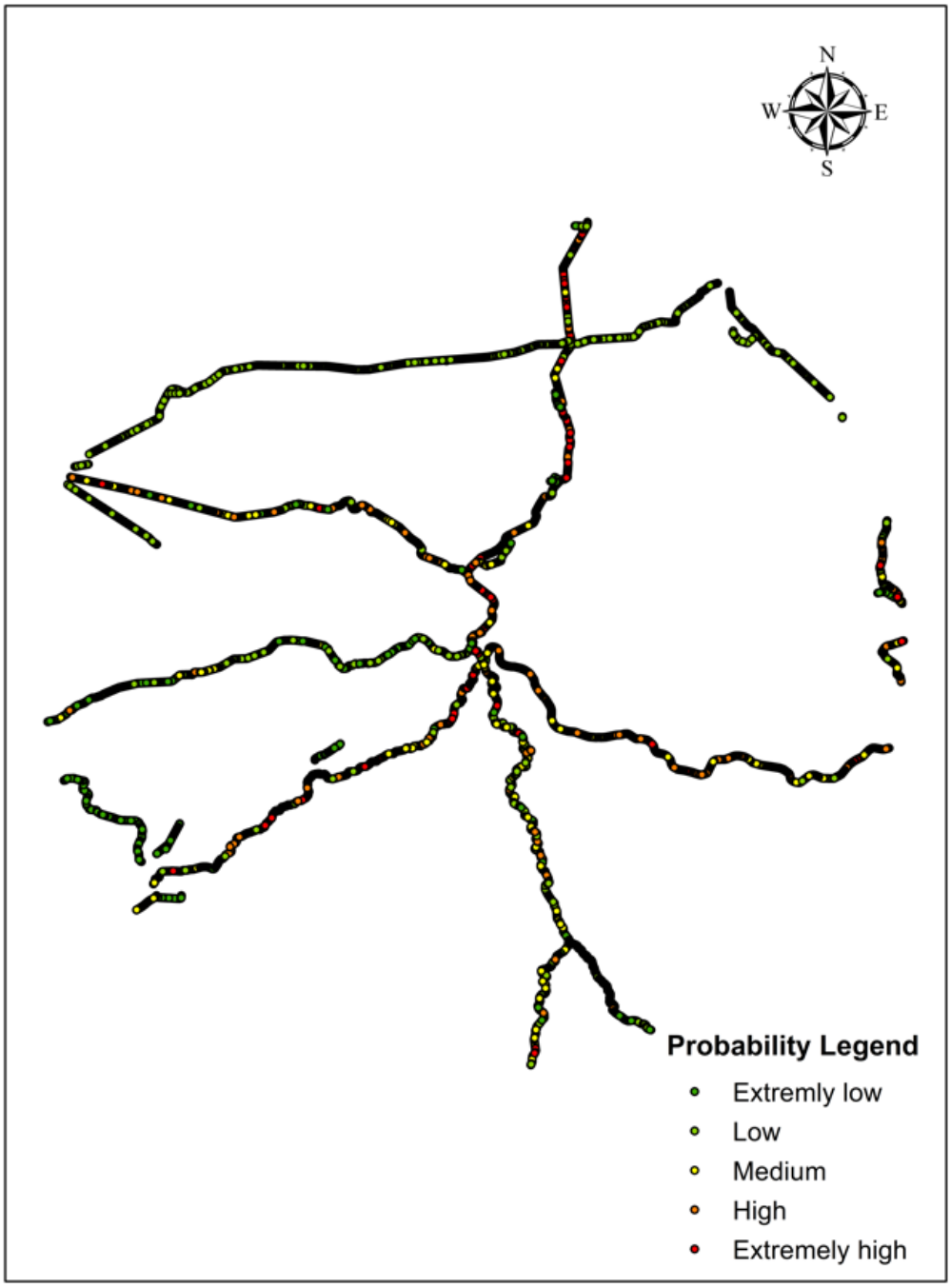


Figure 5. 20 Visualization of Predicted Broken Rail Probabilities Marked with Various Categories

Figure 5.21 shows that the 30 percent of the screened network mileage to identify the locations with relatively higher broken rail probabilities. As summarized in Table 5.8,

the model can identify over 71% of broken rails (weighted by segment length) by performing a screening of 30% of network, which is marked in red (Figure 5.22).

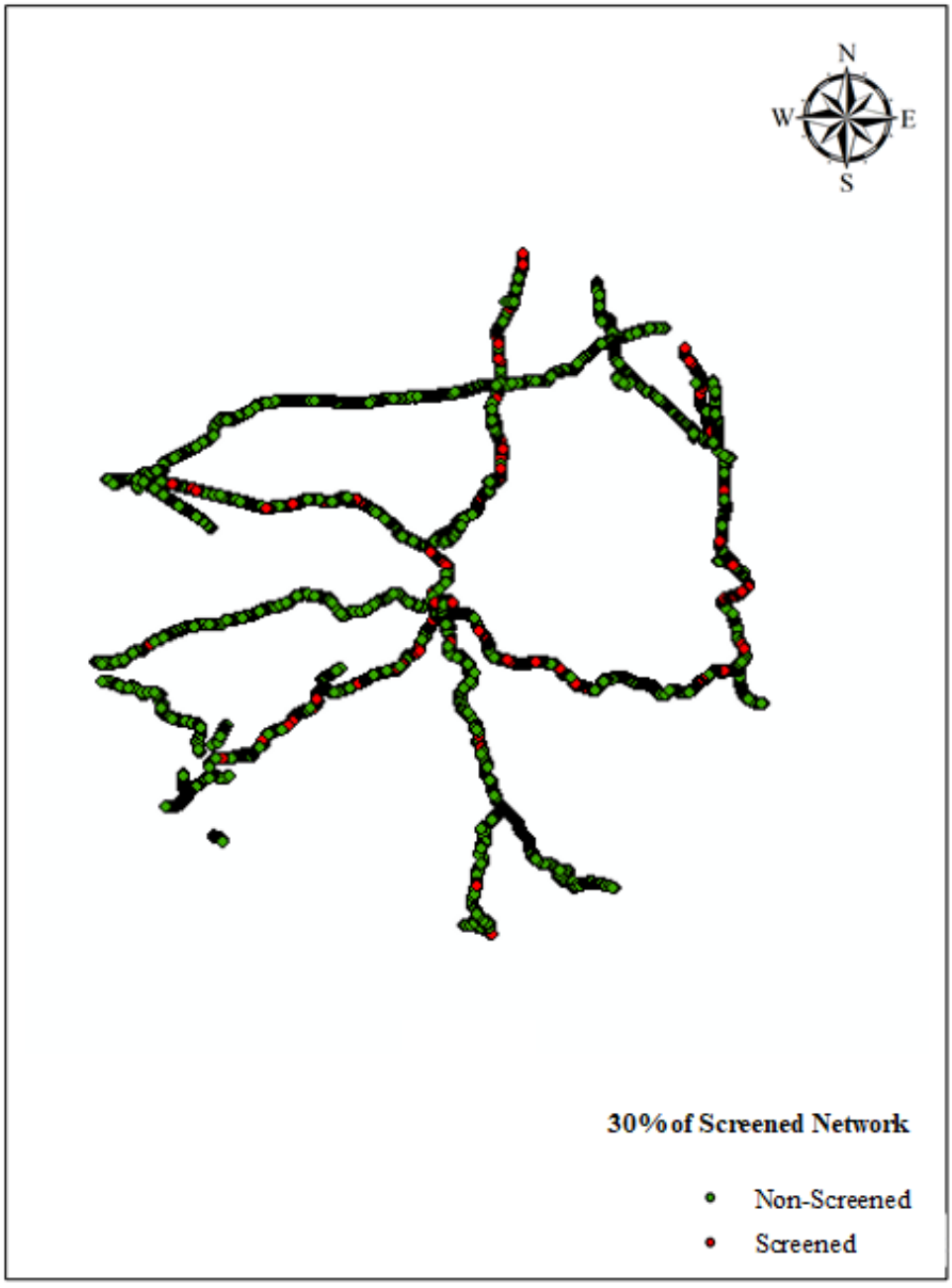


Figure 5. 21 Visualization of Screened Network (30% of Network Mileage) (Partial Display)

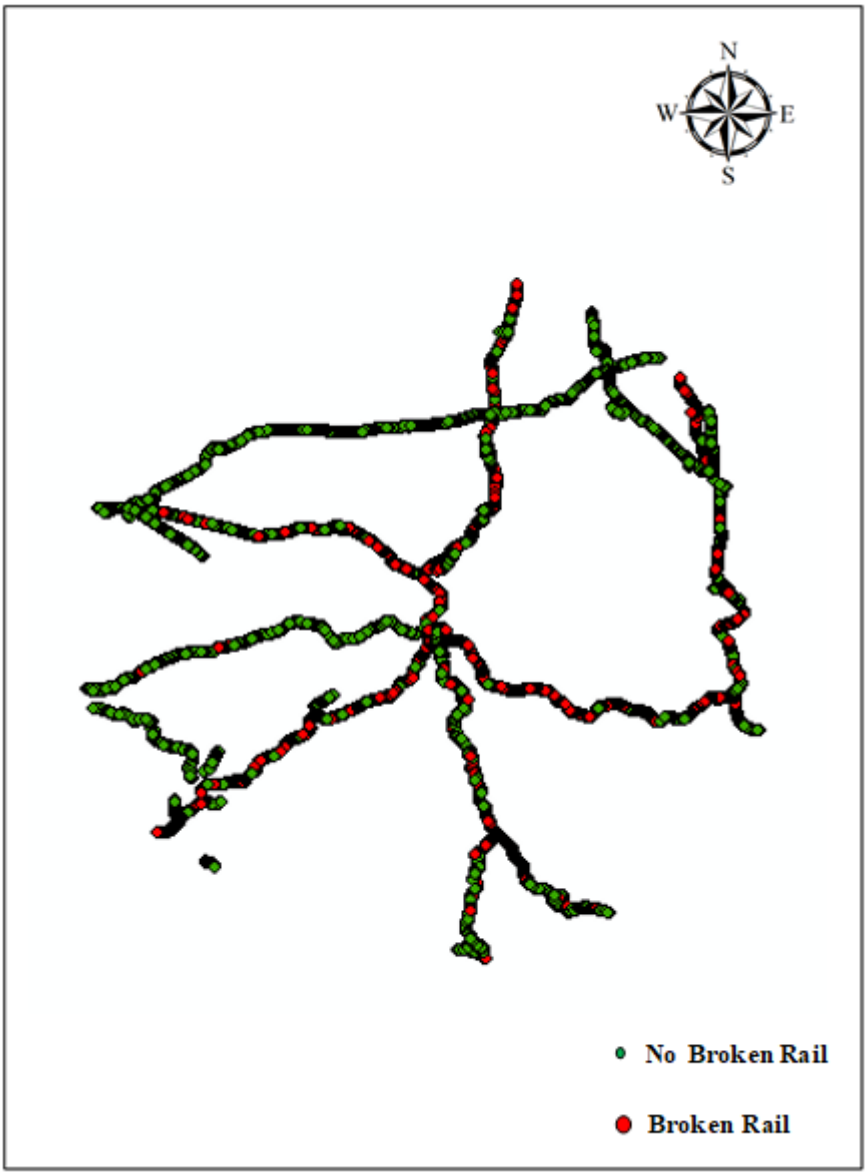


Figure 5. 22 Visualization of Broken Rails within Screened Network (30% of Network Mileage) (Partial Display)

5.9.3 Partial Features of Top 20 Segments with High Predicted Probability of Broken Rails

With ranking the predicted broken rail probability in one year, a list of locations with higher probabilities of broken rails is shown in Table 5.9.

Table 5. 9 Selected Feature Information of Top 20 Segments

Segment ID	Annual Traffic (MGT)	Rail Age (Year)	Rail Weight (lbs/yard)	Speed (MPH)	Curve Degree	Probability
1	53.26	21.01	135	50	0.94	0.392
2	60.26	38.93	139	50	0.35	0.379
3	58.90	10.66	136	50	0.27	0.379
4	38.73	30.38	135	60	0.25	0.378
5	70.17	1.48	136	60	0.11	0.377
6	73.83	27.35	133	57	0.24	0.377
7	57.36	40.17	139	50	0.34	0.377
8	59.83	2.40	136	50	0.34	0.376
9	59.27	36.96	140	50	0.25	0.374
10	44.93	18.95	135	38	1.43	0.370
11	70.90	31.22	136	58	0.00	0.370
12	58.43	31.45	134	50	0.32	0.370
13	74.78	22.48	134	40	1.13	0.369
14	78.91	34.98	122	57	0.00	0.369
15	55.33	26.71	135	50	0.44	0.369
16	56.34	23.60	137	50	0.18	0.368
17	62.45	11.51	136	46	1.00	0.368
18	63.21	21.33	135	50	0.41	0.368
19	67.88	15.91	135	50	1.19	0.368
20	85.87	18.67	135	58	0.73	0.368

5.10 Chapter Summary

This chapter proposes a novel, customized Soft-Tile-Coding-Based Neural Network (STC-NN) model to predict broken rail probability using railroad big data. First, tile-coding approach was applied to handle the exact lifetime values. Then, a soft-tile-coding technique was developed to deal with censored lifetime data when no broken rail was found during the observation period. Then, forward architecture of STC-NN with a SoftMax layer and a soft-tile-coding decoder was designed to fit the training data. After that, the backward architecture of the STC-NN model was presented and a customized training algorithm was designed to tackle the challenging problem of imbalance distribution in the research dataset (only a small portion of segments had broken rails in a short observation period). Feeding Imbalance Ratio (FIR) was proposed as a constraint on the fed data in the training process. Then, the soft-tile-coding decoder was derived to transform a soft-tile-encoded vector into a probability distribution.

A case study was carried out on a network-level dataset covering over 20,000 miles on mainline tracks. A comparison in terms of the prediction accuracy and computational time was conducted between the proposed STC-NN model and alternative machine learning algorithms using AUC as the primary measure of model performance. The main conclusions include:

(1) The output of the STC-NN model is the probability distribution of broken rail occurrence by time. A seasonal fluctuation has been observed, with the peak of the probability density likely occurring in the colder season (in this study, it includes December, January and February).

(2) The STC-NN model performs better for the short-term prediction than long-term prediction. The AUC for the three-year prediction period is around 0.84. For one-month-ahead prediction, the AUC of STC-NN is around 0.86.

(3) The STC-NN model outperforms several commonly used machine learning algorithms, in terms of both prediction accuracy and computational time.

(4) For one-month prediction, the risk-based screening 10%, 20%, and 30% of the total network mileage can "catch" 34%, 55%, and 71% of potential broken rails, respectively.

(5) The STC-NN algorithm is fast and suitable for broken rail prediction even for a short prediction horizon (e.g., one month or three months in advance).

CHAPTER 6

BROKEN RAIL-CAUSED DERAILMENT RISK MODEL

Chapter 5 describes the prediction of broken rail probability given a specified time period. In this chapter, a broken rail-caused derailment risk model is proposed. Three sections are included: (a) statistical analysis of relationship between broken rails and broken-rail derailments, in which the broken-rail derailment probability given a broken rail was calculated with respect to specific variables; (b) estimation of the severity of a broken-rail derailment, in which a decision tree model was developed accounting for several input variables; (c) example application of the broken-rail derailment risk model.

6.1 Overview of Broken Rail-Caused Derailment Risk Estimation

In this research, the definition of risk contains two elements: probability of an event (e.g., broken rail-caused train derailment) and the consequence (e.g., number of cars derailed) given occurrence of an event. As for broken-rail derailment risk, it is calculated through multiplying the broken-rail derailment probability by the broken-rail derailment severity:

$$Risk(DB) = P(DB) * S(DB) \quad (6\text{-}1)$$

where

$Risk(D \cdot B)$= broken-rail derailment risk,

$P(D \cdot B)$= the probability of broken-rail derailment,

$S(D \cdot B)$= the severity of broken-rail derailment,

DB= broken rail-caused derailment.

Broken-rail derailment has a very low probability of occurrence, and thus directly estimating its occurrence is challenging due to the very small sample size. A more practical method is to indirectly estimate broken rail derailment probability based on: 1) the probability of broken rail occurrence, and 2) the conditional probability of a derailment given a broken rail. Using Bayes' Theorem, broken rail derailment probability ($P(DB)$) can be calculated by:

$$P(DB) = P(D|B) * P(B) \tag{6-2}$$

where

$P(D|B)$= probability of broken-rail derailment given a broken rail;

$P(B)$= probability of a broken rail, which can be estimated by the broken rail prediction model (Chapter 5).

6.2 Statistical Relationship Between Broken Rails and Broken Rail-Caused Derailments

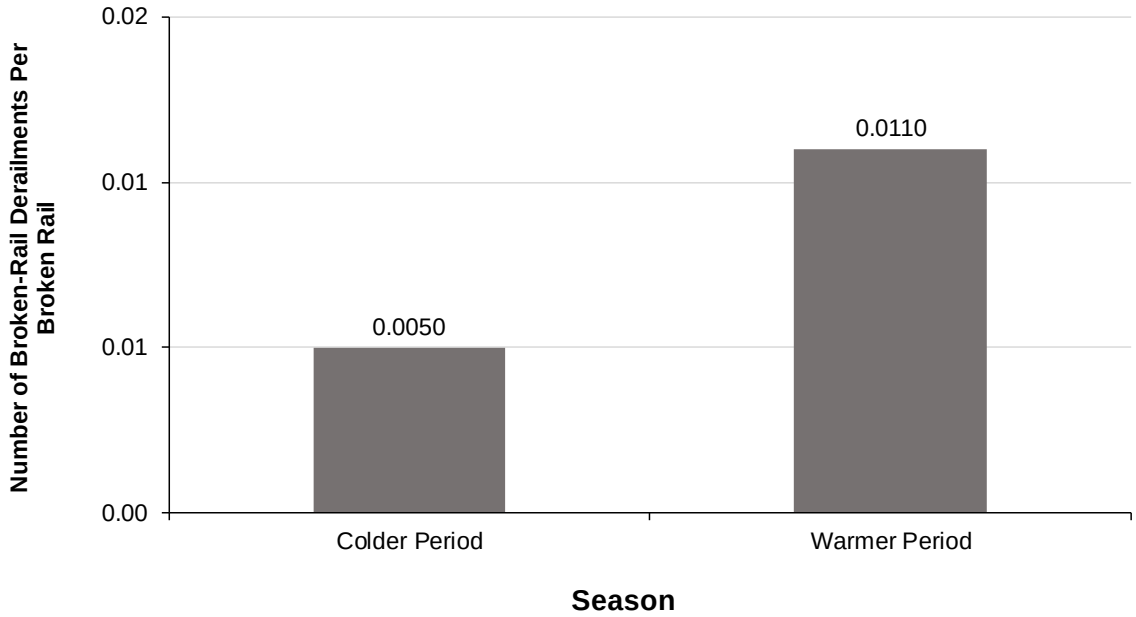
In order to estimate the probability of broken-rail derailment given a broken rail, the statistical relationship was analyzed between broken-rail derailments and broken rails. In this research, the broken rails from 2012 to 2016 were used in the broken rail prediction model. For consistency, the statistical analysis is conducted based on the broken-rail derailments which occurred from 2012 to 2016. From 2012 to 2016, the studied railroad had 25 mainline broken-rail derailments and 4,051 broken rails.

6.2.1 Univariate Statistical Analysis of Broken Rails and Broken-Rail Derailments

There are certain track-related characteristics or traffic conditions that are expected to be associated with the occurrence of broken-rail derailments. This section intends to analyze the relationships between the occurrence of broken-rail derailment per broken rail by single variables.

6.2.1.1 Derailment per Broken Rail by Season

This section considers two seasonal periods: colder period and warmer period (Figure 6.1). In this research, the colder period is from November to April and the warmer period is from May to October. The colder period accounts for approximately 80 percent of broken rails but only 64 percent of broken-rail derailments. The other 20 percent of broken rails occurred during the warmer period, but were associated with 36 percent of broken-rail derailments. In terms of the rate of broken-rail derailment per broken rail, the rate in the warmer period is two times of that in the colder period. This conclusion is consistent with previous research (Reiff, 1997). In lower temperatures, thermally induced tensile force is higher and rails are much more sensitive to repeated loading and are much more likely to break. Even though the colder period involves more broken rails, broken rails under colder temperatures could be easier to be detected (either visually or by electric track circuit continuity) when the rails are in tension and thus are further pulled apart. As a result, more broken rails might be detected and removed, before leading to train derailments. By contrast, in the warmer period, cracks are held more tightly when the rail is compressed and are more difficult to be detected. Therefore, these “weak” rails may not be easily detected but are more likely to cause derailments.



Notes: November to April for the Colder Period; Other Months for the Warmer Period

Figure 6. 1 Number of Broken-Rail Derailments per Broken Rail by Season

6.2.1.2 Derailment per Broken Rail by Track Curvature

Approximately 35 percent of broken rails occurred on tangent tracks, while only 24 percent of broken-rail derailments took place there. Slightly over half of the broken rails occurred at locations with curves lower than 2 degrees, which account for approximately half of the broken rail-caused derailments. In terms of curvature of over 2 degrees, only 13 percent of broken ails were associated with 24 percent of derailments. The ratio between broken-rail derailments and broken rails on curvatures over 2 degrees is two and half times higher than that on tangent tracks (Figure 6.2).

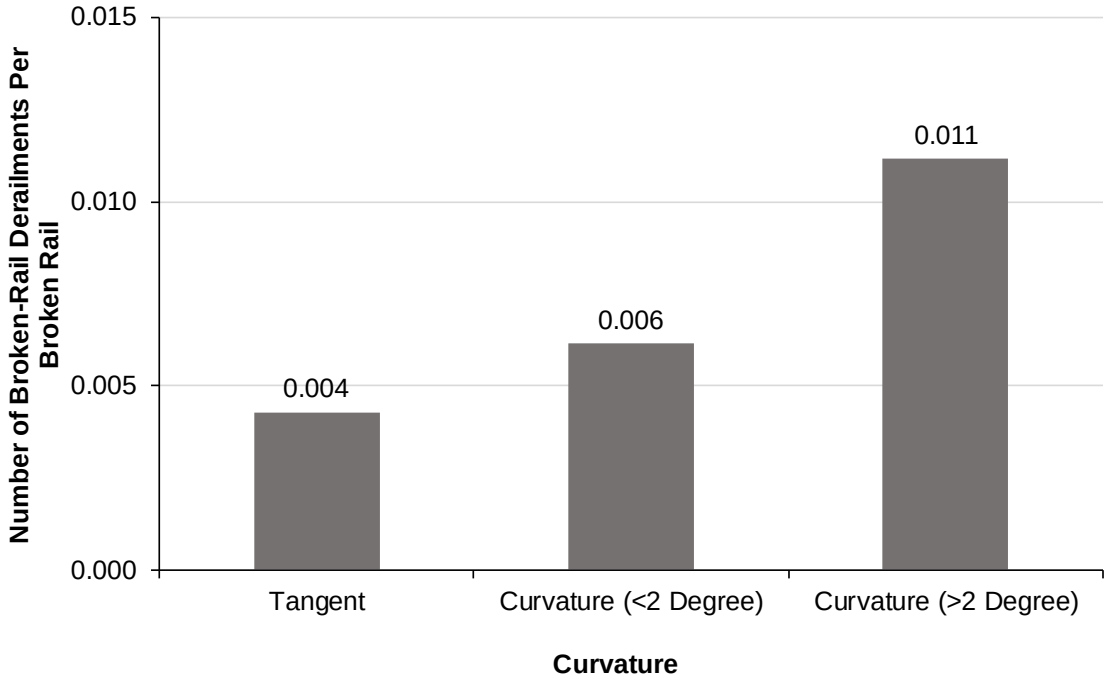


Figure 6. 2 Number of Broken-Rail Derailments per Broken Rail by Curvature

6.2.1.3 Derailment per Broken Rail by Signal

Less than 10 percent of broken rails and more than 50 percent of broken-rail derailments occurred in non-signaled territories (aka. dark territories). The rate measured by the number of broken-rail derailments per broken rail in non-signalized tracks is ten times of that in signalized tracks (Figure 6.3). A possible reason is that signaled trackage uses low-voltage electric current in the rails (known as “track circuits”) to detect the presence of trains in a given section. An important secondary benefit of track circuits is that they enable detection of several types of infrastructure problems, most notably in the context of this study, broken rails (Liu et al., 2017).

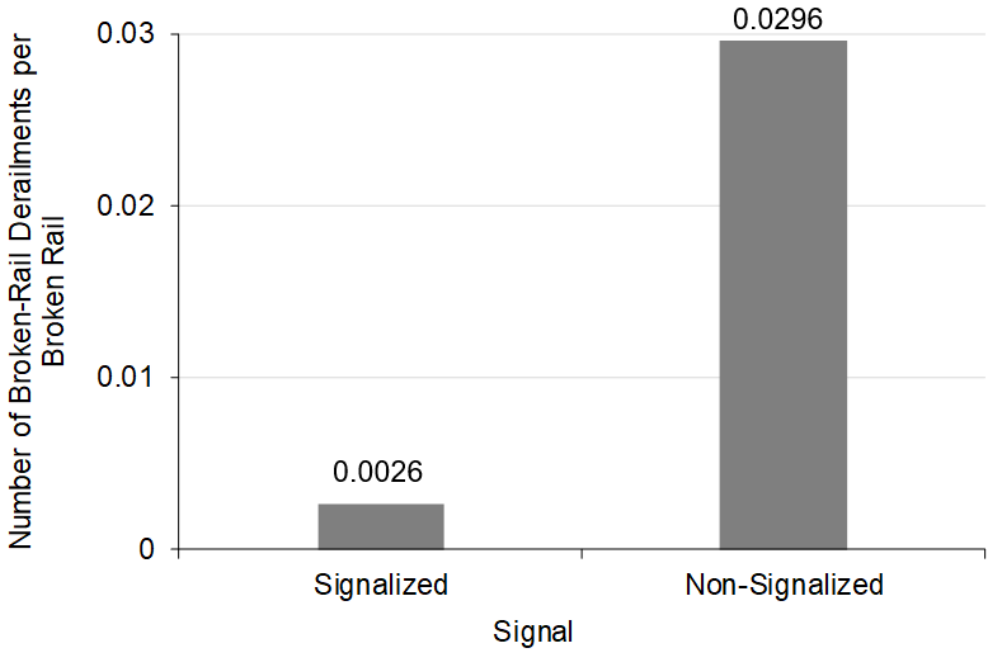


Figure 6. 3 Number of Broken-Rail Derailments per Broken Rail by Signal Setting

6.2.1.4 Derailment per Broken Rail by Annual Traffic Density (in MGT)

According to the annual traffic tonnage crossing over the track, the railway network can be divided into two categories: tracks with lower traffic density and tracks with higher traffic density. Higher traffic is defined as having annual traffic tonnage of more than 20 million gross tonnages (MGT), and if the annual traffic tonnage is below this threshold, the track has a lower traffic density. This classification referred to the previous study (Liu et al., 2017). About three-quarters of broken rails occurred at higher traffic locations, and these resulted in 40 percent of the broken-rail derailments. Although the prior literature found that higher traffic density causes an increase in rail defects or failures (Dick, 2001; Dick et al., 2003), these defects or failures may be addressed more promptly in high-density track territories, due to stricter inspection and maintenance standards.

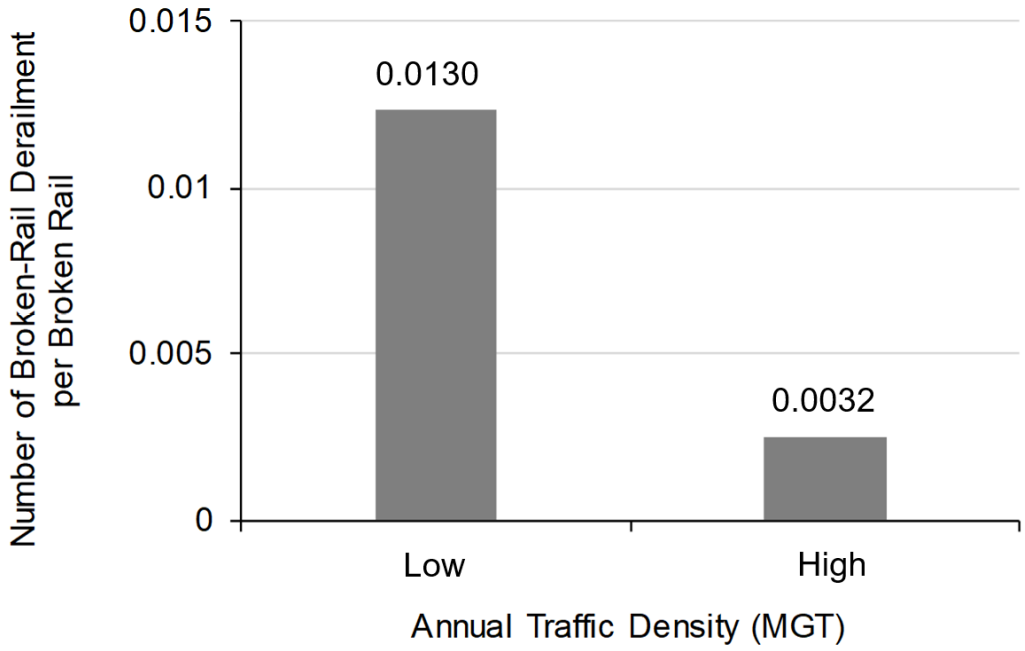


Figure 6. 4 Number of Broken-Rail Derailments per Broken Rail by Annual Traffic Density

6.2.1.5 Derailment per Broken Rail by FRA Track Class

The maximum speeds for FRA Track Class 1 and Class 2 are 10 mph and 25 mph (for freight railroads), respectively. The maximum speeds for FRA Track Class 3, and Class 4 are 40 mph, and 60 mph (for freight railroads), respectively. Approximately 13 percent of broken rails occurred at Track Class 1 and Track Class 2, while over 30 percent of broken-rail derailments took place on these tracks (Figure 6.5). For the combination of Class 3 and above, 86 percent of broken rails resulted in 68 percent of derailments. In terms of number of broken-rail derailments per broken rail, the ratio of Track Class 1 & 2 is three times of that of Class 3, and Class 4 in combination. Similar to the higher traffic territory, Class 3, and Class 4 have more stringent inspection and maintenance requirements

compared to lower track classes. As a result, this leads to a lower probability of derailment given broken rail occurrence.

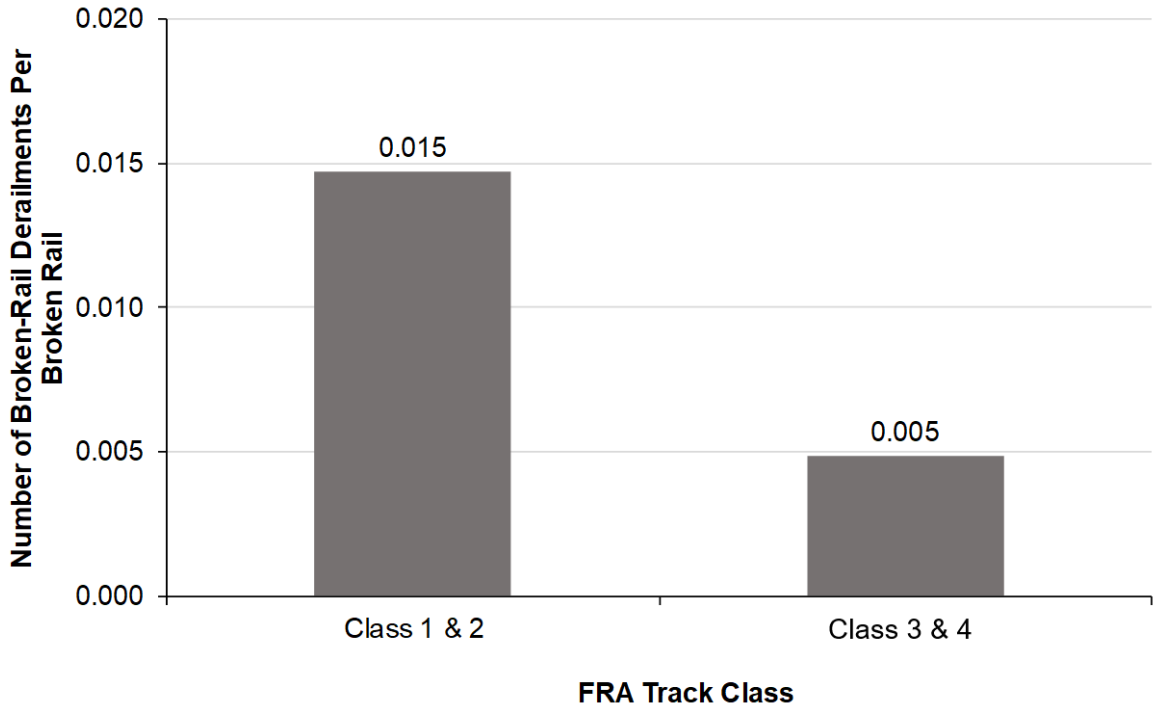


Figure 6. 5 Number of Broken-Rail Derailments per Broken Rail by FRA Track Classes

6.2.2 Multivariate Statistical Analysis of Broken Rails and Broken-Rail Derailments

The analyses above showed that season, traffic, curvature, and signal setting are all related to the conditional probability of derailment given a broken rail. However, the approach in the previous section only considered one variable at a time and ignored the interaction or correlation between multiple variables. This section develops a multivariate analysis of the relationship between broken rails and broken-rail derailments. Note that the analysis might be subject to greater uncertainty due to a smaller sample size in each category.

As shown in Figure 6.6, lower traffic density (< 20 MGT), non-signaled locations have the largest ratio measured as the number of derailments per broken rail. Around 11 percent of broken rails occurred in the lower traffic density, non-signaled territory, which resulted in over half of the broken rails. In the datasets, very few non-signaled tracks are associated with higher traffic tonnages. Only 0.6 percent of broken rails occurred on higher traffic, non-signalized tracks, where there were no broken-rail derailments in the studied period. By contrast, higher traffic, signalized locations account for almost 70 percent of broken rails. However, since the number of broken-rail derailments in this category is 36% of the total, the ratio of number of derailments per broken rail is relatively low. Note that the combination of non-signaled, high traffic is still rare and therefore, the sample size is quite small, leading to a greater statistical uncertainty.

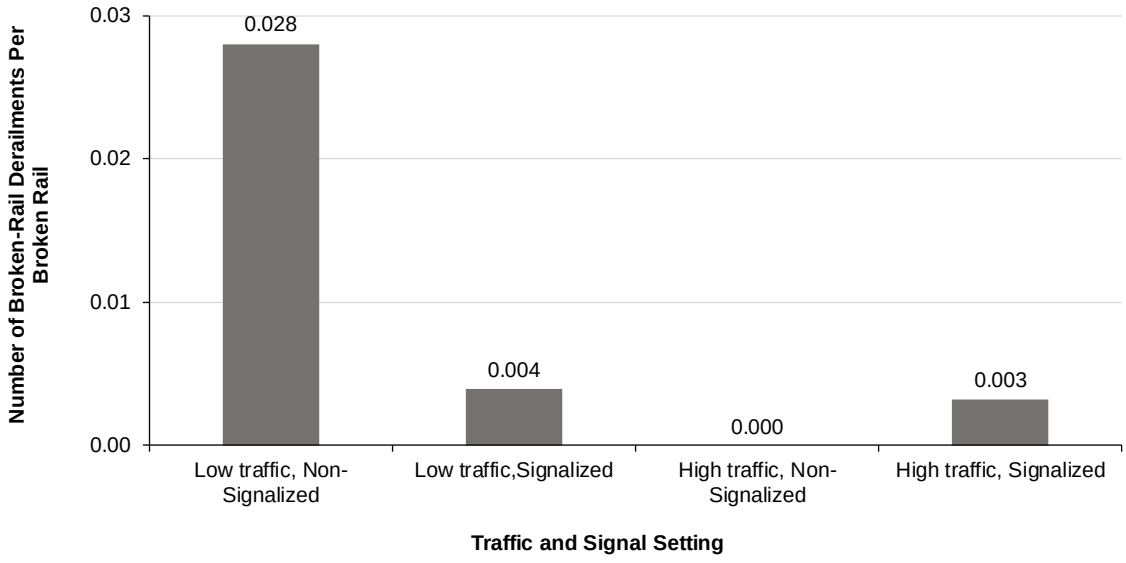


Figure 6. 6 Number of Broken-Rail Derailments per Broken Rail by Annual Traffic Density Level and Signal Setting

Figure 6.7 illustrates the distribution of broken-rail derailments per broken rail by season and signal setting. In general, in both the colder and warmer period, non-signalized

tracks have higher numbers of broken-rail derailments per broken rail. Only 2.5 percent of broken rails occurred on the non-signalized tracks during the warmer period, but they resulted in 20 percent of broken-rail derailments. By contrast, on signalized tracks in the colder period, there is a much lower ratio of broken-rail derailments to broken rails.

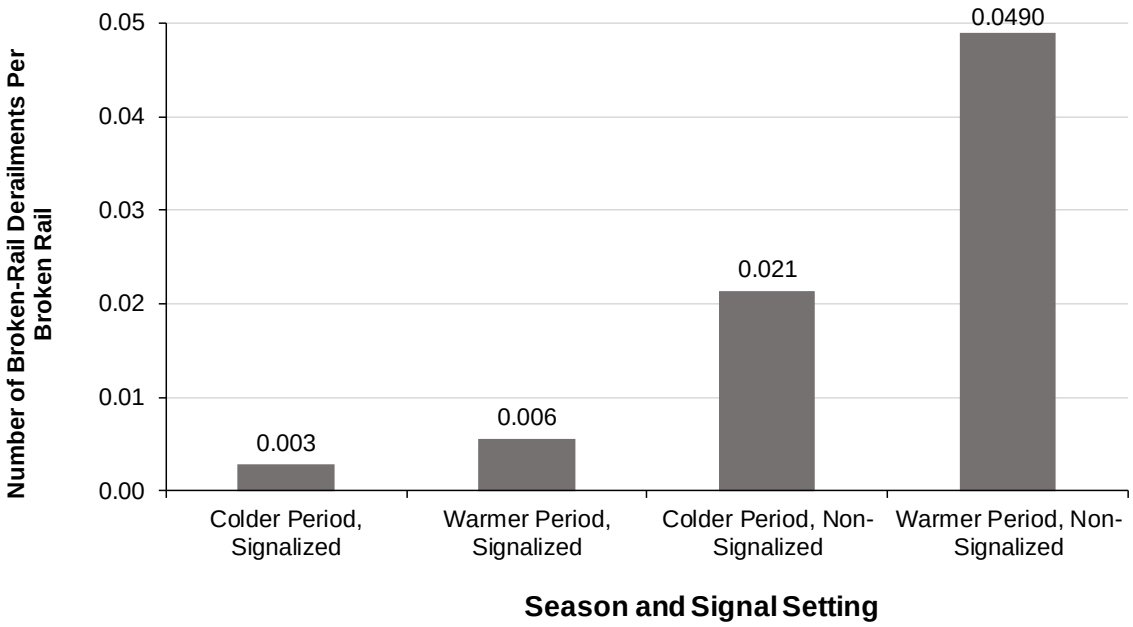


Figure 6. 7 Number of Broken-Rail Derailments per Broken Rail by Season and Signal Setting

6.3 Broken Rail-Caused Derailment Severity Estimation

Statistical analysis has been used in prior studies to model train derailment severity (Liu et al., 2013b; Saccomanno and El-Hage, 1989, 1991; Anderson, 2005, Bagheri, 2009). The most commonly used metric in the study of train derailment is the number of cars derailed per accident. The generic term of "cars" refers to all types of vehicles, including locomotives, railcars and cabooses, unless specifically stated otherwise. Monetary damage and number of casualties are also used to assess derailment severity. However, prior studies (Barkan et al., 2003; Martey and Attoh-Okine, 2018) stated that monetary damage is prone

to substantial variation due to factors such as the cost difference between locomotives and railcars, and differences in repair costs between regular track and special track.

In the analysis of train derailment severity, accident cause has been widely studied (Saccomanno and El-Hage, 1991; Barkan et al., 2003; Liu et al., 2012). Liu et al. (2013b) pointed out that broken rails are likely to involve high-frequency and high-severity derailments, and thus pose greater accident risk than other causes. For example, broken rails, as the most common cause of freight-train derailment on U.S. Class I mainlines, caused an average of 14 derailed cars in a freight-train derailment, which is greater than the average number of cars derailed in a bearing-failure-caused derailment (Liu et al., 2013b). This section will focus on modeling the severity of freight-train derailment caused by broken rails.

6.3.1 Methodology

6.3.1.1 Data Description

This research used broken rail-caused freight train derailment data on the main line of a Class I railroad from 2000 to 2017. In this period data was collected on 938 Class I broken rail-caused freight-train derailments on mainlines in the United States. The generic use of “cars” here refers to locomotives and all types of railcars (laden or empty), unless otherwise specified. Using the collected broken rail-caused freight-train derailment data, the distribution of the number of cars derailed is plotted in Figure 6.8.

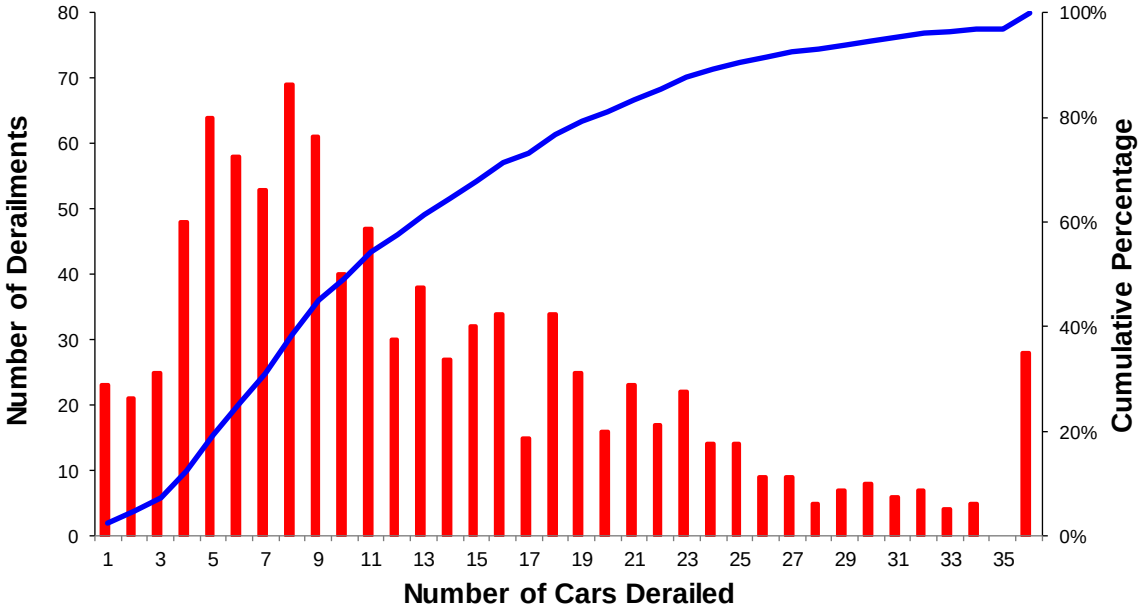


Figure 6. 8 Number of Cars (Railcars and Locomotives) Derailed per Broken Rail-Caused Freight-Train Derailment, Class I Railroad on Mainline, 2000 to 2017

6.3.1.2 Model Variables

The response variable here is the total number of cars derailed (e.g., locomotives, laden railcars and empty railcars) in one derailment. Several factors affect train derailment severity. Based on the review of the literature, the following predictor variables (Table 6.1) were identified for statistical analyses in this research. For example, train derailment speed is the speed of train operation when the accident occurs. The effect of this factor on derailment severity is the first and most widely studied in the literature (e.g., Saccomanno and El-Hage 1991; Anderson 2005; Bagheri et al. 2011; Liu et al. 2013b). It has been found that, given that all other factors are equal, derailment speed is positively associated with the number of cars derailed. This finding is reasonable, as speed is an indicator of a train's kinetic energy.

Table 6. 1 Predictor Variables in Severity Prediction Model

Variable Name	Definition	Type of Variable
TONS	Gross tonnage	Continuous
TRNSPD	Train derailment speed (MPH)	Continuous
CARS_TOTAL	Total number of cars	Continuous
CARS_LOADEDP	Proportion of loaded cars	Continuous
TRAINPOWER	Distribution of train power (distributed or non-distributed)	Categorical
WEATHER	Weather conditions (clear, cloudy, rain, fog, snow, etc.)	Categorical
TRKCLAS	FRA track class	Categorical
TRKDNSTY	Annual track density in MGT	Continuous

6.3.1.3 Decision Tree Model

The methodology developed in this research is built upon a machine learning algorithm called Decision Tree. A decision tree is a type of supervised learning algorithm that splits the population or sample into two or more homogeneous sets based on the most significant splitter / differentiator in input variables (Safavian and Landgrebe, 1991). It can cover both classification and regression problem in machine learning.

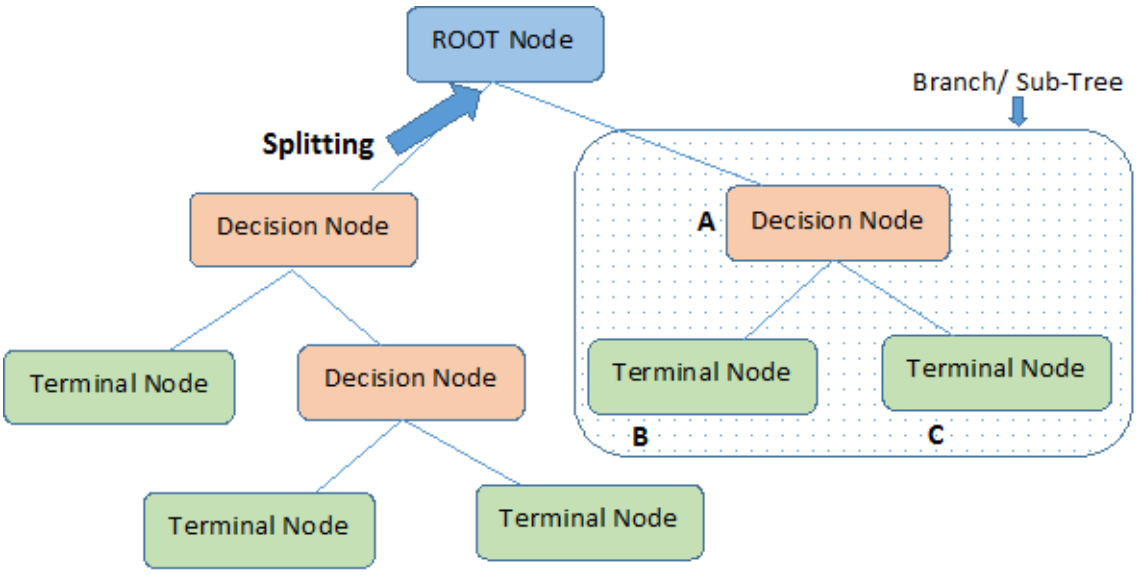


Figure 6. 9 Schematic Architecture of Decision Tree (Jain, 2017)

Figure 6.9 presents the structure of a simplified decision tree. Decision Node A is the parent node of Terminal Node B and Terminal Node C. In comparison with other regression methods and other advanced machine learning methods, decision tree has several advantages (Pal and Mather, 2003):

- It is easy to understand, interpret, and visualize.
- Decision trees implicitly perform variable screening or feature selection. They can quickly identify the most significant variables and relationships between two or more variables.
- They can handle both numerical and categorical data. They can also handle multi-output problems.
- Nonlinear relationships between parameters do not affect tree performance.
- It requires less data cleaning compared to some other modeling techniques. It is not influenced by outliers and missing values to a fair degree.

For example, compared to the Zero-Truncated Negative Binomial that was previously used in the study of train derailment severity (Liu et al., 2013b), which involves several assumptions (e.g., over-dispersed count variables), the decision tree method does not require these prerequisites but can still exclude the impacts from the nonlinear relationship between parameters. KNN (K-nearest neighbors algorithm) is one commonly used machine learning algorithm, but it can only be used in classification problems. Instead, decision tree is applicable for both continuous and categorical response variables. Random forests, gradient boosting, and artificial neural network are three other machine learning algorithms that have been widely used. In particular, random forests and gradient boosting are two algorithms based upon decision tree methods and aim to overcome some limitations in decision tree, such as overfitting. However, in this research, since only 938 broken rail-caused derailments are analyzed, the advantages of these advanced machine learning methods may not be significant due to relatively small sample size. In fact, the prediction accuracy of decision tree is comparable to other methods (random forests, gradient boosting, and artificial neural network) based on the data in this research. The preliminary testing results indicate that decision tree, random forests, gradient boosting, and artificial neural network all have similar prediction accuracy in terms of MSE (Mean Square Error) and MAE (Mean Absolute Error). Moreover, the features of decision tree, such as being simple to understand and visualize, and being a fast way to identify most significant variables, will be highlighted.

There are many specific algorithms to build a decision tree, such as CART (Classification and Regression Trees) using Gini Index as a metric, ID3 (Iterative Dichotomiser 3) using Entropy function and Information gain as metrics. Among these,

CART with Gini Index and ID3 with Information gain are commonly used. In this research, the development of a derailment severity prediction model is based upon the CART algorithm. The Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled, if it was randomly labeled according to the distribution of labels in the subset. The Gini impurity can be computed by summing the probability p_i of an item with label i being chosen, multiplied by the probability of wrongly categorizing that item $(1 - p_i)$. It reaches its minimum (zero) when all cases in the node fall into a single target category. To compute Gini impurity for a set of items with J classes, support i $\in \{1, 2, \ldots, J\}$, and let p_i be the fraction of items labeled with class i in the set.

$$I_G(p) = \sum_{i=1}^{J} p_i \sum_{k \neq i} p_k = \sum_{i=1}^{J} p_i (1 - p_i) = \sum_{i=1}^{J} (p_i - p_i^2) = \sum_{i=1}^{J} p_i - \sum_{i=1}^{J} p_i^2 = 1 - \sum_{i=1}^{J} p_i^2 \quad (6\text{-}3)$$

Where $I_G(p)$ is the Gini impurity; p_i is the probability of an item with label i being chosen; J is the classes of a set of items.

6.3.2 Model Development

First, the importance of each predictor in the database is identified and two measures of variable importance, Mean Decrease Accuracy (%IncMSE) and Mean Decrease Gini (IncNodePurity), are reported. Mean Decrease Accuracy (%IncMSE) is based upon the average decrease of prediction accuracy when a given variable is excluded from the model. Mean Decrease Gini (IncNodePurity) measures the quality of a split for every variable of a tree based on the Gini Index. For both measures, the higher value represents greater importance of a variable in the broken rail-caused train derailment

severity (Figure 6.10). Both metrics indicate that train speed (TRNSPD), number of cars in one train (CARS_TOTAL), and gross tonnage per train (TONS) are the three most significant variables impacting broken rail-caused train derailment severity.

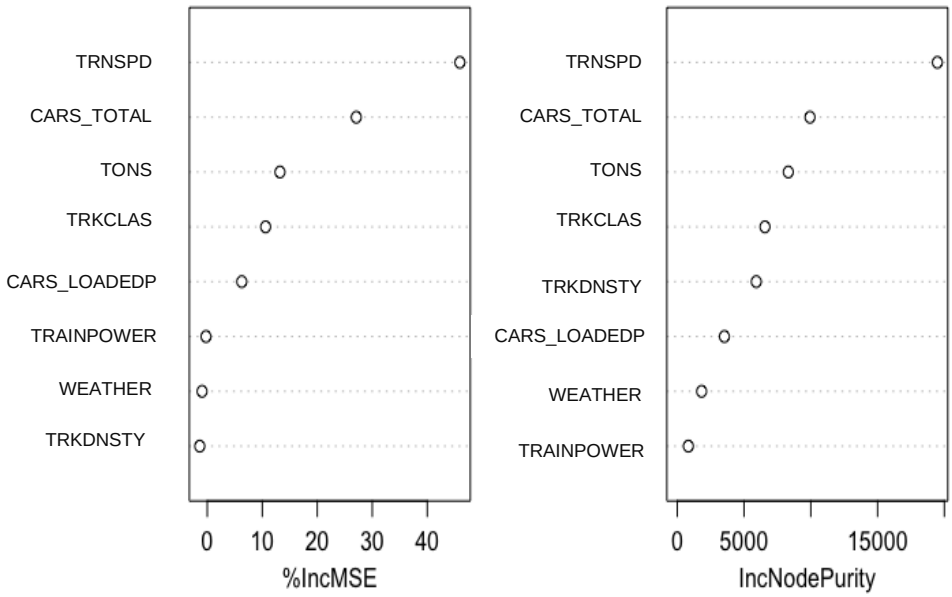


Figure 6. 10 Variable Importance for Train Derailment Severity Data

A decision tree has been developed for the training data (Figure 6.11). The response variable in the developed decision tree is the number of derailed cars. Three independent variables are employed in the built decision tree: TRNSPD (train derailment speed); CARS_TOTAL (number of cars in one train); and TONS (gross tonnage of a train). It indicates that these three factors have significant impacts on the freight-train derailment severity, in terms of number of cars derailed, while other variables (e.g., proportion of loaded cars, distribution of train power, weather condition, FRA track class, and annual track density) are statistically insignificant in the developed decision tree. For example, using the developed decision tree model, for a broken rail-caused freight-train derailment with a speed lower than 20 mph, the expected number of cars derailed is 7.5. Also, if a 100-

car freight train traveling at 30 mph derails due to broken rails, the expected number of cars derailed is 19.

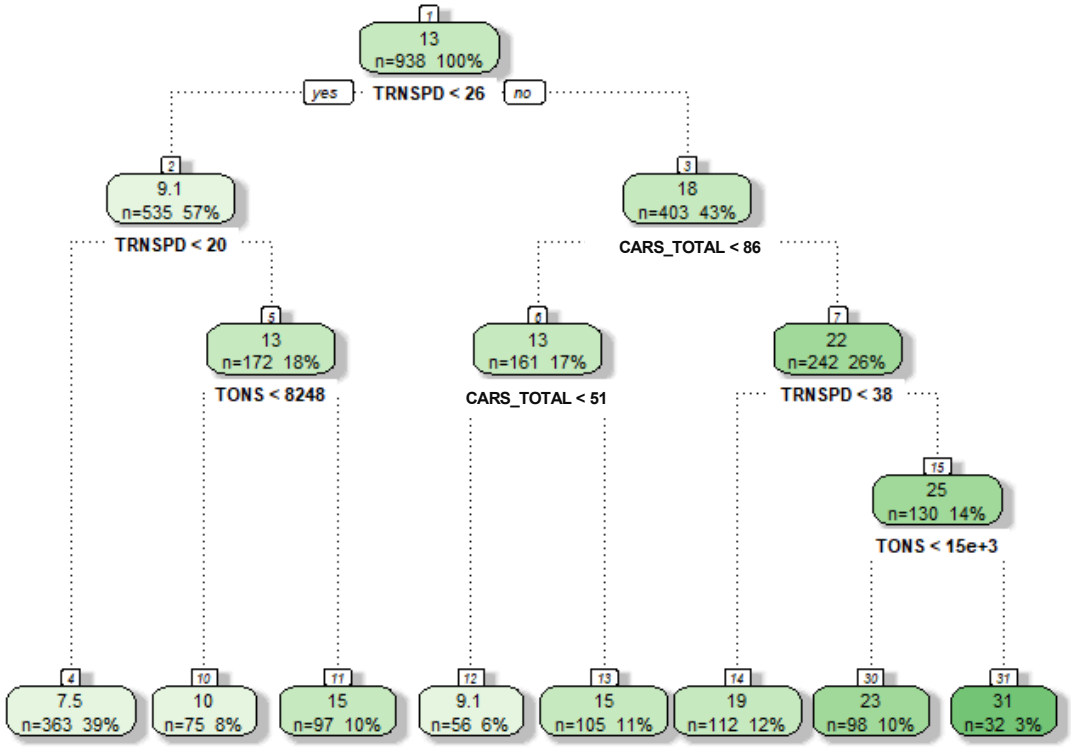


Figure 6. 11 Decision Tree in Broken Rail-Caused Derailment Severity Prediction[6]

To further validate the accuracy and practicability of the developed decision tree, selected broken rail-caused accidents of the studied Class I railroad in the last several years are listed in Table 6.2. The table lists the historical information of the accident, such as train speed (TRNSPD), gross tonnage (TONS), total number of cars in one train (CARS_TOTAL), number of derailed cars, as well as the estimated number of derailed cars via the decision tree model.

[6] TRNSPD: Train derailment speed (MPH); CARS_TOTAL: number of cars in one train; TONS: gross tonnage of a train.

Table 6. 2 Selected Broken Rail-Caused Derailments on the Studied Class I Railroad and Estimated Derailment Severity[7]

No	Gross tonnage (Tons)	Train speed (MPH)	Total number of cars in one train	**Observed number of derailed cars**	**Estimated number of derailed cars**
1	5,000	9	56	**6**	**7**
2	7,229	25	59	**6**	**10**
3	9,873	24	82	**21**	**15**
4	3,284	28	34	**14**	**15**
5	4,217	34	54	**22**	**15**
6	8,190	16	65	**12**	**7**
7	21,297	39	152	**31**	**31**
8	5,448	43	73	**23**	**15**
9	14,107	23	107	**17**	**15**
10	2,300	15	25	**4**	**7**
11	2,272	37	24	**11**	**9**
12	5,764	47	86	**29**	**23**
13	14,847	33	111	**27**	**19**
14	21,118	10	152	**9**	**7**
15	13,869	13	141	**11**	**7**
16	4,866	10	50	**8**	**7**
17	15,000	7	152	**13**	**7**

[7] All fields, excluding the estimated number of derailed cars in the last column using the decision tree model, are the observed values from the FRA REA database (FRA, 2018).

18	6,649	23	96	**2**	**10**
19	13,689	15	190	**15**	**7**
Average				**14.8**	**12.3**

6.4 Example Application of Broken-Rail Derailment Risk Model

In order to estimate the broken-rail derailment risk, calculation steps are illustrated in Figure 6.12:

- Step 1: Use broken rail prediction model to estimate the probability of broken rail $P(B)$.
- Step 2: Estimate the probability of broken-rail derailment given a broken rail $P(D|B)$, then calculate the probability of broken-rail derailment $P(DB)$.
- Step 3: Based on the decision tree model, estimate the severity of broken-rail derailment ($S(DB)$) given specific variables.
- Step 4: Calculate the broken-rail derailment risk $Risk(DB)$.

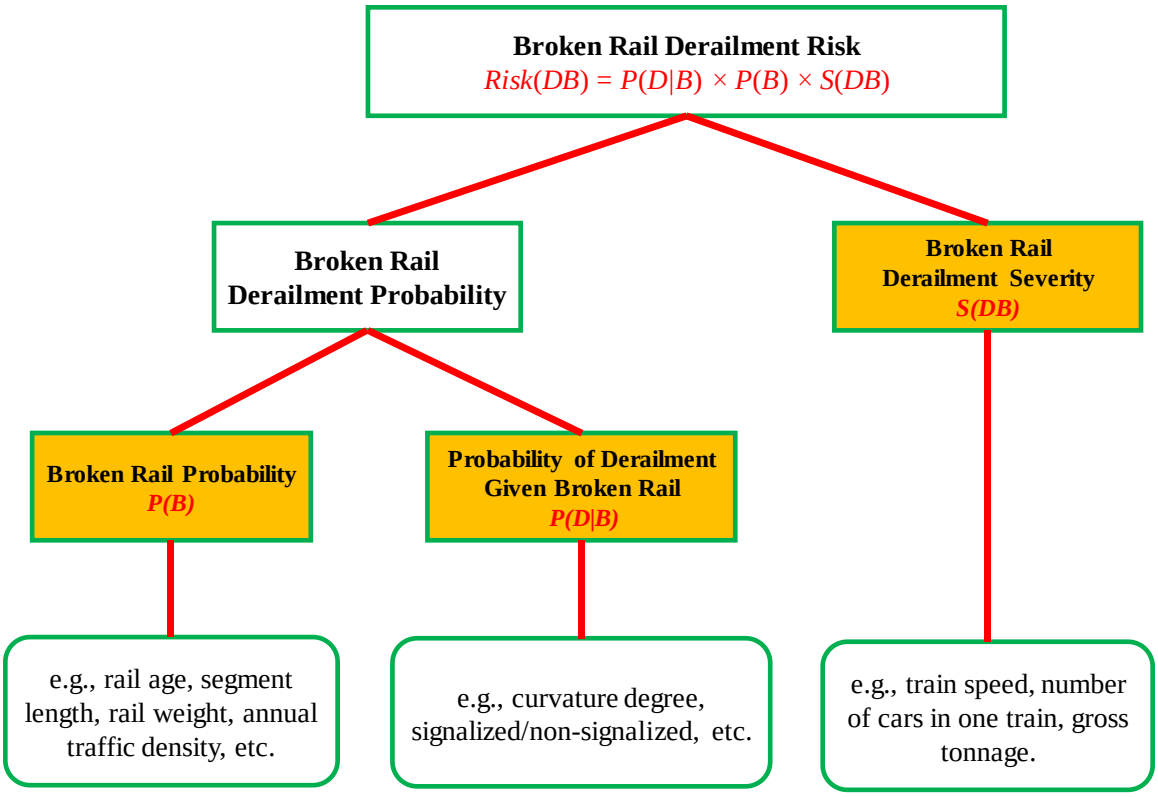


Figure 6. 12 Step-by-Step Broken Rail-Caused Derailment Risk Calculation

A step-by-step calculation example is used to illustrate the application of the broken rail derailment risk model. For illustrative convenience, a 0.2-mile signalized segment is used, with characteristics regarding rail age, traffic density, curve degree and others. More details of the example segment are summarized in Table 6.3. To calculate the severity given a broken-rail derailment on the segment, the train characteristics are also considered (Table 6.4).

Table 6. 3 Selected Characteristics of the Track Segment

Rail age (years)	23
Segment length (miles)	1
Rail weight (lbs/yard)	136

Annual traffic density (MGT)	30
Annual number of car passes	432,000
Curve degree	5.5
Speed	40 mph
Number of rail defects (all types) in last year	2
Number of broken rails in last year	1
Signalized/Non-signalized	Signalized
Presence of turnout	No

Table 6. 4 Train-Related Characteristics

Train operational speed (MPH)	40
Number of cars in one train	100
Gross tonnage	9,000

The broken rail-caused derailment risk can be calculated with the mentioned steps in this example:

- Step 1: Using the broken rail prediction model, the probability of broken rail on this track segment is estimated to be 0.015, $P(B) = 0.015$;
- Step 2: For curvature and signalized track segment, the estimated probability of derailment given a broken rail is 0.006, $P(D|B) = 0.006$. The estimated probability of broken-rail derailment on this particular track segment is calculated by $P(D|B) * P(B) = 0.006 * 0.015 = 0.00009$;

- Step 3: Use the decision tree model to estimate the average number of derailed cars per derailment on this track segment based on the given variables. The calculation procedure is illustrated in Figure 6.11. The estimated number of derailed cars is 23 given a broken-rail derailment on the track segment, with train speed 40 MPH, number of cars in one train 100, and gross tonnages 9,000;
- Step 4: The annual expected number of derailed cars is estimated to be $Risk(DB) = 0.00009 * 23 = 0.00207$.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Summary and Conclusion

This dissertation develops new models to predict broken-rail derailment risk using railroad big data. The risk model is built upon 1) prediction of broken rail probability; 2) estimation of the conditional probability of broken-rail derailment given a broken rail; and 3) derailment severity, measured by the number of cars derailed. Several research tasks were performed:

(1) A literature review was conducted regarding previous research on risk factors as well as data-driven models related to broken rail prediction. Knowledge gaps were identified.

(2) A novel network segmentation approach was developed based on important factors. Automated feature generation and selection techniques were also used to improve model accuracy and generalization.

(3) a novel, customized Soft Tile Coding based Neural Network model (STC-NN) was proposed to predict broken rail probability by location and time horizon.

The following conclusions were made based on this work:

(1) It is found that segment length, traffic tonnage, number of car passes, rail age, and the number of prior detected defects have a larger importance for the prediction of broken rails, based on the data used in this research.

(2) The proposed STC-NN model can be used to predict broken rail probability with good accuracy and flexibility. The algorithm demonstrates that in the colder season, the broken rail probability will be higher, given all else being equal. The proposed STC-NN is more accurate for the near-term prediction (e.g., one month or three months ahead). To predict broken rail occurrence within one month, the AUC of STC-NN is about 0.86, while other models (e.g., random forest, logistic regression, XGBoost, neural network, Cox regression) have the AUC below 0.78. Overall, the STC-NN outperforms these alternative machine learning algorithms in terms of one-month, three-month, and one-year data. The proposed STC-NN is able to "catch" over 71% of broken rails (weighted by segment length) by performing a risk-informed screening of 30% of network mileage with one month as prediction period.

(3) Signaled track in the colder period has the lowest ratio of broken rail-caused derailments to broken rails, while non-signaled track in warmer weather has the highest. This indicates that a more focused visual track inspection might be helpful in the warmer period for non-signaled territories if the goal is to reduce the chance of derailment per broken rail.

(4) In terms of derailment severity, a longer, heavier train traveling at a higher speed is associated with more cars derailed per broken rail-caused derailment.

7.2 Recommendations for Future Work

There are several research areas that the dissertation does not cover, which will leave them for future research. First, by no means the data used in this research is comprehensive and exhaustive. Certain factors (e.g., tie condition, weather condition, human operation) may also be important but the corresponding information is not available yet in this study. In the future, additional variables could be collected to examine whether the model can be further improved. Second, due to data limits, derailment severity model may not be generalized. More derailment data from other railroads should be collected to understand whether there is railroad-specific difference in terms of broken-rail derailment severity. Third, machine learning algorithm allows the efficient utilization of multi-dimensional and multi-variety information resources for the prediction. However, as the major challenge of the machine learning algorithm, it is difficult to interpret. How to interpret the casual effects, causal interaction in high-dimension and the application of the machine learning for decision making will be accounted for in the future research.

APPENDIX A

NOMENCLATURES FOR DATA SUMMARY

Table A. 1 Description of Signal Code

Signal Code	Description
CP	Control point signal
YL-S	Main track yard limits signaled
TWC-ABS	Track warrant control with automatic block signals
COT	Current of traffic-track signaled in one direction
TC	Train control
CP/CSS	Cab signal system
311	Railroad crossings at grade

Table A. 2 Nomenclatures for Rail Defect Type Code

Abbreviation	Description
TDD	Detail Fracture
TW	Defective Field Weld
SSC	Shelling/Spalling/Corrugation
EFBW	In-Track Electric Flash Butt Weld
SD	Shelly Spots
EBF	Engine Burn Fracture

BHB	Bolt Hole Crack
HW	Head Web
HSH	Horizontal Split Head
VSH	Vertical Split Head
EB	Engine Burn – (Not Fractured)
OAW	Defective Plant Weld
FH	Flattened Head
CH	Crushed Head
SW	Split Web
SDZ	Shelly Spots in Dead Zones of Switch
TDT	Transverse Fissure
TDC	Compound Fissure
LER	Loss of Expected Response-Loss of Ultrasonic Signal
BRO	Broken Rail Outside Joint Bar Limits
DWL	Separation Defective Field Weld (Longitudinal)
BB	Broken Base
PIPE	Piped Rail
DR	Damaged Rail

Figure A. 1 Illustration of Top 10 Types of Broken Rails[8]

Table A. 3 Nomenclatures for Geometry Track Exception Type

Subgroup	Geometry Track Exception Type
CROSS-LEVEL/CLIM[9]	CROSS-LEVEL
	CLIM
GAGE	WIDE GAGE
	PLG 24 1ST LEVEL

[8] The picture source is FRA Track Inspector Rail Defect Reference Manual (FRA, 2015).
[9] Cross Level Index Meter

	PLG 24 2ND LEVEL
	GWP 1ST LEVEL
	GWP 2ND LEVEL
	LOADED GAGE
	TIGHT GAGE
CANT	LEFT RAIL CANT
	RIGHT RAIL CANT
	CONC LT RAIL CANT
	CONC RT RAIL CANT
ALIGNMENT	ALIGNMENT LEFT
	ALIGNMENT RIGHT
	ALIGNMENT
	ALIGNMENT LFET 31 FT
	ALIGNMENT RIGHT 31 FT
WARP 31	WARP 31FT
WARP 62	WARP 62 FT
	WARP 62 FT>6IN XLV
SPEED/ELEVATION	EXCESS. ELEVATION
	CURVE SPEED 3IN
	CURVE SPEED 4IN
	RUN OFF LEFT
	RUN OFF RIGHT
	RIGHT VERT ACC

PROFILE/SURFACE	PROFILE RIGHT 62 FT
	PROFILE LEFT 62 FT
	UNBALANCE 4IN
	UNBALANCE 3IN

APPENDIX B

AGGREGATION FUNCTION FOR MERGING SIDES OF TRACK

Table B. 1 Aggregation Functions for Merging the Rail Information on the Same Track

Attribute	Description	Preferred Value
Division	Division information	Either one
Subdivision	Subdivision information	Either one
Prefix	A 3-alphabet coding system as route identifiers	Either one
Track_type	Single track or multiple tracks (SG, track 1, track 2, track 3, track 4)	Either one
Rail_laid _year	The year when the rail was laid	Minimum
Rail_weight	Rail weight measured as pounds per yard	Minimum
Rail_quality	Two categories: new rail and re-laid rail	Worse case
Curve_degree	The curve degree posted at the location	Either one
Spiral_1	The spiral length (feet) at the beginning of the curve	Either one
Spiral_2	The spiral length (feet) at the ending of the curve	Either one

Super-elevation	Super-elevation between two rails on the curve	Either one
Grade_percent	The feet of rise per 100 feet of horizontal distance	Either one
Speed	The maximum allowed speed (mph) at the location	Either one
Signal	Whether track circuits were set at the location (yes or no)	Either one
Turnout_num	Total number of turnouts posted at the location	Either one
Ballast_*time*	The total number of ballast cleaning at the location in the particular time period	Either one
Grinding_*time*	The total number of grinding passes at the location in the particular time period	Mean
Service_failure_*time*	The total number of service failure (broken rails) (including all types) occurred at the location in the particular time period	Sum
Car_passes_*time*	The number of cars passing at the location in the particular time period	Mean
Tonnages_*time*	The gross million tonnages (MGT) experienced at the location in the particular time period	Mean

Defect_*type_time*	The total number of rail defects with specific type at the location in the particular time period	Sum
Geometry_*type_time*	The total number of geometry exception defects with specific type at the location in the particular time period	Sum
Geometry_*time*	The total number of geometry exception defects (including all types) at the location in the particular time period	Sum
Geometry_*priority_time*	The total number of geometry exception defects with the specific priority in the particular time. Geometry exceptions are automatically prioritized based on the deviation of the measure from the class of track being measured.	Sum
Class reduced_*time*	Class reduction due to geometry exceptions in the particular time period. It is calculated by the difference between the original track class and the updated track class.	Maximum
VTI_*type_time*	The total number of vehicle-track interaction exceptions with the specific type in the particular time period	Sum

Measure_VTI_*type_time*	The max measurements corresponding to different vehicle-track interaction exception types in the particular time	Maximum
VTI_priority_*time*	The total number of vehicle-track interaction exceptions with specific priority in particular time period.	Mean

APPENDIX C

BROKEN RAIL-CAUSED DERAILMENT SEVERITY ESTIMATION WITH ALTERNATIVE MODELS

C.1 Zero-Truncated Negative Binomial (ZTNB) Model

Zero Truncated Negative Binomial (ZTNB) model is one popular count data regression models used in accident analysis. This model assumes that the Poisson mean follows a gamma distribution and has been used for analyzing over-dispersed data, in which the variance is greater than mean. Moreover, traditional Negative Binomial model can analyze data with zero counts but the number of cars derailed per derailment accident represents non-negative count data. Therefore, Zero Truncated Negative Binomial (ZTNB) model can calculate the probability of response variable based on positive count data using Bayes's Theorem and account for the exclusion of zeros. The response surface of a Zero Truncated Negative Binomial (ZTNB) model is shown below. A detailed discussion of the ZTNB model can be found in Grogger and Carson (1991):

$$\log(\mu_i) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} \quad \text{(C-1)}$$

Where μ_i is the estimated derailment severity for the ith observation, β_k is the parameter coefficient of the kth predictor variable (k = 0 for intercept), X_{ki} is the value of the kth predictor variable for the ith observation.

After removing insignificant independent variables using P-values as the measure, a Zero Truncated Negative Binomial (ZTNB) model is developed based on a set of 938 broken rail-caused freight-train derailments from 2000 to 2017. For example, below model accounts for main effect only.

$$\log(\mu) = 1.16 + 0.27\text{TRNSPD} + (6.15 \times 10^{-6})\text{TONS} + (6.63 \times 10^{-3})\text{cars\_total}$$

where μ is the estimated derailment severity.

C.2 Artificial Neural Network

Artificial Neural Network is another main tool in machine learning. It is a brain-inspired system which is intended to replicate the way that humans learn. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that transform the input into something that the output layer can use. They are excellent tools for finding patterns which are far too complex or numerous for a human programmer to extract and teach the machine to recognize. The output of the entire network, as a response to an input vector, is generated by applying certain arithmetic operations, determined by the neural networks. In the prediction of broken rail-caused derailment severity, the neural network can use a finite number of past observations as training data and then make predictions for testing data.

The prediction accuracy of these four models, which are Zero-Truncated Negative Binomial, random forest, gradient boosting, and artificial neural network, are presented in the below table. MSE (Mean Square Error) and MAE (Mean Absolute Error) are employed as two metrics.

Table C. 1 Prediction Accuracy of Alternative Models

Prediction Models		**MSE**	**MAE**
Zero-Truncated Negative Binomial	Main component only	52.65	5.21
	With second order and interaction	52.59	4.99
Decision tree		48.55	4.91
Random forest		48.30	4.89
Gradient boosting		52.50	5.00
Artificial Neural Network		55.68	5.23

REFERENCES

[1] Aglan, H., & Gan, Y. X. (2001). Fatigue crack growth analysis of a premium rail steel. *Journal of Materials Science*, *36*(2), 389-397.

[2] Ahlbeck, D. R. (1980). *An investigation of impact loads due to wheel flats and rail joints* (No. 80-WA/RT-1 Conf Paper).

[3] An, R., Sun, Q., Bai, W., Liu, R. & Wang, F. (2017). Grid-Based Analysis of causal factors of railway rail break risk. *Transportation Research Record: Journal of the Transportation Research Board*.

[4] Anderson, R.T., 2005. Quantitative analysis of factors affecting railroad accident probability and severity (Doctoral dissertation, University of Illinois at Urbana-Champaign).

[5] Ashley, J. (2008). Giving Dirty Ballast the Cold Shoulder. *International Railway Journal*, *48*(8).

[6] Association of American Railroad (AAR) (2018). *State of the industry reports/ Report 2: Economic Impact*.

[7] Arthur D. Little, Inc. (ADL) (1996). Risk Assessment for the Transportation of Hazardous Materials by Rail, Supplementary Report: Railroad Accident Rate and Risk Reduction Option Effectiveness Analysis and Data, 2nd rev. *ADL, Cambridge, Mass*.

[8] Bagheri, M., Saccomanno, F., Chenouri, S. and Fu, L., 2011. Reducing the threat of in-transit derailments involving dangerous goods through effective placement along the train consist. *Accident Analysis & Prevention, 43(3),* pp.613-620.

[9] Bai, L., Liu, R., Sun, Q., Wang, F., & Xu, P. (2015). Markov-based model for the prediction of railway track irregularities. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit*, *229*(2), 150-159.

[10] Bai, L., Liu, R., Wang, F., Sun, Q., & Wang, F. (2017). Estimating railway rail service life: A rail-grid-based approach. *Transportation Research Part A: Policy and Practice*, *105*, 54-65.

[11] Barkan, C.P.L., Dick, C.T., Anderson, R.T., 2003. Analysis of railroad derailment factors affecting hazardous materials transportation risk. *Transportation ResearchRecord 1825,* 64–74.

[12] Birk, A. M., Anderson, R. J., & Coppens, A. J. (1990). A computer simulation of a derailment accident: Part I-model basis. *Journal of Hazardous Materials, 25(1-2),* 121-147.

[13] Brouzoulis, J. (2014). Wear impact on rolling contact fatigue crack growth in rails. *Wear*, *314*(1-2), 13-19.

[14] Buckman, J., Roy, A., Raffel, C., & Goodfellow, I. (2018). *Thermometer encoding: One hot way to resist adversarial examples*.

[15] Burstow, M. C., Watson, A. S., & Beagles, M. (2002). Application of the whole life rail model to control rolling contact fatigue. *In Proceedings of the International Conference Railway Engineering 2002*, London, UK.

[16] Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, 70-79.

[17] Cannon, D. F., Edel, K. O., Grassie, S. L., & Sawley, K. (2003). Rail defects: an overview. *Fatigue & Fracture of Engineering Materials & Structures*, *26*(10), 865-886.

[18] Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, *32(5)*, 361-378.

[19] Chattopadhyay, G., & Kumar, S. (2009). Parameter Estimation for Rail Degradation Model. *International Journal of Performability Engineering*, *5*(2), 119-130.

[20] Chattopadhyay, G., Larsson, P. O., & Reddy, V. (2003). Mathematical modelling for optimal rail grinding decisions in maintenance of rails. In Shrivastav et al, Om (Ed.) *Condition Monitoring and Diagnostic Engineering Management - Proceedings of the 16th International Congress*, 27-9 August 2003, Vaxjo, Sweden.

[21] Chattopadhyay, G., Reddy, V., & Larsson-Kråik, P. O. (2005). Decision on economical rail grinding interval for controlling rolling contact fatigue. *International Transactions in Operational Research*, *12*(6), 545-558.

[22] Clayton, P. (1996). Tribological aspects of wheel-rail contact: a review of recent experimental research. *Wear*, *191*(1-2), 170-183.

[23] Coppens, A.J., Wong, J.D.E., Bibby, A., Birk, A.M., Anderson, R.J. (1988). Development of a Derailment Accident Computer Simulation Model. *Transport Canada Report No. TP 9254E.* Prepared for the Transportation Development Centre and Transport of Dangerous Goods, Ottawa, Canada.

[24] Corbin, J. C., & Fazio, A. E. (1981). Performance-based track-quality measures and their application to maintenance-of-way planning. *Transportation Research Record*, (802).

[25] Cuervo, P. A., Santa, J. F., & Toro, A. (2015). Correlations between wear mechanisms and rail grinding operations in a commercial railroad. *Tribology International*, *82*, 265-273.

[26] da Silva, L. F., Stewardson, D. J., De Oliveira, F. M. F., & De Castro, P. M. S. T. (2003). Fatigue crack growth of rails for railways. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *217*(2), 89-97.

[27] Dick, C. T. (2001). *Factors affecting the frequency and location of broken railway rails and broken rail derailments* (Doctoral dissertation, University of Illinois at Urbana-Champaign).

[28] Dick, C. T., Barkan, C. P.L., Chapman, E., & Stehly, M. P. (2002). Predicting the occurrence of broken rails: a quantitative approach. *In Proceedings of the American Railway Engineering and Maintenance of Way Association (AREMA) Annual Conference, Washington DC*.

[29] Dick, C.T., Barkan, C.P.L., Chapman, E., & Stehly, M.P. (2003). Multivariate statistical model for predicting occurrence and location of broken rails. *Transportation Research Record: Journal of the Transportation Research Board*, (1825), 48-55.

[30] Esveld, C. (2001). Modern Railway Track, 2nd Editon. *Delft university of Technology*.

[31] Farris, T. N. (1996). Effect of overlapping wheel passages on residual stress in rail corners. *Wear*, *191*(1-2), 226-236.

[32] Federal Railroad Administration (FRA) (2015), Track Inspector Rail Defect Reference Manual: Reversion 2, *Federal Railroad Administration*, 2015 July.

[33] Fischer, F. D., Daves, W., Pippan, R., & Pointner, P. (2006). Some comments on surface cracks in rails. *Fatigue & Fracture of Engineering Materials & Structures*, *29*(11), 938-948.

[34] FRA Rail. (2017). *Interactive Rail Defect Remediation Table.* Retrieved from https://frarail.wordpress.com.

[35] FRA Rail Equipment Accident (6180.54) database. https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx. (Accessed in November 1, 2017).

[36] Federal Railroad Administration (FRA), 2018. FRA Rail Equipment Accident (6180.54) database. https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx. Accessed April 21, 2018.

[37] Garnham, J. E., & Beynon, J. H. (1992). Dry rolling-sliding wear of bainitic and pearlitic steels. *Wear*, *157*(1), 81-109.

[38] Grogger, J.T., Carson, R.T., 1991. Models for truncated counts. *Journal of Applied Econometrics 6,* 225–238.

[39] Hay, W. W. (1982). *Railroad engineering* (Vol. 1). John Wiley & Sons.

[40] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, *21(9)*, 1263-1284.

[41] He, Q., Li, H., Bhattacharjya, D., Parikh, D. P., & Hampapur, A. (2013). Railway track geometry defect modeling: deterioration, derailment risk and optimal repair. In *Proceedings of the Transportation Research Board Annual Meeting*. *The Academy of Transportation Research Board*.

[42] He, Q., Li, H., Bhattacharjya, D., Parikh, D. P., & Hampapur, A. (2015). Track geometry defect rectification based on track deterioration modelling and derailment risk assessment. *Journal of the Operational Research Society*, *66*(3), 392-404.

[43] Hokstad, P., Langseth, H., Lindqvist, B. H., & Vatn, J. (2005). Failure modeling and maintenance optimization for a railway line. *International Journal of Performability Engineering*, *1*(1), 51.

[44] International Heavy Haul Association. (IHHA, 2001). Guidelines to best practices for heavy haul railway operations: Wheel and rail interface issues. *International Heavy Haul Association.*

[45] Jablonski, D. A., & Pelloux, R. M. (1992). Effect of train load spectra on crack growth in rail steel. *Residual Stress in Rails. Effects on Rail Integrity and Railroad Economics. Kluwer Academic Publ., Dordrecht, The Netherlands*, *1*, 81-98.

[46] Jain, R., 2017. Decision Tree. It Begins Here. https://medium.com/@rishabhjain_22692/decision-trees-it-begins-here-93ff54ef134

[47] Jenkins, H. H., Stephenson, J. E., Clayton, G. A., Morland, G. W., & Lyon, D. (1974). The effect of track and vehicle parameters on wheel/rail vertical dynamic forces. *Railway Engineering Journal*, *3*(1).

[48] Jeong, D. Y., & Gordon, J. E. (2009). Evaluation of rail test frequencies using risk analysis. In *2009 Joint Rail Conference* (pp. 23-30). American Society of Mechanical Engineers.

[49] Jeong, D. Y., Tang, Y. H., & Orringer, O. (1997). Damage tolerance analysis of detail fractures in rail. *Theoretical and Applied Fracture Mechanics*, *28*(2), 109-115.

[50] Jeong, D. Y., Tang, Y. H., & Orringer, O. (1998). *Estimation of rail wear limits based on rail strength investigations* (No. DOT-VNTSC-FRA-98-13). United States. Federal Railroad Administration.

[51] Jeong, D. Y. (2001). *Progress in rail integrity research* (No. DOT/FRA/ORD-01/18). Federal Railroad Administration. United States.

[52] Jovanovic, S. (2004, October). Railway track quality assessment and related decision making. *In 2004 IEEE International Conference on Systems, Man and Cybernetics* (IEEE Cat. No. 04CH37583) (Vol. 6, pp. 5038-5043). IEEE.

[53] Judge, T. (2000). Finding the right profile. *Railway Age*, *201*(12).

[54] Kalousek, J., & Magel, E. (1997). Achieving a balance: the magic wear rate. *Railway Track and Structures*.

[55] Kalousek, J., Sroba, P., & Hegelund, C. (1989). Analysis of rail grinding tests and implications for corrective and preventative grinding. In *Fourth International Heavy Haul Railway Conference 1989: Railways in Action; Preprints of Papers, The* (p. 193). Institution of Engineers, Australia.

[56] Kassa, E., Andersson, C., & Nielsen, J. C. O. (2006). Simulation of dynamic interaction between train and railway turnout. *Vehicle System Dynamics*, *44*(3), 247-258.

[57] Kassa, E., & Nielsen, J. C. O. (2008). Stochastic analysis of dynamic interaction between train and railway turnout. *Vehicle System Dynamics*, *46*(5), 429-449.

[58] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems* (pp. 3146-3154).

[59] Kim, J. K., & Kim, C. S. (2002). Fatigue crack growth behavior of rail steel under mode I and mixed mode loadings. *Materials Science and Engineering: A*, *338(1-2)*, 191-201.

[60] Kumar, S. (2006a). *Study of rail breaks: associated risks and maintenance strategies*. Luleå tekniska universitet.

[61] Kumar, S. (2006b). *A study of the rail degradation process to predict rail breaks* (Doctoral dissertation, Luleå tekniska universitet).

[62] Lessmann, S., Baesens, B., Mues, C., & Pietsch, S. (2008). Benchmarking classification models for software defect prediction: A proposed framework and novel findings. *IEEE Transactions on Software Engineering, 34(4),* 485-496.

[63] Lichtberger, B. (2005). Track compendium. *Eurailpress Tetzlaff-Hestra GmbH & Co. KG, Hamburg*, *634*.

[64] Liu, X., Dick, C. T., Lovett, A., Saat, M. R., & Barkan, C. P. (2013a). Seasonal effect on the optimization of rail defect inspection frequency. *In ASME 2013 Rail Transportation Division Fall Technical Conference* (pp. V001T01A008-V001T01A008). American Society of Mechanical Engineers.

[65] Liu, X., Saat, M.R., Barkan, C.P.L. (2012). Analysis of causes of major train derailment and their effect on accident rates. *Transportation Research Record 2289,*154–163.

[66] Liu, X., Saat, M. R., Qin, X., & Barkan, C. P. (2013b). Analysis of US freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. *Accident Analysis & Prevention, 59,* 87-93.

[67] Liu, X., Saat, M. R., & Barkan, C. P. (2017). Freight-train derailment rates for railroad safety and risk analysis. *Accident Analysis & Prevention, 98*, 1-9.

[68] Magel, E. E., & Kalousek, J. (2002). The application of contact mechanics to rail profile design and rail grinding. *Wear, 253(1-2),* 308-316.

[69] Martey, E.N. and Attoh-Okine, N. (2018). Bivariate Severity Analysis of Train Derailments using Copula-Based Regression Models. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 4(4),* p.04018034.

[70] Mohammadi, R., He, Q., Ghofrani, F., Pathak, A., & Aref, A. (2019). Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects. *Transportation Research Part C: Emerging Technologies, 102*, 153-172.

[71] Muster, H., Schmedders, H., Wick, K., & Pradier, H. (1996). Rail rolling contact fatigue. The performance of naturally hard and head-hardened rails in track. *Wear, 191*(1-2), 54-64.

[72] Olofsson, U., & Nilsson, R. (2002). Surface cracks and wear of rail: a full-scale test on a commuter train track. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 216*(4), 249-264.

[73] Orringer, O. (1990). Control of rail integrity by self-adaptive scheduling of rail tests (No. DOT/TSC/FRA-90-02). *United States. Federal Railroad Administration*.

[74] Orringer, O., Tang, Y. H., Gordon, J. E., Jeong, D. Y., Morris, J. M., & Perlman, A. B. (1988). *Crack propagation life of detail fractures in rails* (No. DOT/FRA/ORD-88/13). United States. Federal Railroad Administration.

[75] Pal, M. and Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment*, *86*(4), pp.554-565.

[76] Rahman, F., & Devanbu, P. (2013, May). How, and why, process metrics are better. *In 2013 35th International Conference on Software Engineering (ICSE)* (pp. 432-441). IEEE.

[77] Reddy, V. (2004). *Modelling and analysis of rail grinding and lubrication strategies for controlling rolling contact fatigue (RCF) and rail wear* (Doctoral dissertation, Queensland University of Technology).

[78] Reiff, R.P., 1997. Proceedings of Rail Defect and Broken Rail Defects Expanded Workshop, *Transportation Technology Center, Pueblo, Colorado,* July 1997.

[79] Roney, M. D., & Ebersohn, W. (2001). Maintaining Optimal Wheel and Rail Performance, Guidelines to Best Practices for Heavy Haul Railway Operations: Wheel and Rail Interface Issues. *International Heavy Haul Association, Virginia, USA*.

[80] Saccomanno, F. F., & El-Hage, S. (1989). Minimizing derailments of railcars carrying dangerous commodities through effective marshaling strategies. *Transportation Research Record*, *1245(34-51)*, 39-41.

[81] Saccomanno, F.F. and El-Hage, S.M. (1991). Establishing derailment profiles by position for corridor shipments of dangerous goods. *Canadian Journal of Civil Engineering, 18(1),* pp.67-75.

[82] Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics, 21(3),* 660-674.

[83] Schafer, D. H. (2008). *Effect of train length on railroad accidents and a quantitative analysis of factors affecting broken rails* (Doctoral dissertation, University of Illinois at Urbana-Champaign).

[84] Schafer, D. H., & Barkan, C. P. L. (2008a). A hybrid logistic regression/neural network model for the prediction of broken rails. In *Proceedings of the 8th World Congress on Railway Research, Seoul, Korea*.

[85] Schafer, D. H., & Barkan, C. P. (2008b). A prediction model for broken rails and an analysis of their economic impact. In *Proceedings of American Railway Engineering and Maintenance of Way Association (AREMA) Annual Conference*.

[86] Schupp, G., Weidemann, C., & Mauer, L. (2004). Modelling the contact between wheel and rail within multibody system simulation. *Vehicle System Dynamics*, *41*(5), 349-364.

[87] Sebes, M., Ayasse, J. B., Chollet, H., Pouligny, P., & Pirat, B. (2006). Application of a semi-Hertzian method to the simulation of vehicles in high-speed switches. *Vehicle System Dynamics*, *44*(sup1), 341-348.

[88] Sherstov, A. A., & Stone, P. (2005, July). Function approximation via tile coding: Automating parameter choice. *In International Symposium on Abstraction, Reformulation, and Approximation (pp. 194-205)*. Springer, Berlin, Heidelberg.

[89] Shyr, F. Y. (1993). *Combining laboratory and field data in rail fatigue analysis* (Doctoral dissertation, Massachusetts Institute of Technology).

[90] Shyr, F. Y., & Ben-Akiva, M. (1996). Modeling rail fatigue behavior with multiple hazards. *Journal of Infrastructure Systems*, *2*(2), 73-82.

[91] Skyttebol, A., Josefson, B. L., & Ringsberg, J. W. (2005). Fatigue crack growth in a welded rail under the influence of residual stresses. *Engineering Fracture Mechanics*, *72*(2), 271-285.

[92] Soeleiman, S., & Rucinski, J. (1991). Rail Grinding Model for ROA. *BHP Research Melb. Lab. Report No. BHPRML/CM11/91/001, BHP Melbourne Research Lab., Melbourne, Australia*.

[93] Sourget, F., & Riollet, A. M. (2006). PROBARAIL: a statistical tool to improve preventive maintenance on rails. *In Proceedings of the 7th World Congress on Railway Research*.

[94] Stock, R., & Pippan, R. (2011). RCF and wear in theory and practice—The influence of rail grade on wear and RCF. *Wear*, *271*(1-2), 125-133.

[95] Sun, Y. Q., Cole, C., & Boyd, P. (2011). A numerical method using VAMPIRE modelling for prediction of turnout curve wheel–rail wear. *Wear*, *271*(1-2), 482-491.

[96] Sun, M., Wang, Y., Zhang, X., Liu, Y., Wei, Q., Shen, Y., & Feng, N. (2014, May). Feature selection and classification algorithm for non-destructive detecting of high-speed rail defects based on vibration signals. *In 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings (pp. 819-823). IEEE*.

[97] Tantithamthavorn, C., McIntosh, S., Hassan, A. E., & Matsumoto, K. (2016, May). Automated parameter optimization of classification techniques for defect prediction models. *In 2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE) (pp. 321-332). IEEE*.

[98] Thelen, G., & Lovette, M. (1996). A parametric study of the lubrication transport mechanism at the rail-wheel interface. *Wear*, *191*(1-2), 113-120.

[99] Trawiński, B., Smętek, M., Telec, Z., & Lasota, T. (2012). Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. *International Journal of Applied Mathematics and Computer Science*, *22*(4), 867-881.

[100] Vale, C., & Calçada, R. (2013). A dynamic vehicle-track interaction model for predicting the track degradation process. *Journal of Infrastructure Systems, 20(3),* 04014016.

[101] van den Bosch, R. A. (2002). Rail Grinding Strategies on Netherlands Railways. *Rail Engineering International*, *31*(1).

[102] Vesković, S., Tepić, J., Ivić, M., Stojić, G., & Milinković, S. (2012). Model for predicting the frequency of broken rails. *Metalurgija-Zagreb*, *51*(2), 221.

[103] Xu, J., Wang, P., Wang, L., & Chen, R. (2016). Effects of profile wear on wheel–rail contact conditions and dynamic interaction of vehicle and turnout. *Advances in mechanical Engineering*, *8*(1), 1687814015623696.

[104] Yang, T. H., Manos, W. P., & Johnstone, B. (1972a). A Study Continuation of Derailment Behavior Final Report (Phase 08 Report on Computer Derailment Study). RPI/AAR Report RA-08-1-12 (R-135). *Railroad Tank Car Safety Research and Test Project. Association of American Railroads, Washington, DC*.

[105] Yang, T. H., Manos, W. P., & Johnstone, B. (1972b). *Dynamic analysis of train derailments (No. 72-WA/RT-6 Paper)*.

[106] Zarembski, A. M. (2005). *The art and science of rail grinding*. Simmons-Boardman Books.

[107] Zarembski, A. (2009). Guidelines for broken-rail risk. *Railway Track and Structures*, *105*(10).

[108] Zarembski, A. M., & Bonaventura, C. S. (2010). Dynamic effects of track surface condition on vertical wheel/rail forces and energy consumption. *In 2010 Joint Rail Conference*. *American Society of Mechanical Engineers*, *1-6*.

[109] Zarembski, A. M., Einbinder, D., & Attoh-Okine, N. (2016). Using multiple adaptive regression to address the impact of track geometry on development of rail defects. *Construction and Building Materials*, *127*, 546-555.

[110] Zarembski, A. M., & Palese, J. W. (2010). Evaluation of the effectiveness of rail grinding on reducing rail defects on North American Class 1 railroad. *In Annual Conference of American Railway Engineering and Maintenance-of-Way Association, Orlando, Fla*.

[111] Zarembski, A. M., Palese, J. W., & Euston, T. (2005). Monitoring grinding effectiveness. *Railway Track and Structures*, *101*(6).

[112] Zerbst, U., Lundén, R., Edel, K. O., & Smith, R. A. (2009a). Introduction to the damage tolerance behaviour of railway rails–a review. *Engineering Fracture Mechanics*, *76*(17), 2563-2601.

[113] Zerbst, U., Schödel, M., & Heyder, R. (2009b). Damage tolerance investigations on rails. *Engineering Fracture Mechanics*, *76*(17), 2637-2653.

[114] Zhao, J., Chan, A. H. C., & Stirling, A. B. (2006). Risk analysis of derailment induced by rail breaks-a probabilistic approach. *In Reliability and Maintainability Symposium, 2006. RAMS'06. Annual*, IEEE, 486-491.

[115] Zhao, J., Chan, A., & Burrow, M. (2007a). Probabilistic model for predicting rail breaks and controlling risk of derailment. *Transportation Research Record: Journal of the Transportation Research Board*, (1995), 76-83.

[116] Zhao, J., Chan, A. H. C., Roberts, C., & Madelin, K. B. (2007b). Reliability evaluation and optimisation of imperfect inspections for a component with multi-defects. *Reliability Engineering and System Safety*, *92*(1), 65-73.

[117] Zong, N., Wexler, D., & Dhanasekar, M. (2013). Structural and material characterization of insulated rail joints. *Special Issue: Electronic Journal of Structural Engineering*, *13(1)*, 75-87.