Development and Evaluation of Interpretable Machine Learning Models for Mitigating Winter Road Safety – An Empirical Investigation

by

Zehua Shuai

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

TRANSPORTATION ENGINEERING

Department of Civil and Environmental Engineering
University of Alberta

ABSTRACT

In Canada, winter crashes account for a significant portion of crashes each year. This thesis investigates the utility of machine learning (ML) for understanding and mitigating winter road risks. Despite their potential to achieve high predictive performance in the face of complex data structures, ML models often lack interpretability as their predictions lack sufficient explanations for users to validate. Numerous studies have utilized ML models in traffic safety modeling; however, the transparency of the model predictions remains a challenge. To address this problem, this study developed highly interpretable ML models with the assistance of explainable artificial intelligence (XAI), which helps determine feature contributions and directionality towards the ML predictions. The first part of this thesis analyzed the characteristics of snowstorm-related crashes, while the second part developed a winter crash frequency model. Both high-complexity ML models were then evaluated using SHapley Additive exPlanations (SHAP) for interpretability to understand the inner working mechanisms of the model predictions.

In the first part, a model for the classification of crash-inducing snowstorm events was built using a dataset of 231 snowstorm events occurring over 21 friction testing routes in the City of Edmonton. The issue was addressed by integrating SHAP with a Support Vector Machine (SVM) model. Using the Radial Basis Function (RBF) kernel, the SVM model achieved an accuracy rate of 87.2% and a high recall rate of 80%. SHAP global explanations revealed that duration, road length, and precipitation were the most significant factors influencing crash-inducing snowstorms, along with some counterintuitive feature characteristics. To understand these counterintuitive features more clearly, local explanations were applied to closely examine representative snowstorm events, confirming the model's applicability in practical scenarios and informing future enhancements.

This study also highlighted the critical role of maintenance activities, such as plowing and anti-icing, in mitigating accident risks.

In the second part, an in-depth analysis of winter crash frequency was conducted using a dataset of 26,970 winter crashes over four years period. In the data collection step, Ordinary Kriging (OK) was evaluated as a valuable tool to interpolate traffic volume at unknown locations. The analysis first explored spatial patterns through Hot Spot Analysis (HSA), identifying high and low crash clusters. High crash frequency areas were associated with high traffic volume, high functional road class, and commercial land use, while low crash frequency areas were typically residential with lower traffic volumes and speed limits. Next, both micro and macro level variables were fused to build a crash frequency model. Three high-performance tree-based models – XGBoost, Random Forest, and LightGBM – were compared. XGBoost emerged as the best-performing model with a testing R^2 value of 92.67%, MAE of 3.64, and RMSE of 5.77. With the larger dataset, significantly more stable SHAP analysis results were obtained, enhancing the understanding of feature interactions. The global analysis indicated that road type, speed limit, and the presence of traffic enforcement cameras contributed most to the model. Key characteristics between high and low crash frequency locations were differentiated using local explanations.

The framework presented in this thesis underscores the importance of integrating interpretability techniques in practical applications to enhance winter road safety. More interpretable models provide greater insights into the fairness and trustworthiness of the model decisions, enhancing the understanding of winter road safety, and aiding maintenance personnel in effective decision-making processes and resource allocations. This thesis also recommends larger datasets for more stable models and consistent predictions, ultimately improving model reliability and decision-making processes.

PREFACE

The work presented in this thesis is either published or is under preparation of submitting to various journals in the areas of Transportation Engineering.

Journal Papers Accepted/Published

1. **Shuai, Z.,** Kwon, T. J., & Xie, Q. (2024). Using Explainable AI for Enhanced Understanding of Winter Road Safety: Insights with Support Vector Machines and SHAP. *Canadian Journal of Civil Engineering*, (ja). https://doi.org/10.1139/cjce-2023-0446

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Tae J. Kwon, for his invaluable support and guidance throughout my graduate studies. His mentorship has been profoundly impactful, and his dedication to research has left a lasting impact on me.

I am immensely grateful to the professors in the Department of Transportation Engineering for their expertise and guidance, which have broadened my horizons and enriched my knowledge. I would also like to extend my heartfelt thanks to Dr. Tony Qiu, Dr. Stephen Wong, and Dr. Japan Trivedi, for being members of my defense committee and for providing valuable insights and feedback on my research results.

I would also like to express my heartfelt appreciation to our research group members Mingjian Wu, Simita Biswas, Qian Xie, Xueru Ding, Michael Urbiztondo, and Nancy Huynh for their support and assistance. Their contributions and fellowship played an integral role in the successful completion of my research.

Finally, I would like to express my sincere gratitude to my family and friends for their unwavering support and understanding throughout the research process. Their encouragement has been a constant source of motivation for me.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURES

AI	Artificial Intelligence
ADT	Average Daily Traffic
AAWDT	Annual Average Weekday Traffic
ATE	Automated Traffic Enforcement
BI	Black Ice
CS	Cold Spots
CNN	Convolutional Neural Network
CGAN	Conditional Generative Adversarial Networks
CART	Classification And Regression Tree
D	Snowstorm Duration
Dew	Dew Point Temperature
EB	Empirical Bayes
EFB	Exclusive Feature Bundling
G	Grip value
GOSS	Gradient-based One-Side Sampling
GDPR	General Data Protection Regulation
FP	False Positive
FN	False Negative
H	Humidity
HS	Hot Spots
HSA	Hot Spots Analysis
I	Intensity of Precipitation
LIME	Local Interpretable Model-agnostic Explanations
LightGBM	Light Gradient-Boosting Machine
Length	Road Length
ML	Machine Learning
MAE	Mean Absolute Error
MSE	Mean Standardized Error
Maint	Snow plowing
MVK	Million Vehicle Kilometers
NB	Negative Binomial
OK	Ordinary Kriging
P	Precipitation
PDP	Partial Dependence Plot
R^2	Coefficient of Determination, R squared
RF	Random Forest
RBF	Radial Basis Function
RMSE	Root Mean Squared Error
RMSSE	Root Mean Square Standardized Error
SVM	Support Vector Machine
SHAP	SHapley Additive exPlanations
T	Temperature
T15	Temperature below -15 degree
TP	True Positive

TN	True Negative
W	Wind speed
WRM	Winter Road Maintenance
WRC	Winter Road Crash
XAI	eXplainable Artificial Intelligence
XGBoost	eXtreme Gradient Boosting

Chapter 1: INTRODUCTION

1.1 Background

Winter road conditions are responsible for a large percentage of road accidents in North America. According to the U.S. Department of Transportation, approximately 15% of fatal crashes, 19% of injuries, and 22% of property damage incidents occurred in the presence of adverse weather between 2007 and 2016 [1]. Similarly, the Royal Canadian Mounted Police reported that nearly 30% of collisions occurred in wet, snowy, or icy road conditions in Canada [2]. This high incidence of winter-related accidents highlights the challenges of maintaining road safety during adverse weather conditions. Analysts have estimated the economic toll of winter-related crashes at $42 billion [3], underscoring the criticality of understanding how diverse weather conditions influence motor vehicle crashes. By gaining insights into the causes of these crashes, appropriate preventive measures and maintenance activities can be effectively developed to reduce these impacts to road users.

The complexity of winter road accidents arises from both micro-level and macro-level variables. The micro-level variables typically reflect on the immediate influence factors on crashes, such as collision records, weather-related factors, road characteristics, and traffic exposure. Collision records often contain information about the at-fault party, violation type, and the circumstances of the accident. The significance of low ambient temperatures, heavy rainfall, snow accumulation, and strong wind speeds is commonly highlighted in literature as major contributors to the multifaceted nature of winter collisions [4]. The consistent changes in these weather variables often cause rapid variations in road surface properties [5], directly influencing the pavement friction, exacerbating driving conditions [6], and ultimately leading to collisions. Studies have shown that driving on snowy roads can result in a 30% reduction in driving speed [7], and can cause traffic delays up to 50% [8]. Furthermore, road design varies between regions due to the changes in land use and travel demand, leading to varying levels of traffic exposure. Therefore, micro-level factors are critical in analyzing the characteristics of winter traffic crashes.

Winter road accidents are also influenced by macroscopic variables including land use patterns and spatially identified high and low crash clusters, which capture the characteristics of the spatial unit and geographical area. Numerous studies have shown that the different land use patterns

influence travel behaviors, often altering the flow of traffic [9]. Research has also revealed that driver behaviors are closely associated with land use patterns, where dense and populated areas typically require more attentive driving [10, 11]. Furthermore, different purpose [12] of land use also necessitate changes in roadway design to accommodate varying capacities and demands [13].

Spatial analysis plays a crucial role in crash frequency analysis as many scholars have assessed an inherent spatial autocorrelation among crashes [14, 15]. Typically, spatial analysis is often conducted using hot spot analysis (HSA), which helps identify the hot spots (HS) and cold spots (CS), thus detecting the clusters of crashes with similar spatial patterns. However, the consideration of spatiotemporal dependencies tends to be overlooked [16], highlighting the significance of conducting spatial analysis on winter crashes. Establishing a spatial variable can effectively determine whether a crash location occurs in regions without any spatial pattern, or in areas with intense, spatially clustered high and low crash frequencies. Furthermore, spatial clustering serves as an effective visualization tool, reflecting the characteristics for intense high crash frequency locations and intense low crash frequency locations [17]. By integrating a spatial variable, a comprehensive crash prediction model can be developed, facilitating a deeper understanding of the relative influence of various variables on winter crashes.

The abovementioned characteristics emphasized the pressing necessity for a comprehensive investigation to understand the magnitude of these feature contributions towards winter crashes. Most research examines crash characteristics throughout the year, often not separating winter accidents from general traffic incidents. This oversight could dilute the specificity and effectiveness of safety measures tailored to winter conditions. Previous efforts have extensively focused on statistical models that are often constrained by assumptions [18], potentially limiting the power of the available crash dataset. In contrast, the power of ML models lies in analyzing the underlying patterns within the entire dataset [19], leveraging their exceptional ability to deal with non-linearity and non-parametric datasets [20]. With the continuous evolution of ML models over the past two decades, these advancements provide great opportunities for researchers to develop robust and accurate models [21].

Although ML models often demonstrate satisfactory performance and achieve high levels of model accuracy, they usually involve a critical trade-off that compromises model transparency [22]. This lack of interpretability, cited by a few studies [23, 24], fundamentally limits the utility and

trustworthiness of these ML models in real-world applications. Without insights into how and why specific predictions are made, assessing the reliability of a given model becomes challenging. This gap in understanding becomes particularly important in scenarios such as road safety, where understanding the causal factors of collisions is as crucial as predicting their likelihood. Furthermore, the performance of the model also depends on the quality of the training data [25]. Suboptimal data input can result in a deterioration of model outcomes. Moreover, many scholars also highlight the significance of other factors that affect model outcomes, such as model normalization and model generalization [26, 27].

In light of these challenges, this thesis initially collects a wide range of winter crash micro- and macro-level independent variables extensively studied in the literature. Given the complexity and non-linear nature of the dataset, an ML model becomes the natural choice to build the foundation for a prediction model. Utilizing the processed and fused data, models that fit our dataset are developed and compared based on their performance. Then, the optimal model is evaluated using eXplainable Artificial Intelligence (XAI) techniques by first identifying the global significance of individual features to the model, then by conducting local assessments to evaluate the reliability of individual predications. Providing both global and local explanations for the model predictions assist in validating the outcomes, rendering the previously opaque decisions more interpretable.

1.2 Problem Statement and Motivation

The aforementioned statistical data and empirical evidence emphasize the necessity of formulating a comprehensive winter crash prediction model. It is critical to quantify and compare the risk factors associated with winter vehicular collisions. The significant loss of lives and property damage underscores the urgent need for in-depth exploration of the factors contributing to winter crashes.

To understand the multifaceted nature of the crashes, many existing studies have developed crash prediction models based on full-year data [28, 29, 30], often neglecting to focus specifically on winter road crashes. Despite this, previously described difficulties and challenges mentioned also marks a clear need to having a better understanding of winter road crash (WRC) associated risk factors. Given this context, it is important to have the insights and analytics available for harsh winter city such as Edmonton. According to historical data provided by Canadian Climate Normals,

Edmonton experiences an average January temperature of -10.4 ℃, with temperatures potentially dropping as low as -48.3 ℃. Additionally, consistent snowfall results in an average snow depth of up to 17mm during winter months [31]. With such knowledge, we can enhance our understanding in winter road safety, which in turn improves maintenance personnel's ability to allocate resources effectively and implement preventive measures to minimize harm to road users.

ML models can be promising tools used in performing both classification and regression tasks. Unlike conventional statistical parametric modeling approach, ML models leverage robust data capabilities without the constraints of traditional assumptions. This data driven approach is extensively explored in many recent studies which have proven that the prediction power of ML model often helps obtain the better model performance and accuracy. However, due to their high reliance on the data they are trained on, this drawback can potentially impose challenges and hinders the power of the ML models. Additionally, their inherently complex internal decision-making processes make them difficult for human to interpret. The lack of model transparency also impedes the veracity of the model outcomes as understanding the reasoning behind model predictions is as important as their accuracy. To mitigate this black box nature, XAI techniques can assist in understanding the rationale behind the predictions and provide transparency to the model decision-making process.

These challenges and the current gaps in research serve as the driving force behind this thesis. The complexity in understanding winter vehicle crashes, shortcomings of the existing prediction models, the potential drawbacks of the ML models, and the urging need for transparent model solutions collectively emphasize the necessity for in-depth exploration. A thorough investigation is conducted with the goal of improving how we are predicting the winter crashes. Having transparent model predication outcomes ultimately assists maintenance personnel in making more informed and accurate decisions regarding resource allocations.

1.3 Research Objectives

The primary objective of this thesis is to leverage the power of ML models to develop robust models in crash predictions and uncover the ML inner working mechanism to the winter road maintenance personals for practical applications. To achieve this outcome, this thesis is divided into two phases to provide a comprehensive understanding of winter crash characteristics. The first

phase is to assess the underlying factors that contribute significantly to accidents during the snowstorm events. The second phase involves a preliminary assessment on the feature interactions in HS and CS, then a crash frequency regression model was built to predict the number of crashes and assess the key factors that differentiate the high and low crash frequency locations.

To achieve these goals, the research outlines the following specific tasks:

1. Identify and understand the critical risk factors contributing to crash-inducing snowstorm events.
2. Establish ML models to predict the crash-inducing snowstorm events and winter crash frequencies.
3. Explore the inner workings of ML 'black box' models using XAI techniques.
4. Provide global instance explanations to understand the feature contributions to risky snowstorm events.
5. Provide local explanations by examining representative crash-inducing snowstorm events to analyze feature interactions.
6. Apply spatial analysis to preliminarily evaluate the feature interactions at identified hot and cold spots.
7. Conduct a comparative analysis of ML algorithms to identify the optimal crash frequency models.
8. Enhance the understanding of feature behaviors under high and low winter crash frequencies using SHAP for global and local explanations.

By pursuing these tasks, this thesis aims to deliver unique methodological frameworks for analyzing the characteristics of crash-inducing snowstorm events and estimating winter crash frequencies that are both accurate and intuitive. Through these efforts, this thesis provides practical insights that could improve decision-making in winter road maintenance (WRM), ultimately leading to enhanced safety and efficient resource allocation.

1.4 Thesis Structure

This thesis is split into 6 chapters. Chapter 1 introduces the objective of the study and the motivation behind the work.

Chapter 2 outlines the literature review, where existing research on estimating crash frequency and the several approaches are discussed. These approaches include the use of conventional parametric statistical modelling, non-parametric ML approaches, and the application of XAI techniques. The risk factors that pose significant influence on the snowstorm events and winter crashes are investigated.

Chapter 3 presents the foundations of the methodologies employed in the study, which can be divided into four distinct sections. Firstly, an overview of the different machine learning methods. Subsequently, spatial clustering for the HSA. Thirdly, an overview of Kriging estimation techniques, and lastly, the SHAP techniques.

Chapter 4 presents the first phase of the work, providing a comprehensive description of the risk factors involved in collision occurred snowstorm events. An ML classification model, SVM, is then developed to classify snowstorm events with accidents involved. The performance of different kernel functions are also evaluated and compared. Subsequently, the optimal model is subjected to examination via the SHAP XAI approach.

Chapter 5 presents the second phase, beginning with a spatial analysis to identify crash prone locations in the city and visualize the feature reactions on the hot spots mapping. A series of crash frequency models are then constructed to predict the specific number of crashes by integrating both micro and macro-level predictors. The ML models are subsequently calibrated and compared using error metrics to select the most optimal model. Finally, using SHAP XAI techniques, the inner workings of the optimal model are revealed, detailing the interactions between variables, and illustrating behavioral differences amongst risk factors at high and low crash frequency locations.

Chapter 6 concludes the thesis with a summary and discussion of the research outcomes and contributions. Additionally, the chapter mentions the encountered challenges and limitations while providing suggestions for further research directions.

Chapter 2: LITERATURE REVIEW

Chapter two aims to provide an overview of the different model approaches in predicting crashes for general traffic safety practices. The chapter is further divided into four different sections. The first section evaluates the risk factors that significantly contribute to the winter crashes from both a micro and macro level. The second section discusses the parametric, statistical model that researchers conventionally applied to develop the predictive models. It then discusses, the contemporary ML-based approaches that often result in better performance and high efficiency are discussed. In the third section, the challenges involving ML model interpretability are explored. In the final section, a summary is provided, noting the limitations of the reviewed literatures.

2.1 Risk Factors Affecting Crashes

There are a variety of influencing factors that contribute to the crashes. They can be categorized as two types of variables: microscopic and macroscopic. The use of micro-level features helps increase the granularity of the dataset. Macro-level variables reflect larger scale influences of demographic, socioeconomics, and network attributes, providing insights for long-term, large-scale safety planning.

For the micro-level factors, the current literature focuses on detailed and granular variables such as traffic volume, vehicle speed, road surface conditions, and weather data to assess crash risks. Wang et al. [32] established and compared three types of crash frequency models based on average daily traffic (ADT-based), average hourly traffic (AHD-based), and microscopic traffic at 5 min interval for safety analysis. The result showed for all three models that the traffic exposure, speed variation, road length, and the existence of diverging segments are positively significant. Additionally, Abdel-Aty et al. [33] assessed crash occurrences on the urban freeway using road geometric characteristics and microscopic traffic variables to establish their crash frequency model. Their findings showed the median type, pavement surface types, and average annual daily traffic (AADT) as critical factors contributing to crashes. Peng et al. [34] investigated the effect of reduced visibility using real-time microscopic data, concluding that reduced visibility would significantly increase the crash risk especially for rear-end crashes and the risk impact varies for different vehicle types and lanes. Under heavy rainfall conditions, Jung et al. [35] has evaluated

the contributing factors to the crashes. In this study, they collected the multiple data sources from microscopic geographics and secondary weather data. It was identified that posted speed limit change, presence of off-ramps, and changes of pavement surface material has significant influence on the crashes in rainy weather. Similarly, Seeherman and Liu [36] examined the effects of snowfall on traffic safety. The study found that heavy snowfall critically impacted on crashes, and the effects of mixed precipitation conditions were less impactful than entirely snowy conditions. Hermans et al. [37] investigated the impact of severe weather conditions on the number of crashes by incorporating a total of 17 climate factors, concluding that an increase in maximum wind gusts positively influences crash occurrences, while global radiation and sunshine duration negatively impacted crash occurrences.

With regards to the macro-level variables, existing studies examine broader contextual factors such as land use, road networks, socioeconomics, and spatial interactions to understand their impact on crashes. Chand et al. [38] considered macroscopic variables from demographic, vehicle utilisation, environmental, responder variables, and street network features to establish the crash frequency model and found that highly connected and dense road networks were often associated with higher crash frequency. Additionally, a comparative analysis conducted by Huang et al. [39] compared macro-level Bayesian CAR model with micro-level Bayesian spatial joint model. With an integration of both models, the specific treatment strategies could be provided to different screening categories. Moreover, Pervaz et al. performed an integrated framework to model crashes at macro and micro-level, including roadway and traffic factors, land use, built environment and sociodemographic characteristics for model estimation [40]. The researchers demonstrated an improved model performance compared to a non-integrated macro model. Furthermore, Wang et al. [41] investigated the effect of zonal factors in estimating crash risk for different transportation modes. The study found that zonal factors play an important role in model performance on non-motorized crashes and significantly affect intersection crashes.

2.2 Methods for Crash Analysis and Estimation

This section breaks down two different modeling approaches in traffic safety literature. The first section concerns the parametric modeling approach, which was prevalent in early modelling frameworks. This type of approach is often constrained due to underlying assumptions that hinder

the flexibility and accuracy of the model. In response, many literatures have explored the ML modeling approach, which has led to high-performance models that fully utilize data which are discussed in the second section.

2.2.1 Parametric Crash Prediction Models

Traditionally, researchers have analyzed factors affecting collisions through a parametric modeling framework that assumes an inherent data distribution. Typical data and methodological related issues for statistical analysis of crash frequency modeling often involve over-dispersion, under-dispersion, time-varying explanatory variables, temporal and spatial correlations, and so on [42]. For example, the Poisson model is a basic model for simple estimation; however, it cannot handle over-dispersion and under-dispersion, and is significantly influenced by low sample means and smaller sample sizes.

Regardless, researchers Jovanis and Chang [12] proposed to use a Poisson Regression modeling approach to analyze the number of accidents and miles traveled. The authors revealed that automobile accidents are much more sensitive to environmental conditions than truck accidents. Using a similar approach, Jones at al. [43] analyzed urban freeway accident frequency and duration, which can be used to assist in the ongoing development of Seattle's accident management system. Besides, Wei and Lovegrove [44] leveraged a Negative Binomial (NB) regression model to predict macro level cyclist collisions, and found that an increase in total lane kilometres, intersection density, and arterial-local intersection percentages increases the bicycle-auto collisions. To examine accident occurrences at signalized intersections, Chin and Quddus [45] utilized a random effect NB model to handle spatial and temporal effects in their data. Their study found that the total approached volumes, number of phases per cycle, presence of uncontrolled left-turn lanes, and the presence of surveillance cameras are major influential factors.

Additionally, multivariate models, generalized estimating equation models, negative multinomial models also are major statistical model types in analyzing the crash frequency data [42]. A previous effort conducted by Abohassan et al. [46] examined the effect of maintenance, weather, surface conditions, and road safety variables using Structural Equation Modeling and Path Analysis. This approach highlighted the interdependency between these variables, offering a deeper understanding of how different factors interact to influence winter road conditions.

While traditional parametric methods have significantly contributed to our understanding of road safety, their dependence on assumed pre-assumptions can constrain their flexibility and accuracy [47]. The complexity of winter road conditions, with intricate interactions between variables such as weather, traffic, and maintenance, calls for more flexible and data-driven approaches.

2.2.2 Machine Learning Predictive Models

In response to these challenges, ML algorithms have emerged as powerful alternatives. These algorithms make minimal assumptions about relationships between predictors and the desired output variable, leading to higher performance in most applications [48]. Due to significant breakthroughs in artificial intelligence (AI), ML has gained lots of favor for researchers for developing comprehensive, accurate and efficient models.

In recent studies, Li and Abdel-Aty [49] developed a real-time crash occurrence prediction model using temporal attention-based deep learning and trajectory fusion, showing reasonable performance for crash occurrence prediction. Additionally, Li et al. [50] used SVM to predict crash frequency on rural frontage roads in Texas with different sample sizes and performed a comparative analysis with a conventional NB model. Results show that SVM possesses a better prediction performance. However, the reasons behind why an increased data size can result in a decrease in model performance remain unclear. Researchers Zarei et al. [51] used Conditional Generative Adversarial Networks (CGAN) for Empirical Bayes (EB) analysis of road crash hotspots demonstrating the robustness of the CGAN-EB model than traditional NB-EB model. Hu et al. [52] developed a deep learning, convolutional neural network (CNN), and decision tree-based approach using connected vehicle data, finding that the CNN model achieved best accuracy to predict the crash risk at intersections.

In addition to the previous mentioned complex ML approaches, more interpretable tree-based models are also popular in predicting crashes. Chang and Chen [53] modelled freeway crash frequency with Classification and Regression Trees (CART) and compared the results with a NB regression model, leading to a marginally higher prediction accuracy. Noting marginal improvement, later researchers have adopted ensemble learning methods, which often combines multiple weak classifiers to improve predictive power, decrease sensitivity to noise, and provide feature importance for the model. Zhang et al. [54] developed a random forest (RF) model for real-

time crash occurrence prediction on freeways using crowdsourced probe vehicle data which has found better performance against competing models such as logistic regression, XGBoost, and SVM. Furthermore, Goswamy et al. [55] developed a XGBoost model for investigating the factors affecting injury severity at pedestrian crossing locations with rectangular rapid flashing beacons, finding superior performance over the random parameter discrete outcome models.

Overall, ML algorithms, with their ability to handle non-linearity and non-parametric datasets, offer higher performance than traditional parametric models [56]. While ML models excel at predicting outcomes, they often do so at the expense of model transparency such as deep learning and ensembles [57]. Such a lack of interpretability, as cited by a few studies [23, 24], fundamentally limits the utility and trustworthiness of these ML models in real-world applications. This gap in understanding becomes particularly important in scenarios such as road safety, where discerning the causal factors of collisions on a road segment is as crucial as predicting their likelihood.

2.3 Methods for Enhancing Machine Learning Interpretability

Due to advancements of ML models and their improved predictive accuracy, ML has led to widespread adoptions in industrial applications [58]. However, such enhancements in predictive power have often been achieved through an increase in model complexity [57]. There is a clear trade-off between the performance of a ML model and its ability to produce explainable and interpretable predictions. This phenomenon is often referred as the “black box” problem. To address the need to tackle the “black box” nature of ML models, there has been a surge of interest in interpretable ML: a system that provides explanations for their outputs [59]. Systems whose decisions cannot be well-interpreted are difficult to trust, especially in sectors such as healthcare, where moral and fairness issues have naturally arisen [57]. Without understanding how ML model predictions are made, there is a significant challenge in adopting these models in applications. The prediction outcomes need to emphasize their resolution in the decision-making process. With this vision in mind, the need for building trustworthy, fair, robust, but also highly performance models for real-world applications has led to the revival of the field of eXplainable Artificial Intelligence (XAI) - a field focused on the understanding and interpretation of the behaviour of AI systems.

Challenges with Complex Machine Learning Model Predictions

According to Doshi-Velez and Kim, metrics established in ML such as classification accuracy, are not a complete description of most real-world tasks [59]. Understanding the basic interpretability requirement, several typical challenges arise in academia regarding ML predictions such as high-stakes decision impacts, societal concerns, and regulations. In terms of high-stakes decision impacts, Wexler [60] has highlighted the danger of deploying automated criminal justice systems at every stage, such as incorrect bail decisions that might release potentially dangerous criminals. Similarly, faulty pollution models may incorrectly deem dangerous situations as safe [61]. These forementioned cases all indicate a partial relevance to computer-aided modeling. Regarding social concerns, Doshi-Velez and Kim [59] specified it can be optimized through interpretability, highlighting downstream tasks such as fairness, privacy, reliability, causality, and trust. Moreover, new regulations require verifiability, accountability, and transparency in algorithm decisions. The recently enforced European General Data Protection Regulation (GDPR) provides subjects with the right to an explanation of algorithmic decisions involving their data [62].

Model Interpretation Using Shapley Additive exPlanations

There are typically two types of model interpretability: intrinsic and post hoc. The intrinsic interpretability involves understanding the constraints imposed on the complexity of the ML model itself, whereas the post hoc interpretability involves applying methods to analyze the trained model after it has been built [63]. The intrinsic interpretability, also known as model-specific interpretation, refers to models that are interpretable by themselves during the in-model phase. Post hoc interpretability, or model-agnostic interpretation, involves explanation methods applied after the model training phase to provide insights into the model's decision-making process [61].

Model-specific interpretation methods are limited due to the reliance on the unique internal structure of the model which tends to have simple structures such as the linear, decision tree, and rule-based models [57]. Regardless, when it comes to models with higher complexity, model agnostic interpretation can be very useful. Currently, SHapley Additive exPlanations (SHAP) has emerged as a powerful model-agnostic tool, bringing explanations to complex models. The method leverages concepts in game theory to enhance interpretability through the computation of importance values for each feature that composes an individual prediction [64].

SHAP has been used in various domains. In the field of healthcare, Moreno-Sanchez [65] developed an interpretable model to predict heart failure survival using ensemble ML trees, identifying XGBoost as the best amongst models such as Decision Trees, Random Forest, Extra Trees, AdaBoost, Gradient Boosting, and XGBoost. SHAP analysis was used to gain the interpretability of the prediction model developed. Similarly, Athanasiou et al. [66] developed an explainable personalized risk prediction model for cardiovascular diseases in patients with Type 2 Diabetes Mellitus, utilizing XGBoost alongside the Tree SHAP technique to generate individual explanations for model decisions. In finance and marketing sectors, researchers Lin and Gao [67] applied group SHAP to evaluate the different abilities of companies, which presented excellent performance in local interpretation and with decreased the computation time. Moreover, in the marketing field, Meng et al. [68] utilized an XGBoost model along with SHAP to predict and explain online review helpfulness. This comprehensive application of SHAP across diverse fields underscores its utility in enhancing a model interpretability and transparency, making it a valuable tool for understanding complex ML models.

2.4 Summary

This chapter has conducted a comprehensive literature review on winter crash risk factors, crash prediction models, and methods for improving the interpretability of ML models. Although parametric models provide a solid foundation for crash prediction, they often struggle with complex, high-dimensional data. In contrast, ML models offer superior performance and flexibility, but this often comes at the expense of interpretability. Recent advances in explainable AI, such as SHAP, provide promising solutions to bridge this gap, enhancing the transparency and trustworthiness of ML models. By combining both powerful predictive models with transparent decision-making mechanisms, we can ensure and appreciate a model's fairness and trustworthiness. Future research should focus on integrating these interpretability techniques into practical applications to improve winter road safety.

Chapter 3: METHODOLOGY

The previous chapters discuss the limitations in existing crash frequency models due to the underlying assumptions behind the statistical parametric model, the limitations on prediction variables, and how model transparency remains unknown due to the complexities of ML models. To address these identified gaps, this thesis established a series of machine learning based frameworks to enhance winter road safety in the city of Edmonton (**Figure 3-1**).

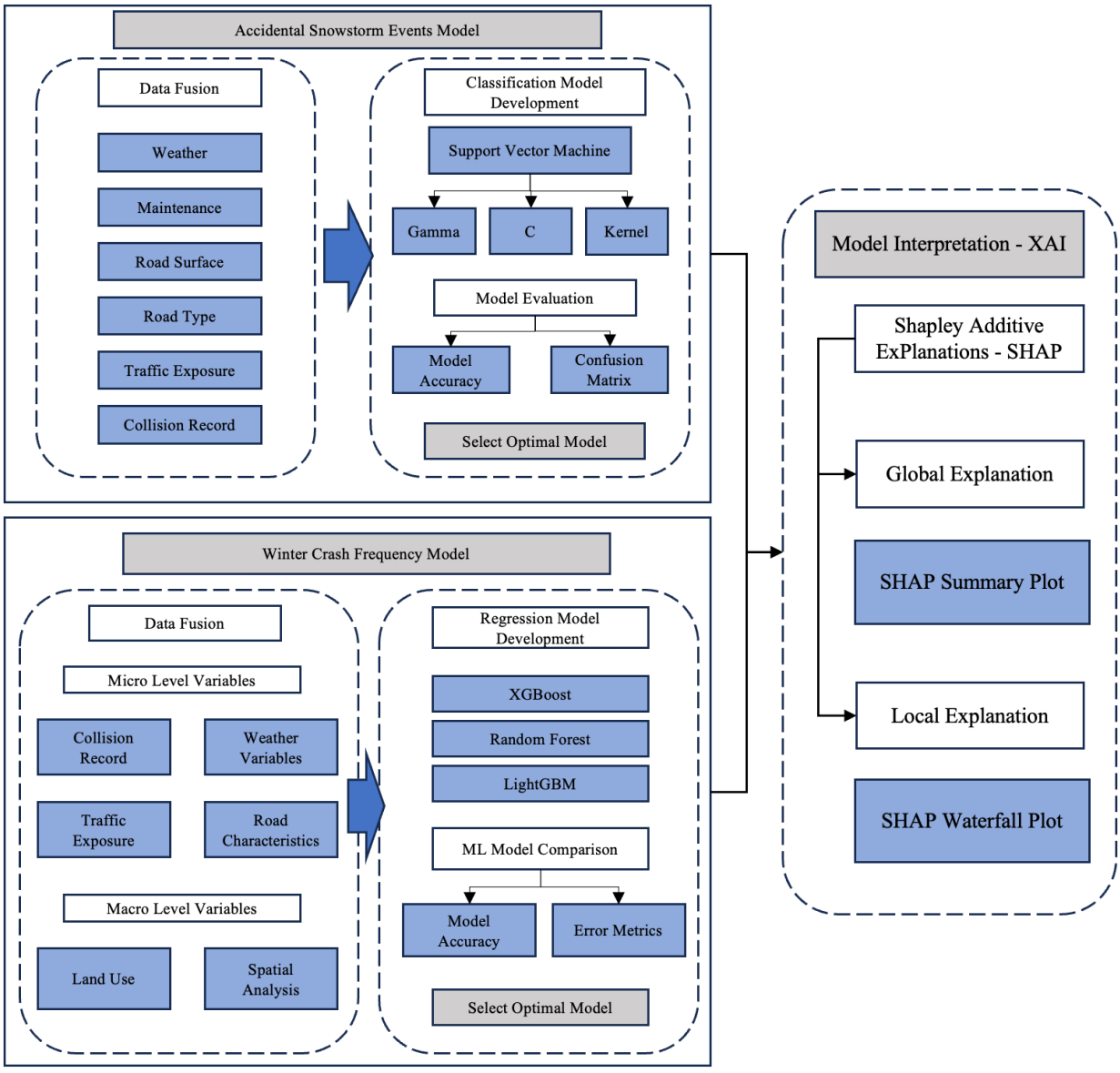


Figure 3-1 Methodological Framework

This thesis performs two phases of analysis to mitigate winter driving difficulties and provide insights in hazardous snowstorms and winter crash characteristics using XAI techniques. In the first phase, the binary classification model was built to determine the risk factors that lead to crash-inducing snowstorm events according to the examined friction testing routes. The second phase considers both macro and micro level variables to give a holistic evaluation on the entire city of Edmonton winter crash risk factors, and how various features interplay differently with each other at high and low crash frequency locations. Both phases have developed ML models, and through comparisons in model performance metrics, the model with high predictive power are assured. Most importantly, for the selected model, the thesis also applies XAI techniques to increase the interpretability and improve the models' transparency. This comprehensive workflow allows the model prediction outcomes to be explained, providing insight on how a model established the outcome by further decomposing the contributions of each of the individual features. Within this chapter, we conduct a thorough examination of the principles behind the applied method.

3.1 Spatial Analysis via Getis Ord Gi*

Getis-Ord Gi* is a spatial analysis technique used to identify the hot and cold spots within a geographic area. The study performed spatial analysis using the Hot Spots Analysis tool in ArcGIS. The output of this tool creates a new output feature class with z-score, p-value, and confidence level bins (Gi_Bin). The tool functions by examining each feature within the context of its neighboring features, creating a spatial weight matrix based on the proximity between locations. For the statistically significant positive z-scores, the larger the z-score is, the more intense the clustering of high values (hot spots) [69], which in our context means a cluster with high crash frequency. Whereas a negative z-score, indicates an intense cluster of low values (cold spots) signifying a cluster with low crash frequency.

The method is commonly used in transportation engineering. In traffic safety, it is commonly used to locate crash-prone sites and give a spatial assessment of the feature influences from the spatial distribution of events. In this use case, it helps identify regions within the city with high and low crash frequencies. By overlaying additional features on top of the hot spots map, we can conduct a preliminary assessment of the influences of these features, thereby enhancing our understanding

of crash patterns in winter. Additionally, this investigation aids in the subsequent development of an ML crash frequency model.

The following **Equations 3.1 to 3.3** are used for calculating the Getis-Ord local Statistics.

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{[n\sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2]}{n-1}}} \quad (3.1)$$

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (3.2)$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \quad (3.3)$$

Where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features.

3.2 Ordinary Kriging

To understand the concept of Ordinary Kriging (OK), it is essential to first introduce the fundamentals of kriging. Kriging operates on the principle that spatially correlated data can be used to predict values at unsampled locations. It helps to weigh the surrounding measured values, and derive a prediction for an unmeasured location. The general formula for both interpolators is formed as a weighted sum of the data as described in **Equation 3.4.**

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \quad (3.4)$$

Where $\hat{Z}(s_0)$ is the measured value at the ith location, λ_i is an unknown weight for the measured value at the ith location, s_o is the prediction location, and N is the number of measured values.

Amongst all kriging approaches, OK is selected for our study. As OK is the most used form of kriging methods due to a combination of its simplicity and high accuracy [70]. It assumes a constant but unknown mean across the study area. The model is expressed in **Equation 3.5**.

$$\boldsymbol{Z(s) = \mu + \varepsilon(s)} \quad \textbf{(3.5)}$$

Where μ is an unknown constant, $\varepsilon(s)$ is the residual. Kriging weights are also needed to estimate values at unknown points. Spatial autocorrelations imply an internal spatial relationship between sample points. Kriging weights therefore depends on both the distance between observations and prediction location and the overall spatial arrangement of the observations. The objective of OK interpolation is to determine the optimal kriging weights that minimize the estimation variance which is illustrated in **Equation 3.6**.

$$\boldsymbol{\hat{Z}(s) = \mu + \sum_{i=1}^{N} \lambda_i \left[Z(s_i) - m(s_i)\right]} \quad \textbf{(3.6)}$$

Where $\hat{Z}(s)$ is OK estimate, $Z(s_i)$ is the observation at location s_i, λ_i is the unknown Kriging weight for the observation at location s_i, s is the location for estimation, N is the number of observations, and $m(s_i)$ is the expected values of $Z(s_i)$.

Beside the kriging weight, it is also important to construct a high quality semivariogram to determine the accuracy of the estimation results. Semivariance represents the reciprocal of the spatial autocorrelation, which is calculated using **Equation 3.7**. It illustrates how similarity between values changes with separation distance.

$$\boldsymbol{\gamma(h) = \frac{1}{2}\left[z(x_i) - z(x_j)\right]^2} \quad \textbf{(3.7)}$$

Where $\gamma(h)$ is the semivariance between sample points x_i and x_j in h distance, and z is the feature value. Afterwards, the semivariogram is compared using three commonly used model forms: Spherical, Gaussian, and Exponential. The fitted semivariogram model provides three spatial parameters: sill, range, and nugget [71]. The sill represents the semivariance level where the model begins to plateau. The range is the lag distance at which the semivariance reaches the sill, beyond which spatial autocorrelation is considered non-existent. Finally, the nugget represents the spatial

variability at distances smaller than the shortest distance between observations, often termed measurement error.

3.3 Machine Learning Algorithms

ML algorithms are a powerful tool for both classification and regression tasks when it comes to the predictive power and robustness of a model. Over the years, ML models have evolved dramatically, which surpass some of the conventional traffic safety models that rely on a statistical approach, such as the negative binomial model, Poisson model, etc. However, the problem of these parametric models is often the constraints caused by the predefined assumption, which limits the full utilization of the collected data. This primarily limits the potential of abundant traffic data, hindering the pathway to a high-performance model. ML is a heavily data driven and non-parametric model approach which overcomes the problem from conventional statistical models. Because of their high performance, it has quickly gained researchers' interest in building robust predictive models.

In this section, we discuss the models applied for this study and provide an overall description of each model, including its concepts and its characteristics. As previously mentioned, this thesis has two phases. The first phase applied SVM to classify whether snowstorm event induced crashes. The resulting model is then compared with different kernel functions and evaluated based on the performance metrics such as accuracy, precision, recall and F1-score. In the second phase, some of the most popular tree-based ML algorithms are applied to perform the regression task for predicting crash frequency. Namely, these are Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Light Gradient Boosting Machine (LightGBM). The three models are compared based on their percentage variance explained and error metrics.

3.3.1 Support Vector Machine

SVM is a supervised learning method that can be used for classification and regression [72]. The model is renowned for its ability to model high-dimensional relationships [73]. According to Vapnik [74], two of the most outstanding qualities for SVM are the power to find the global minima (because of convex optimisation), as well as its decreased proclivity to overfitting (because of structural risk minimisation principle). Additionally, many literatures also conclude that for studies with a small dataset size, the performance of SVM (shallow structure model) for classification

tasks tends to produce better classification results than deep structure models [75, 76]. Hence, due to the limited total of 231snowstorm events within the dataset of our study, SVM is selected to classify the crash-inducing snowstorms.

According to WS Noble, to grasp the essence of SVM classification, one should understand four foundational concepts: separating hyperplanes, maximum-margin hyperplanes, the soft margin, and the kernel function [77]. A separating hyperplane is essentially a line that separates different classes of samples. SVM selects a hyperplane that maximizes the distance between it and the nearest data points. The concept of soft margin allows for a trade-off between hyperplane violations and the margin size. In general, the hyperplane that separates the data cannot be found in 2-D space; a projection must be made to transform the data into a higher-dimensional space to find a separating hyperplane. This projection into higher dimensional space can become computationally expensive. Thus, a technique called "kernel trick" is proposed as a solution to this problem. The kernel trick is a formula that quantifies the relationship between variables in a higher dimension by formulation, eliminating the need to transform the data. There are several forms of the kernel trick, and some of the most common are Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid [78, 79]. These kernel tricks are explored in our analysis.

Equation 3.8 is the principal equation that solves the SVM classification problem.

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i$$

$$\text{Subject to } y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i \tag{3.8}$$

$$\zeta_i \geq 0, i = 1, \ldots, n .$$

In other words, in a given set of training vectors $x_i \in \mathbb{R}^P, i = 1, \ldots, n$, two classes use a vector $y \in \{1, -1\}^n$. The goal of this equation is to find $w \in \mathbb{R}^P$ and $b \in \mathbb{R}$, such that the prediction given by $\text{sign}(w^T \phi(x_i) + b)$ is mostly correct among the samples. The equation aims to maximize the margin, minimize $\|w\|^2 = w^T w$, and apply a penalty when a sample is misclassified. In an ideal case, the value $y_i(w^T \phi(x_i) + b)$ is less or equal to 1 for all the samples. However, a perfect separatable hyperplane is not always available. Thus, distance ζ_i helps by allowing some samples

to fall in the correct margin boundary. C is the penalty term acting as an inverse regularization parameter [80].

3.3.2 Tree-Structured Models

Extreme Gradient Boost

XGBoost [81] is a powerful and efficient tool that implements gradient boosted decision trees, which are essentially an ensemble method that build the model in a stage-wise manner. It is a gradient boosting algorithm that involves building an ensemble of weak learners. This means the model starts with a weak learner, then after every iteration, a new predictor comes to account for the errors of the previous model, which is then added into the ensemble. The iteration stops once the number of boosting rounds ends.

The model consists mostly of two important parts, a training loss function and regularization. The first part is a training loss function, denoted as l, which measures the difference between the actual value and the predicted values. Typically, a differentiable convex function is used. The second term, Ω, is the regularization term, which penalizes the complexity of the model, helping smooth the final learnt weights to avoid overfitting [82]. The objective function is shown in **Equation 3.9**.

$$\boldsymbol{Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i{}^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + constant} \qquad \textbf{(3.9)}$$

Where l is the loss function, y_i is the target value, $\hat{y}_i$ is the prediction value, Ω is the regularization, f_t is the t-th regression tree.

Equation 3.10 provides the definition of the tree, $f(x)$. And **Equation 3.11** defines the complexity of the tree, $\omega(f)$.

$$\boldsymbol{f_t(x) = \omega_{q(x)}, \omega \in R^T, q: R^d \rightarrow \{1, 2, \ldots, T\}} \qquad \textbf{(3.10)}$$

Here, ω is the vector of scores on leaves, q is a function assigning each data point to the corresponding leaf, and T is the number of leaves.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \quad (3.11)$$

Where the γ and λ are the regularization parameters that prevents the model from overfitting.

Random Forest

Random Forest (RF) is an ensemble learning method introduced by Breiman in 2001 [83]. It builds multiple decision trees and combines their results to enhance predictive performance. Unlike XGBoost, which minimizes an explicit objective function through sequential boosting, RF utilizes the bagging method (bootstrap aggregating), and feature randomness to create an uncorrelated forest of decision trees. For each decision tree in the forest, a random subset of features and training data is used. This helps diversify the trees which helps reduce the variance and prevent overfitting. In regards to the RF regression model built for our study, the predictions from each individual tree are averaged to generate the final output. Here, an intuition diagram for RF is illustrated in **Figure 3-2.** By averaging the trees, the model reduces the impact of the noise and outliers in the training data, hence leading to better predictions.

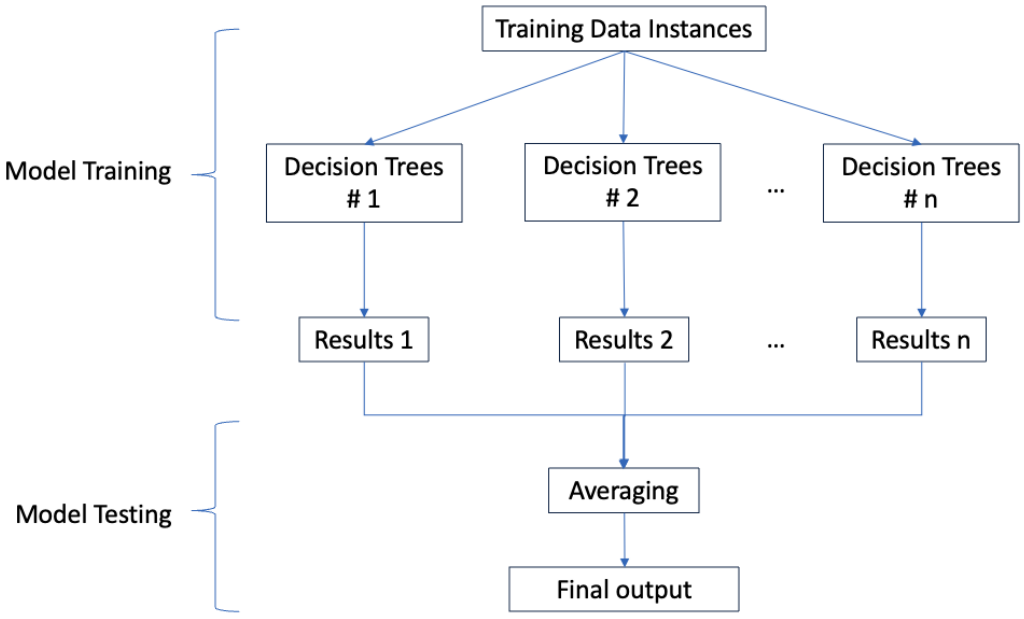


Figure 3-2 The Intuition Diagram for Random Forest Regression Model

To better capture the underlying patterns of the data, the model tends to grow more trees and increases the depth of each tree. Although this does improve the model performance, this inherently increases the complexity of the model. Hence, it once again highlights the importance of understanding the inner working mechanisms of the model, which helps us to validate the truthfulness of the generated outcome.

Light Gradient Boosting Machine

Light Gradient Boosting Machine (LightGBM) is another popular gradient boosting framework designed to enhance training efficiency, particularly when dealing with high-dimensional features and large datasets [84]. Developed as an improvement over XGBoost, LightGBM aims to accelerate model training while maintaining accuracy. To achieve this, LightGBM introduces two novel techniques: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB).

GOSS is applied to retain instances with large gradients while randomly dropping those with small gradients. This selective sampling helps optimize memory usage and significantly reduces training time by focusing computational resources on the most informative instances. This technique differentiates LightGBM from most decision tree learning algorithms that often grow trees level-wise. Instead. LightGBM grows trees leaf-wise, [85] which can potentially lead to faster convergence [86]. Exclusive Feature Bundling (EFB), on the other hand, addresses the issue of high-dimensional feature spaces by bundling mutually exclusive features together. This reduces the number of features, further improving training efficiency.

By combining these techniques with the gradient boosting framework, LightGBM achieves faster training speeds, lower memory usage, and improved accuracy. These enhancements make LightGBM a "light", yet powerful tool for machine learning tasks involving large and complex datasets.

3.4 SHapley Additive exPlanations

A model agnostic XAI technique, SHAP is designed to solve the interpretability problem that plagues most ML approaches. By combining SHAP and ML models, we can accurately identify collision risks and understand the reasoning behind the model's selection.

The main objective of our analytical framework is to determine the contribution of each predictor to a single observation instance [64]. SHAP is based on cooperative game theory, where players in a game form coalitions, and the rewards gained are distributed to the players. It is unlikely that each player contributes equally to the successful outcome. Hence, a system is needed to determine the contribution of the players in the coalition. Analogously, in an ML model, the players are the features, and the SHAP value is the amount by which the features contribute to the prediction output [87]. This approach is highlighted in **Equation 3.12**, where additive feature attribution methods have an explanation model that is a linear function of binary variables.

$$\boldsymbol{g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i} \tag{3.12}$$

Where g is the explanation model, $z' \in \{0,1\}^M$ is the coalition vector, M is the maximum coalition size, and $\phi_i \in \mathbb{R}$ is the feature attribution for a feature j – the Shapley values [64].

The Shapley values are calculated using the following **Equation 3.13**,

$$\boldsymbol{\phi_i(x) = \sum_{S \subset F\{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)\right]} \tag{3.13}$$

Where x is the observation input, $\phi_i(x)$ is the Shapley value for the corresponding feature i in game f. F represent the feature set. For the trained model, f_S meaning subset of feature S and $f_{S\cup\{i\}}$ means the subset of feature S and i. Some restrictions were represented as x_S, meaning the restricted input of x given the subset of feature S. Similarly, $x_{S\cup\{i\}}$ means the restricted input of x given the subset of feature S and $\{i\}$ [87].

As SHAP values describe feature importance for a single instance, it is also possible to quantify global importance of a given feature by taking the average of the absolute SHAP value for all instances with respect to a particular feature. This process is depicted in **Equation 3.14**,

$$I_j = \frac{1}{n}\sum_{i=1}^{n}\left|\phi_j^{(i)}\right| \tag{3.14}$$

Where I_j is the feature importance of the j th feature and $\phi_j^{(i)}$ is the Shapley value of the j th feature of the i th sample.

3.5 Summary

This chapter provides an overview of the methods applied in this thesis. The application of ML serves as the main theme throughout this work, addressing classification and regression tasks aimed at improving winter road safety. Consequently, the chapter introduces the mathematical principles behind each model and justifies their implementation.

The ML algorithms section introduces four widely adopted models: XGBoost, RF, LightGBM, and SVM. By understanding the working principles and limitations of each model, we can observe the differences and similarities between them. It is noteworthy that all four models involved in the study have complex structures. Advanced tree-based models often increase the number of trees and the tree depth to enhance predictive power at the expense of increasing model complexity. SVM, known for its ability to handle high-dimensional unstructured datasets, uses kernel functions to project data into a high-dimensional space to find the best-fitting hyperplane. This mathematical transformation increases the model's complexity as well. Besides the ML methods, we also presented the concept of HSA, which helps identify the crash-prone locations, and the power of OK for data interpolation.

While the increase in ML complexity significantly improves model performance, understanding why a model makes a particular prediction and the evidence supporting that prediction remains challenging. To address this, we introduce the concept of XAI and a powerful explainability tool known as SHAP, which relies on game theory. The application of SHAP helps reflects the feature importance in a global perspective and provides local explanations to individual predictions. In the following two chapters, we will demonstrate the predictive power of ML algorithms, assisted by XAI, to interpret and evaluate the validity of the results. This approach allows us to gain valuable

insights into the practical implementation of the framework and contributes to an enhanced understanding of winter road safety.

Chapter 4: DEVELOPING A MACHINE LEARNING-BASED APPROACH FOR IDENTIFYING CRASH-INDUCING SNOWSTORM EVENTS[1]

In this chapter, the risk factors contributed to the crash-inducing snowstorm events are analyzed at the friction testing routes in the city of Edmonton. The first section gives the background of the study area and the data collection. This section also describes the location of the friction testing routes in the city, showcasing how data was collected and prepared before the modeling phase. The second section focused on ML classification model development using a SVM model and compared its performance metrics under different kernel functions. In the third section, the established complex model is explained using SHAP XAI techniques, where the global explanations of all features are examined to give an overview of the key features contributing to the model outcome and the impacts of high and low feature values to the prediction. In the final section, the representative crash-inducing snowstorm event instances are assessed using SHAP local explanations to uncover general patterns and characteristics of these events, making the model results transparent and interpret for maintenance personnel.

4.1 Study Area and Data Preparation

4.1.1 Study Area

The study area selected for this research is the city of Edmonton, given its cold and prolonged winter season stretching from November to March. The climatic conditions in Edmonton are characterized by significant snowfall, icy roads, and fluctuating temperatures, further reinforcing its suitability for this study. The datasets, collected over two winter seasons (2017/2018 and 2018/2019) across an extensive road network in Edmonton, include a total of 234 friction tests conducted using the Vericom Brakemeter 4000. In total, 21 routes (13 arterials and 8 collectors) were tested across the city shown in **Figure 1**.

[1] Shuai, Z., Kwon, T. J., & Xie, Q. (2024). Using Explainable AI for Enhanced Understanding of Winter Road Safety: Insights with Support Vector Machines and SHAP. Canadian Journal of Civil Engineering, (ja).

These tests required achieving a speed of 30 km/h before initiating the braking process, resulting in a friction measurement based on the device's internal G-force sensor once the driver came to a complete stop. While friction measurements are intermittent and cannot capture all road condition variations, they are a pragmatic method of data collection. These measurements were collected after snow events over two years and serve as a vital sample that reflects the dynamic nature of winter road surfaces in urban environments.

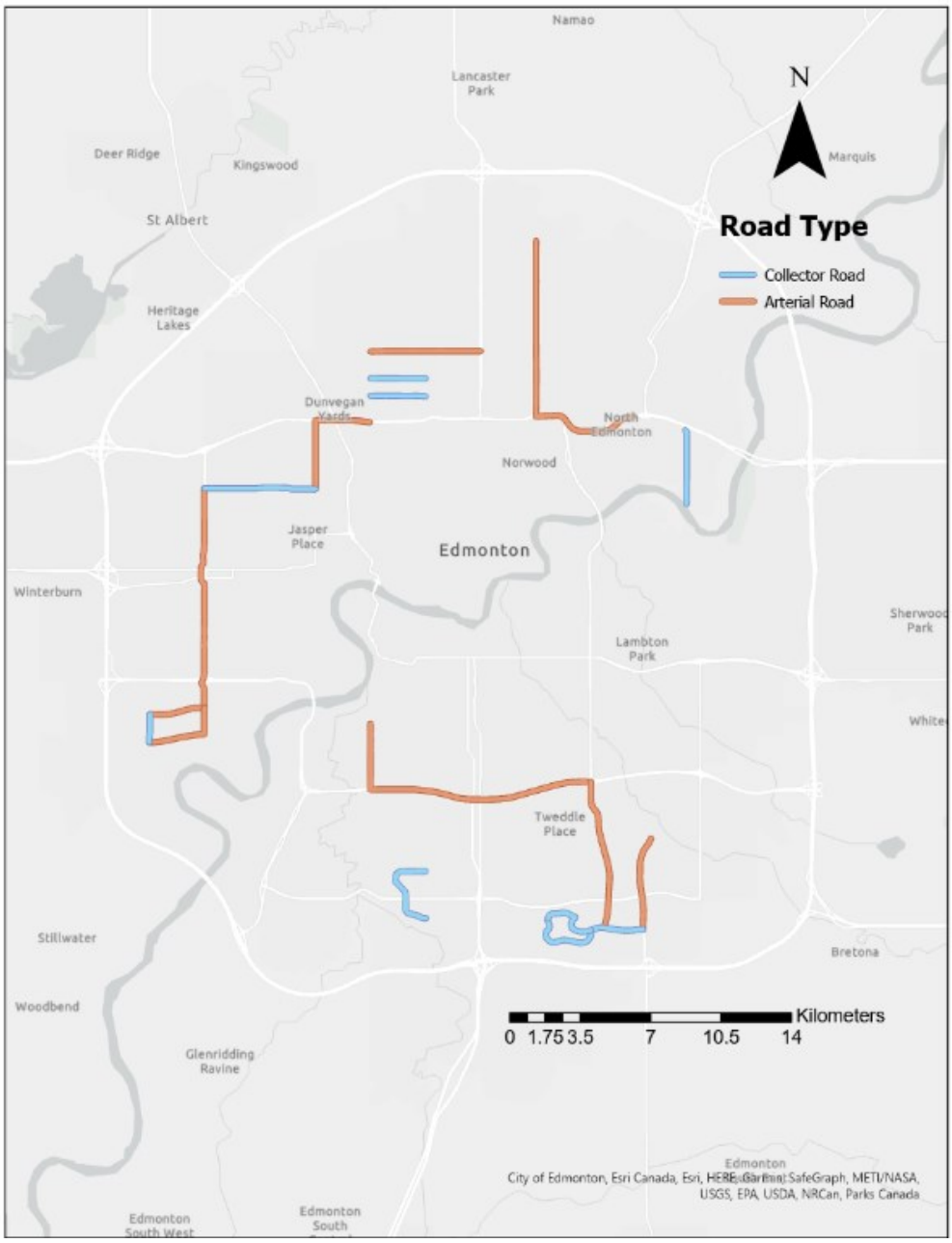


Figure 4-1 Friction Testing Routes

4.1.2 Data Preparation

An overview of the variables used and descriptive statistics for each variable are presented in **Table 4-1** and **Table 4-2**, respectively. In total, the dataset contains 16 predictors categorized into weather, maintenance status, road surface, road type, and traffic exposure variables. After data cleaning, 231 unique snowstorm events were retained for analysis. Our study adopted an event-based approach, with the friction-tested road segments as the spatial units, and the snowstorm events duration as the temporal units. By correlating friction test timestamps and location where the test conducted with specific snowstorm events, we integrated those factors to assess their collective impact on winter road safety.

In terms of weather-related variables, we used hourly weather data to define each snowstorm event. For each event, average ambient temperature, dew temperature, humidity, and wind speed during its occurrence are considered. Given the relatively short duration of these events and the granularity of hourly data collection, the average of these variables provides a reliable estimate of weather conditions. To improve the accuracy of the weather data, readings were collected from the closest weather station to the friction test route. The start and end times of snowstorm events are systematically determined by the presence of precipitation combined with an average temperature of 5 degrees Celsius or below. The event is deemed to have ended following a lapse of three hours without precipitation [88]. Moreover, the study accounts for the characteristics of each friction test road segment including road type and road length. Regarding traffic exposure variables, this study used average daily traffic (ADT) as an indicator of general traffic volume due to AADT being unavailable. In addition to weather and traffic exposure categories, three additional supplementary variables (Black Ice, T15, and Intensity) were included. Black Ice is a binary variable that is assigned a value of "1" when the mean dew point temperature is within a 2-degree range of the mean ambient temperature during a snowstorm. T15 is another binary variable representing events where the average temperature is -15°C or lower. Precipitation intensity was derived using the value of precipitation divided by its corresponding duration [6]. The inclusion of intensity offers a critical measure of the storm's potential impact on road safety which is distinct from mere precipitation. This differentiation allows us to distinguish between the mere presence of precipitation and its severity, which can vary considerably from one event to another.

To enhance the model's precision, we defined the dependent variable, "occurrence of an accident", as a conditional statement. Accidental cases are identified by the presence of crashes, either at the midblock or intersection, during a snowstorm event. In the absence of such crashes, the event is classified as non-accidental. Analyzing both intersection and midblock crashes is crucial for understanding how varying weather conditions impact the likelihood of crashes at different road sections. Typically, intersections experience a higher frequency of crashes, largely due to the presence of traffic controls. In contrast, midblock crashes provide insights into accidents occurring under steady traffic flows.

Table 4-1 Summary of Predictor Variables

Variable Category	Variable Name	Variable Description	Data Type
Weather Variables	T	Temperature in Celsius	Continuous
	H	Humidity in %	Continuous
	P	Precipitation in mm	Continuous
	W	Wind speed in km/h	Continuous
	D	Duration of snowstorm in hours	Continuous
	I	Intensity of precipitation in mm/h	Continuous
	T15	Was the average snowstorm temperature below -15 Celsius?	Discrete
	Dew	Dew point temperature	Continuous
	BI	Was the average temperature during a snowstorm within two degrees of the dewpoint?	Discrete
Maintenance Status Variables	AI	Was anti-icing conducted?	Discrete
	Maint	Was the road plowed?	Discrete
Road Surface Variables	G	Pavement friction	Continuous
Road Type Variables	Arterial	Is the road segment arterial?	Discrete
Traffic Exposure Variables	Length	Length of the road segment in km	Continuous
	ADT	Average daily traffic volumes for the road segment	Continuous
	MVK	Million vehicle kilometers	Continuous

Table 4-1 details the 16 predictors used to predict collision likelihood. Descriptive statistics reflecting the diversity and variability of the datasets are provided in **Table 4-2** From temperature (T) ranging from -31.05 to 1.85 degrees, humidity (H) from 58.8% to 94.66%, and various other measures including road type, maintenance status, and traffic exposure, the datasets showcase a

wide span of conditions. Standard deviation values further underline the variability within these parameters, emphasizing the comprehensiveness of the data. This comprehensive data collection underscores the study's potential to provide deeper insights into the factors affecting road safety in winter conditions.

Table 4-2 Descriptive Statistics of the Dataset

	Min	Mean	Max	Std	25%	50%	75%
T	-31.05	-17.09	1.85	7.36	-22.04	-18.53	-10.82
H	58.8	76.56	94.66	8.95	70.6	76.6	83.55
P	0	1.93	11.8	2.97	0	0.2	2.9
W	1.75	10.68	21.64	4.46	7.44	10.76	13.94
D	4	16.45	98	15.67	4	11	25
I	0	0.10	0.36	0.11	0	0.05	0.19
T15	0	0.61	1	0.49	0	1	1
Dew	-33.78	-20.48	-6.58	7.58	-25.93	-21.63	-14.11
BI	0	0.14	1	0.35	0	0	0
Arterial	0	0.74	1	0.44	0	1	1
AI	0	0.14	1	0.35	0	1	1
Maint	0	0.35	1	0.48	0	1	1
G	0.2	0.45	0.82	0.14	0.33	0.44	0.54
Length	0.83	3.32	7.3	1.64	1.82	3.07	4.28
ADT	1433	25339.45	74523	23406.75	5922	20198	43015
MVK	0.95	37.14	170.09	47.13	5.24	20.23	50.01

4.2 Binary Classification Model Development

4.2.1 Classification Model Hyperparameters Calibration

A critical step in the SVM model development is selecting the right Kernel to transform the dataset. As mentioned previously, there are several ways of performing this operation. In this study, we focused on the four most popular methods: Linear, RBF, Sigmoid, and Polynomial. In addition to finding the most suitable Kernel, a regularization parameter “C” must also be tuned. The significance of the “C” parameter is that it controls the size of the soft margin or tolerance for misclassification, which directly impacts model accuracy [89]. For the 'rbf', 'sigmoid', and

'polynomial' kernels, gamma plays a crucial role in defining the decision boundary. In this study, utilizing the python package 'sklearn', the gamma setting of 'auto' denotes that its value is automatically determined as the inverse of the number of features in the input data (1/n_features).

To find the most optimal model, a grid search was performed for each of the four Kernel functions to identify the best "C" value between 1 and 101. The evaluation process involves training the model using 80% of the data and validating it using the remaining unseen 20%. The performance accuracy captured using the unseen 20% validation data is used for the comparison.

By following the process mentioned above, we identified the best-performing model for each Kernel. We document this in **Table 4-3**.

Table 4-3 Comparison of SVM Prediction Performance Using Different Kernels

Performance Metrics				
Kernel	Linear	RBF	Sigmoid	Poly
C	1	16	72	5
Gamma	N/A	Auto	Auto	Auto
Accuracy	79.0%	87.2%	87.2%	83.0%
Recall	33.3%	80.0%	63.6%	60.0%
Precision	11.1%	44.4%	77.8%	33.3%
F1-Score	16.6%	57.1%	70%	42.9%

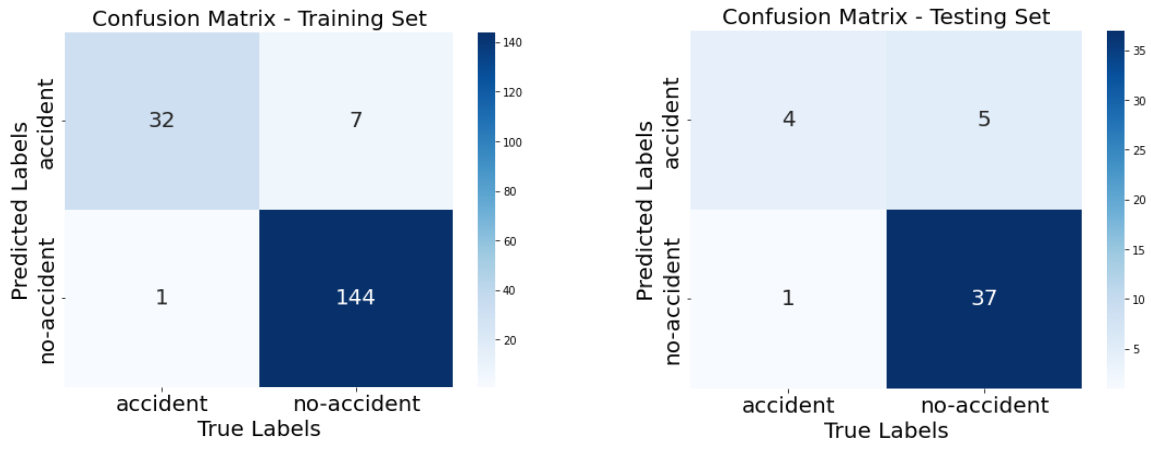
Table 4-3 highlights that the RBF Kernel outperformed all other approaches. It delivers the best performance across all metrics, achieving accuracy at 87.2%, precision at 44%, recall at 80%, and an F1-Score of 57.1%. To understand these performance metrics: accuracy is derived using the ratio of correctly predicted instances to the total instances; recall is the proportion of actual positives correctly identified; precision is the proportion of positive identifications that were actually correct; and F1-Score is the harmonic mean of precision and recall. From this table, it is observed that the Sigmoid Kernel is performing with the same model accuracy with a slightly lower recall to compensate for a higher Precision and F1-Score.

4.2.2 Classification Model Performance

To further evaluate the performance of the SVM (RBF-based) model, a confusion matrix, shown in **Figure 4-2**, is used to determine the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). As shown below, the model has a high classification in

the accident and non-accident categories, with an accuracy of 87.2%. Moreover, of the remaining 12.8% of misclassifications, the vast majority are FPs, i.e., our model is rather conservative, which can be beneficial from a safety point of view. Although these misclassifications are not ideal, the consequences associated with these errors are less severe.

Concerning the Sigmoid Kernel, the reduction in recall makes the model less appealing compared to the high recall obtained in the RBF Kernel. Moreover, the optimum 'C' value observed in the Sigmoid Kernel is significantly higher than the fine-tuned RBF Kernel, which poses a significant risk of overfitting with the Sigmoid Kernel. Thus, considering high recall and lower C value, the RBF kernel is selected for our SVM model.



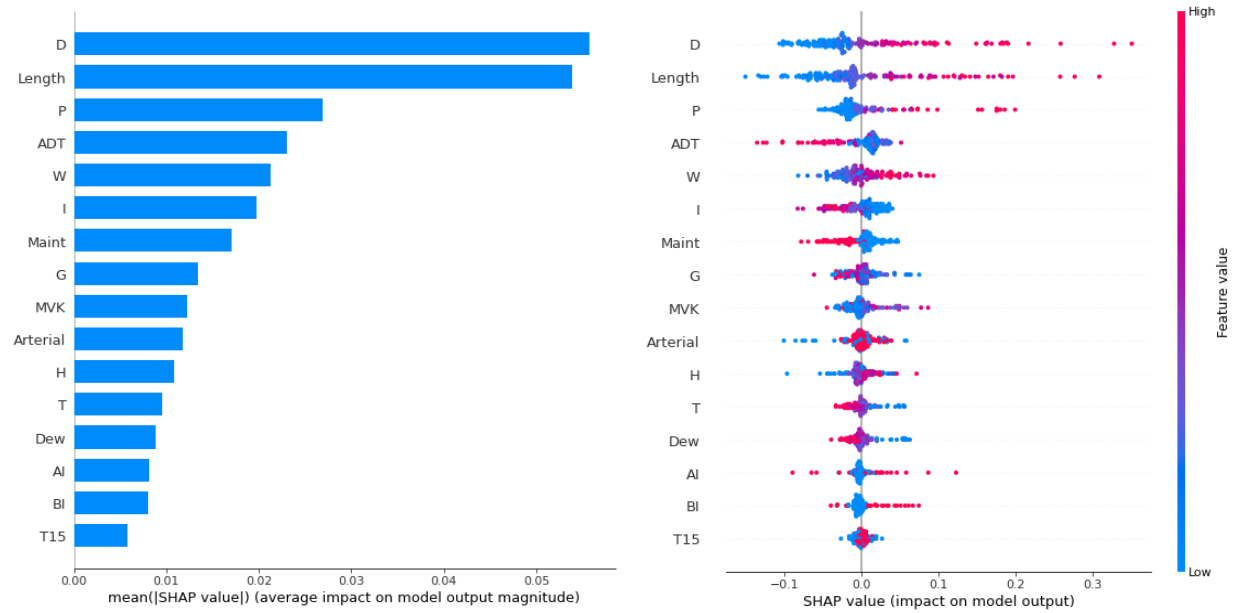
(a) (b)

Figure 4-2 (a) Training Set Confusion Matrix and (b) Testing Set Confusion Matrix

4.3 Global Interpretation of Variable in the Crash-Inducing Snowstorm Events Model

Recall that the most prominent issue associated with ML models is that they are opaque in their decision-making process, which is evident in our developed SVM model. Although the model that we developed can identify road segments where collisions are likely, the model does not tell us much about why accidents are more likely to happen on this road segment. As a result, it is difficult to gauge the intuitiveness of the model and to pinpoint the actions needed to improve road safety.

The SHAP interpreter allows the user to assess the importance of the predictor features. By using SHAP, we can identify the factors responsible for crash-inducing snowstorm events, allowing decision-makers to develop remediation procedures that target the identified factors. There are two ways to visualize feature importance using SHAP; one is through the feature importance diagram (**Figure 4-3 a**), and the other is through the SHAP summary plot (**Figure 4-3 b**).



(a) (b)

Figure 4-3 (a) SHAP Feature Ranking Plot. (b) SHAP Summary Plot.

Based on **Figure 4-3 (a)**, the two most important predictors for collisions are snowstorm duration (D) and road length (Length); their importance is more than double that of the next most important feature. Many severe winter-related car crashes have shown that a longer road can induce fatigue or stress. Moreover, the severity and prolonged snowstorms can be problematic as well [90]. Weather forecasts often discourage driving on snowy days since snow can reduce visibility and create slippery road conditions. This issue poses a significant challenge to snowplowing operations.

Apart from these two predictors, other factors such as precipitation (P), Average Daily Traffic (ADT), wind speed (W), and intensity of precipitation (I) play a significant role in predicting accident occurrence. In winter cities like Edmonton, high wind speed can affect the stability and control of vehicles. In addition, the intensity of precipitation can often impair drivers' visibility and amplify the effects on traffic flow which in turn increases the crash rate [91, 92]. Besides,

research has shown that there is a higher relative accident risk associated with intensive precipitation or slippery roads [93]. This is often because wet and icy conditions reduce pavement friction, resulting in a decrease in tire-pavement grip. Moreover, ADT serves as an additional indicator, reflecting traffic volume and density where increased vehicle interactions often lead to higher risk.

Another notable finding is the relative importance of plowing operations (Maint) over anti-icing (AI). Both of these are common responses to winter weather conditions, but our findings suggest that plowing operations are more effective, possibly due to infrequent deployment of anti-icing operations in the city. This finding aligns with previous research on mitigating snowstorm accidents, which emphasized the importance of proactive winter maintenance during severe snowstorm events [91]. Friction value (G) was also found to be significant with moderate importance, further supporting our previous findings that G directly affects collisions [6].

The remaining variables have relatively lower contributions. In descending order of significance, they include million vehicle kilometers (MVK), arterial road (Arterial), humidity (H), ambient temperature (T), dew point temperature (Dew), black ice (BI), and ambient temperature below -15 degrees (T15).

Similar to the feature importance diagram, the SHAP summary plot **Figure 4-3 (b)** also provides the same feature importance ranking on left side of the plot. In this case, a larger spread of the red and blue dots highlights the importance of a feature. Besides indicating feature importance, the SHAP summary plot shows the directionality of the predictors to the outcome based on the SHAP value, determining whether a predictor contributes positively or negatively to a prediction. Taking the snowstorm duration (D) feature as an example, blue dots are located on the left side of the vertical axis with negative SHAP value, indicating that low values of duration contribute to the prediction of “no-collision”. We concede that we would reach the same conclusion on the feature importance by focusing only on the size of the spread in each predictor. However, the value of this diagram lies in its ability to clearly show how these features contribute to the prediction analysis. The concentration of blue dots on the left side of the graph for D, Length, and P makes it evident that lower values of snowstorm duration, road length, and precipitation are more beneficial for

road safety. Conversely, it also shows that a high true value of these three features creates conditions for road accidents.

For ADT, W, and I, high wind speed is shown to lead to collisions, whereas a higher ADT and precipitation intensity lead to a lower likelihood of collisions. The issue with this observation is that the relationship found for I and ADT is unintuitive. This is perplexing as research shows that ADT is positively correlated with collision frequency, which is logical because high ADT means that there are more interactions between vehicles. Likewise, higher precipitation intensity should lead to more contaminated road segments. The presence of water, snow, and ice should increase the likelihood of collisions due to the loss of friction. Comparatively, precipitation (P), which should be positively correlated with the intensity of precipitation, behaves as expected, with lower P leading to lower collision probability. When looking at maintenance (Maint) and friction (G), a similar pattern emerges where densely accumulated low feature values appear on the right side of the plot. This indicates that increased road friction and active road maintenance contribute to a reduced accident risk, while findings from MVK also demonstrate that reduced traffic exposure is associated with a decrease in the likelihood of collisions. On the other hand, Arterial feature behaves in a mixed manner where SHAP is close to zero. However, the blue dots tend to fall to the left with lower SHAP values. This inclination means that non-arterial roads often lower collision odds.

Moving along to additional features with lower feature significance, it is observed that high humidity (H) tends to increase accident rates. This is because humidity can lead to slippery roads due to a loss of friction. Dew point temperature and ambient temperature displayed a similar pattern, indicating lower values make the road more prone to accidents. Since low temperature is associated with severe weather conditions, the obtained pattern is reasonable. In terms of the remaining variables, AI and BI, it's difficult to interpret them due to red and blue dots appearing on both sides of the graph, meaning their importance varies depending on the instance in question. Finally, it appears that both blue and red dots of T15 are primarily concentrated around zero SHAP value, indicating a lower level of influence for this specific feature.

While a summary plot offers valuable global insights, it may present some counterintuitive feature relationships such as those observed with ADT and precipitation intensity. These can arise due to

the SVM model's sensitivity to outliers [94] or the small sample size of our dataset, which may not fully capture the complex, nonlinear interactions in the data. Moreover, the nature of SHAP's interpretation, grounded in cooperative game theory, inherently assumes a linear additive contribution of features [64], which may not always align with the nonlinear decision boundaries formed by SVMs.

4.4 Interpretation of Representative Crash-Inducing Snowstorm Events

Up to this point, we used SHAP to provide global explanations for our developed SVM model. In Section 4.4, SHAP waterfall plots are employed to further explain individual instances. This local evaluation allows for a more detailed analysis, complementing the broader model tendencies presented in Section 4.3, and is crucial for confirming the model's applicability in practical scenarios and informing future enhancements. We note that the presence of certain outlier instances may complicate the explanations provided. For example, we observed that ADT and intensity of precipitation displayed an illogical relationship with collision occurrence.

To make sense of this apparent contradiction, a SHAP waterfall plot was employed, allowing us to investigate individual instances. Since SVM provides confidence probabilities for the predictions made, we selected the top four most confident predictions (True Positives). The choice of these instances ensures that the analysis focuses on the most statistically significant predictions, thereby reflecting the core behaviors of the model in a manageable and coherent way. The prediction probabilities of the top four instances are 84.94%, 71.91%, 69.93%, and 69.26%. Correspondingly, **Figure 4-4** presents four waterfall plots arranged by descending prediction confidence. To illustrate this, **Figure 4-4 (a)** corresponds to the instance with an 84.94% prediction probability, while **Figure 4-4 (d)** represents the 69.26% probability instance. Since we are dealing with a single instance, each plot represents the contribution of a specific feature value to the final prediction. For example, in **Figure 4-4 (a),** the first row, P=11.8, indicates that the precipitation feature value is 11.8 for this particular instance. The red arrow on the right indicates that precipitation contributes positively to collision occurrence. The magnitude of the bar shows that it is the most important predictor. Had the color of the bar been blue, this would have indicated that the feature contributed negatively to the final prediction.

When looking at the four plots, ADT appears to contribute positively in instances where it is an impactful variable. In **Figure 4-4 (a)** and **(d)**, ADT nudges the prediction towards "accident", which aligns with our previous findings. However, **Figure 4-4 (b)** ADT appears to be counterintuitive where it shows a blue bar with a high feature value, which means that a high ADT negatively contributes to collision odds. This conflicts with the previous observation. Such discrepancies may originate from the use of limited datasets or the inherently local interpretability of SHAP. Lundberg and Lee (2017) have noted that using Kernel SHAP could potentially violate its local accuracy or consistency, which in turn could lead to unintuitive behaviors. Other researchers also identified a propensity for the Kernel SHAP approximation method to overlook feature dependence due to its reliance on marginal expectation [95].

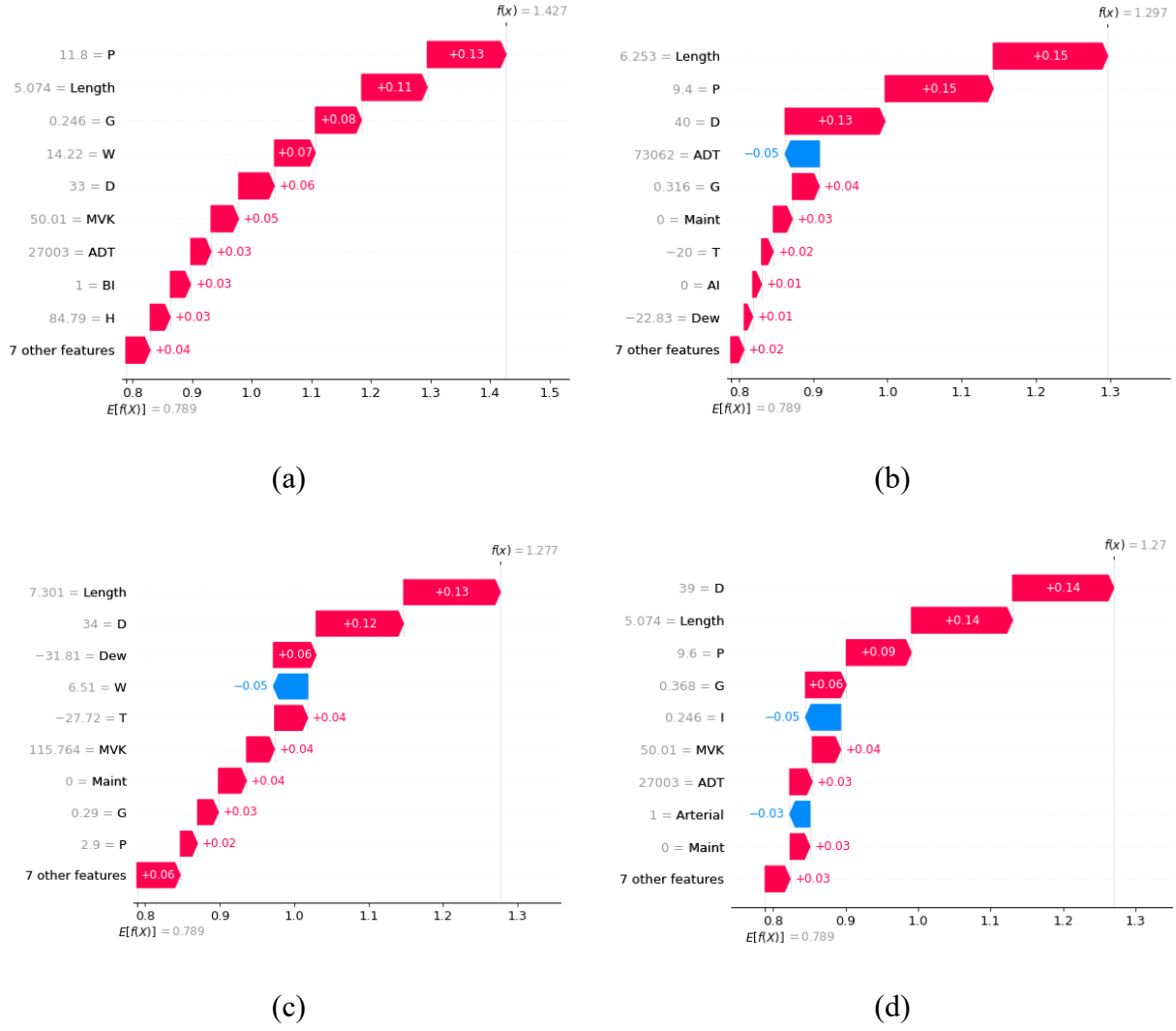


Figure 4-4 SHAP Waterfall Plots. Top 4 Instances Based on High Prediction Probability from True Positive. (a) Instance 65, (b) Instance 23, (c) Instance 51, (d) Instance 128.

Regarding intensity of precipitation (I), it is featured in only one of the four instances depicted in **Figure 4-4 (d)**, where an intensity value of I=0.246 contributes negatively, suggesting that higher precipitation intensity may reduce the probability of collisions. This is counterintuitive, given that high-intensity precipitation leads to dangerous driving conditions. These factors should increase the risk of a collision. Similar to ADT, the counterintuitive finding regarding precipitation intensity further highlights the limitations of the SHAP analysis in accounting for the complexities of real-world conditions that affect collision probabilities. This suggests that SHAP values should be interpreted with caution, especially when they contradict widely accepted phenomena.

Another notable observation is the contribution of friction. The waterfall plot shows how friction (G) contributes to collision prediction. Among all four cases, it is consistently shown that a low friction value increases the collision odds, further supporting our previous findings.

In summary, waterfall plots are useful for understanding the behavior of the model, especially when global explanations do not adequately clarify feature contributions. This visualization technique provides a detailed view into the inner workings of complex ML models, enabling a thorough exploration of the influence of each feature. This then allows users to validate model predictions with their domain expertise. Nonetheless, it is important to recognize the limitations that may arise from the dataset's size or Kernel SHAP's oversight of feature interdependence. These limitations can result in counterintuitive conclusions that do not fully capture the complexity of the underlying data.

Chapter 5: EVALUATING WINTER CRASH FREQUENCIES USING MACHINE LEARNING TECHNIQUES

The previous chapter explored the crash-inducing snowstorm events characteristics and how weather features act differently in specific instances of crash-occurred snowstorm events. The interplay among the harsh weather variables illustrates the significance of active winter road maintenance. With the pattern of the leading significant risk factors in mind, maintenance personnel can effectively use available resources to mitigate the storm impact to the road users.

In this chapter, the focus is now on evaluating the general winter crash behaviors at high and low crash frequency locations. To understand this, this chapter has been broken down into five sections. The first section composes the discovery of the winter crashes associated risk factor from both the micro and macro level.

In the second section, the winter crashes are analyzed using Hot Spot Analysis (HSA) in ArcGIS, a spatial analysis tool that helps locate crash prone regions in the city. A preliminary assessment was conducted to investigate the feature behaviors at hot spots (HS) and cold spots (CS). To achieve this, we perform an ArcGIS visualization by overlaying the features on top of the identified hot and cold spots.

Moving on to the third section, with the significance of the features in mind, the crash frequency machine learning (ML) model is then built using the cleaned and processed data. Three widely adopted tree-based ML models are chosen for a comparative assessment of their performance. Upon the completion of the training for each model, they are then evaluated based on their explained variance and error metrics.

The fourth and fifth sections leverage the power of XAI to enhance the understanding of the ML model decision making progress. In the fourth section, the global interpretation is first examined to quantify the impact of the feature values for different risk factors towards the model prediction. The global explanation gives an idea of the key features that contribute most to the model outcome, but it does not provide any explanations to the outcome of each instance. Hence, in the fifth section, the SHAP local interpretation are adopted to discover the interplay of the features at high and low

crash locations, enhancing the understanding of the similarities and differences between these two types of crash locations.

5.1 Data Processing

Similar to the previous chapter, this chapter also presents the development of an ML model that improves winter road safety. According to the city of Edmonton's weather data report, the annual snowy period often begins from November to March of the following year [96]. Hence the winter crash frequency model is developed based on this duration for four years of the collision dataset, from 2016 to 2020.

The study initially starts with an exploration on the winter crashes related risk factor in addition to the previous chapter identified weather variables. To conduct a thorough and comprehensive analysis on the ML model, additional features are further explored at the micro and macro level. The micro level features crash record information and estimated traffic volume are incorporated with previously identified weather variables. The macro level features involve the spatial variables and land use variables. Such a combination between the micro and macro level elements has proven effective in improving the model performance and provides a holistic understanding of the behavior of the risk factors [97].

To understand the winter crashes from a spatial context, HSA is conducted to observe the crash prone locations in the city. Acknowledging locations of these HS and CS, we can better understand the feature involvement in these locations by overlaying the investigated features on top. This helps us in conducting the preliminary analysis on the significance of the features. After removing the missing information and cleaning the feature data, the total number of crashes we analyzed was 26,970, which are later used for developing a ML-based predictive model.

5.1.1 Micro-Level Variables

In terms of the micro-level variables, the study has mainly considered the following variables: crash record variables, weather variables, road characteristics, and traffic volume variables. The crash record data provides comprehensive information about the crash environment and collision details. These include a detailed record of the location of the crash record, the crash severity, the road conditions, relevant driver information, and the driving maneuver. With respect to the weather

variables, the data is acquired from the City of Edmonton's open data portal. These features include the ambient temperature, dew point temperature, humidity, wind speed, and precipitation. Moreover, road characteristics are also obtained to describe the geographic components of the collision. This includes the collision road type, the condition of the road surface, the function class of the road where the accident happened, the regulated speed limit, the presence of automatic traffic enforcement cameras, the presence of preventive measure medians, and the snow ice clearing priority of the route. Finally, the traffic volume is estimated using the ordinary kriging (OK) approach. Since hourly traffic volume records are unattainable, the traffic volume estimation performed in this study instead utilizes the annual average weekday traffic volume (AAWDT). While traffic volume variables could pose a limitation to the study, AAWDT often provides a reliable estimate of the overall traffic volume pattern across the entire city road network.

5.1.2 Macro-Level Variables

The forementioned micro-level variables helps capturing the specific conditions and contextual factors at the time and location of the crash. Such granular details enable the model to obtain the dynamic aspects of the crash frequency. In addition to the micro level variables, macro-level variables encompass broader contextual factors such as land use and spatial variables. These variables are essential as they often provide a boarder also wider information about the environment which crashes occur. A land use variable includes information about the surrounding area, such as residential, commercial, industrial, agricultural zones. Specific land use regulations often shape the landscape of infrastructures, which in turn affects traffic behavior and patterns. Regarding spatial variables, the analysis identifies hot and cold spots of crash-prone locations, providing a clear representation of the intensity of high and low crashes clusters.

5.1.3 Traffic Volume Estimation

The traffic volume variable is a critical component in transportation planning. It often serves as a key indicator to estimate traffic exposure and provides insights for city planners to evaluate crash locations. However, collecting sufficient traffic volume data across an entire city is challenging, as it often demands significant manpower and infrastructure development [98], which can be costly [99]. Due to the limited coverage of traffic counters in the city, this study employed OK to estimate traffic volume from 2016 to 2020 for all winter crashes.

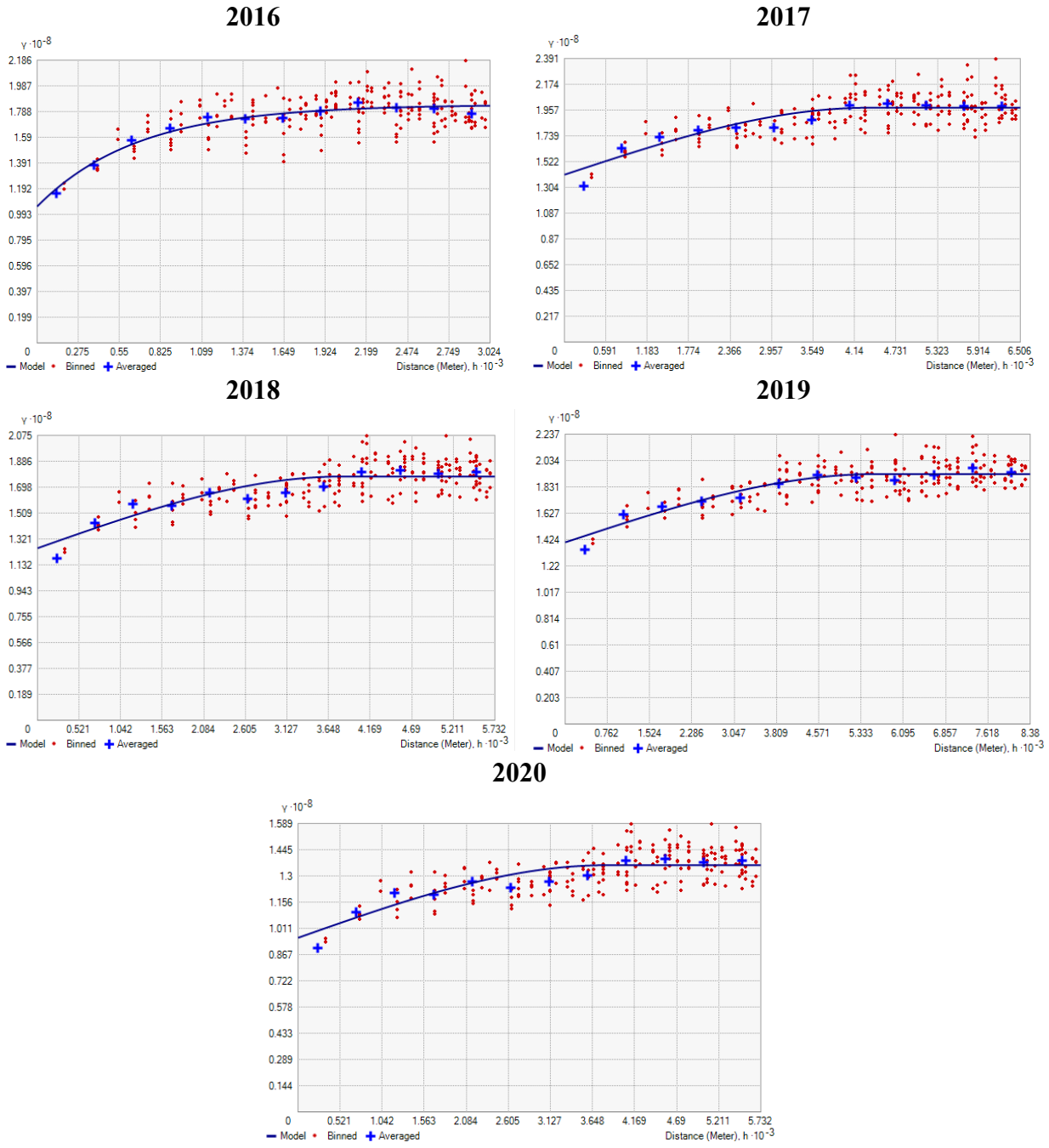


Figure 5-1 Semivariogram from 2016 to 2020

OK is a widely adopted geostatistical interpolation technique based on the assumption of stationarity, meaning that the variance and mean are constant. In this study, the semivariogram was constructed using the ArcGIS geostatistical analyst tool. The generated semivariograms, showcased in **Figure 5-1**, help determine the spatial autocorrelation of the interpolated data points

and the accuracy of the estimated AAWDT results. Three commonly used functional forms are compared when modeling the empirical semivariogram: Spherical, Exponential, and Gaussian. To validate the estimation performance of these functional forms, two statistical measures; namely, Mean Standardized Error (MSE) and Root Mean Square Standardized Error (RMSSE), are evaluated. The closer the MSE is to 0 and the RMSSE is to 1, the higher the accuracy of the estimated results. **Table 5-1** provides a summary of these metrics for the years 2016 to 2020, demonstrating the model's good overall accuracy.

Table 5-1 Summary of Yearly AAWDT Semivariogram Models

Year	Model	Partial Sill	Range	Nugget	MSE	RMSSE
2016	Exponential	78,743,419	2,015.670	105,561,048	0.00425	1.12409
2017	Spherical	56,482,751	4,337.141	141,224,344	-0.00030	1.09708
2018	Spherical	52,194,757	3,821.394	125,405,395	-0.00081	1.09807
2019	Spherical	52,689,223	5,586.850	140,297,708	0.00027	1.08134
2020	Spherical	39,984,723	3,821.394	95,997,382	-0.00080	1.09827

5.1.4 Descriptive Statistics for All Features

Table 5-2 presents an illustration of the 22 features involved in training a crash frequency model. It categorizes these features into micro-level and macro-level variables, providing descriptions and corresponding data types for each feature. **Table 5-3** also provides detailed descriptive summary statistics of these predictors, showcasing the diversity and variety of features. Incorporating both macro and micro-level variables, it enables a thorough understanding of winter crashes, facilitating better resource allocation during the long, harsh winter season in city of Edmonton. The crash count, which serves as the dependent variable, is determined using ArcGIS Collection Event tool, as discussed in the HSA before conducting the preliminary assessment.

Table 5-2 Variable Descriptions

Variable Category	Variable Name	Data Type	Variable Description
Micro Level Variables			
Collision Record	FAT_AND_INJ	Binary	Fatal injury crashes (2) or PDO (1)
	AGE	Numerical	Driver's age
	ROAD_TYPE	Binary	Intersection crash (1) or midblock crash (0)
	ROAD_SURFACE	Ordinal	Dry (3), wet (2), icy (1)
	VLTN_IMP_MANEU	Binary	Improper driving maneuver violation
	VLTN_ROW	Binary	Right of way violation
	VLTN_LOC	Binary	Loss of control violation
Weather Variables	Temperature	Numerical	Ambient temperature (°C)
	Dew	Numerical	Dew point temperature (°C)
	Humidity	Numerical	Humidity (%)
	Wind_Speed	Numerical	Wind speed (km/h)
	Total Precipitation (mm)	Numerical	Average daily precipitation (mm)
Road Characteristics	Function_Class	Ordinal	Road function class: least (0) to most important (4)
	Speed_Limit	Numerical	Speed limit (km/h) at crash location
	ATE	Binary	Traffic enforcement camera presence (within 150 meters)
	Snow_Ice_Clearing_priority	Ordinal	Route clearing priority: most (1) to least important (4)
	Median	Binary	Presence of median island
Traffic Exposure Variable	AAWDT	Numerical	Estimated average annual weekday traffic volume
Macro Level Variables			
Land Use Variables	LU_Residential	Binary	Residential land use
	LU_Commercial	Binary	Commercial land use
	LU_Industrial	Binary	Industrial land use
Spatial Variables	Gi_Bin	Ordinal	Spatial bins: +3/+2/+1 (hot spots), -3/-2/-1 (cold spots), 0 (normal)

Table 5-3 Variable Summary Statistics

Variable Name	Min	Mean	Max	Std	25%	50%	75%
FAT_AND_INJ	1	1.1	2	0.3	1	1	1
AGE	14	39.49	102	16.26	26	36	50
ROAD_TYPE	0	0.71	1	0.45	0	1	1
ROAD_SURFACE	0	1.61	3	0.88	1	2	2
VLTN_IMP_MANEU	0	0.63	1	0.48	0	1	1
VLTN_ROW	0	0.17	1	0.38	0	0	0
VLTN_LOC	0	0.08	1	0.27	0	0	0
Temperature	-37.6	-8.52	19.4	9.46	-15.6	-7.3	-0.6
Dew	-41.3	-12.74	7.4	9.18	-19.6	-11.4	-5.3
Humidity	6	68.84	100	20.44	61	72	81
Wind_Speed	0	10.46	51	6.41	6	10	14
Total Precipitation (mm)	0	0.72	9	1.49	0	0.6	1
Function_Class	0	1.52	4	1.19	1	2	2
Speed_Limit	20	52.34	110	12.63	40	50	60
ATE	0	0.13	1	0.33	0	0	0
Snow_Ice_Clearing_ priority	1	1.62	4	1.08	1	2	2
Median	0	0.55	1	0.5	0	1	1
AAWDT	475.339	13646.751	47469.834	5662.527	9513.287	12995.011	17087.622
LU_Residential	0	0.34	1	0.47	0	0	1
LU_Commercial	0	0.2	1	0.4	0	0	0
LU_Industrial	0	0.03	1	0.17	0	0	0
Gi_Bin	-3	0.75	3	1.42	0	1	1

5.2 Preliminary Feature Assessment for Hot and Cold Spots

As described earlier, a spatial analysis is first conducted to evaluate the characteristics of the high collision clusters and low collision clusters which is derived from HSA. To perform this task, the ArcGIS Getis-Ord Gi* method is applied. This method identifies HS and CS regions, facilitating a preliminary assessment of the features contributing to these spatial clusters.

To deploy the HSA, the four-year winter crash data is first projected on the map using the recorded longitude and latitude for each crash. The Collection Events method is then conducted to combine the coincident points with the exact same X and Y centroid coordinates [100]. This helps generate the crash counts at the same locations, which serve as the dependent variable of the study. This method converts the individual instances to weighted points, aiding the subsequent spatial analysis. Finally, the HSA is performed by specifying the following: conceptualization of spatial relationships, distance method, distance band or threshold distance, and false discovery rate (FDR). A fixed distance band is selected to define the spatial relationships between features, meaning each feature is analyzed within a consistent fixed distance. Euclidean distance is applied for the distance method to measure spatial relationships as it is a common choice for easy and straightforward interpretation. The distance band or threshold distance is not specified, as ArcGIS automatically determines an appropriate threshold distance by ensuring each feature has at least one neighbor, preventing isolated points from effecting the analysis. Lastly, the FDR option is unchecked as this serves as a preliminary assessment intended to identify the possible trends and patterns rather than to confirm definitive results, hence applying FDR is not essential. Using these settings ensures that the analysis captures meaningful spatial patterns for winter crash frequency in Edmonton.

Upon completion of the HSA, the output feature class returns z-score, p-value, and confidence level bin field (Gi_Bin). The z-scores and p-values are measures of statistical significance and the Gi_Bin field identifies statistically significant hot and cold spots. Feature in the +/- bins reflect statistical significance with a 99% confidence level; features in the +/- 2 bins reflect a 95% confidence level; +/- 1 bins reflect a 90% confidence level; and bin 0 represent statistically insignificant [101]. The visualization of the resulting HS and CS are shown in **Figure 5-2,** where the HS are shown in red and CS in blue, with deeper color representing the higher confidence.

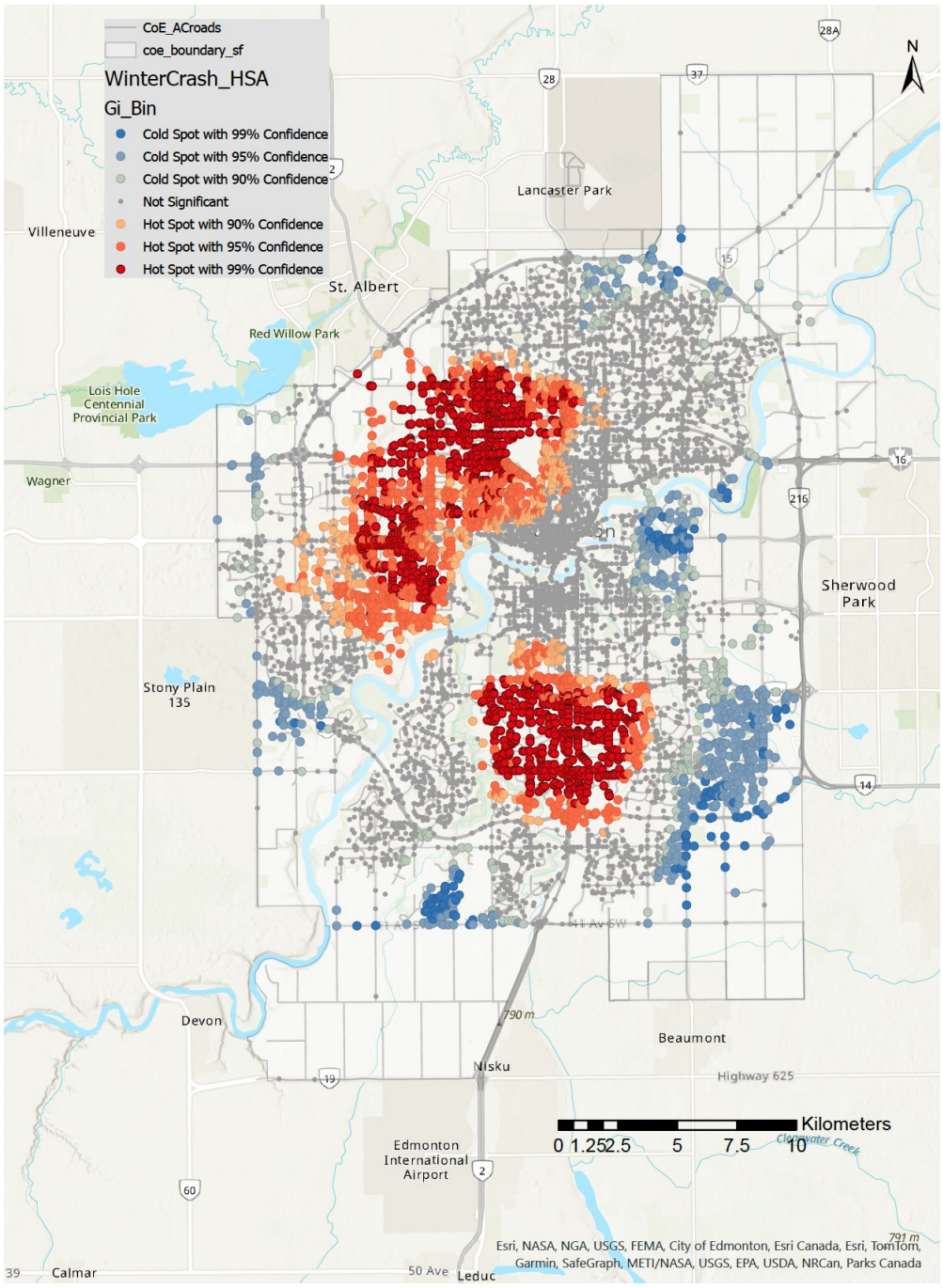


Figure 5-2 Hot Spots Analysis in the City of Edmonton

To fully comprehend this derived spatial pattern mapping, the additional analyses are performed using visually prominent land use features **(Figure 5-3)**, road characteristic features (**Figures 5-4** to **5-7**) and traffic exposure features (**Figure 5-8**) as layers on top of HS and CS to provide visual cues towards understanding the feature interactions at hot and cold spots.

Looking at the land use visualization shown in **Figure 5-3**, it can be seen that HS are mostly concentrated around the commercial zones, whereas CS fall in agricultural zones and their borders, which typically intersect with residential neighbourhoods.

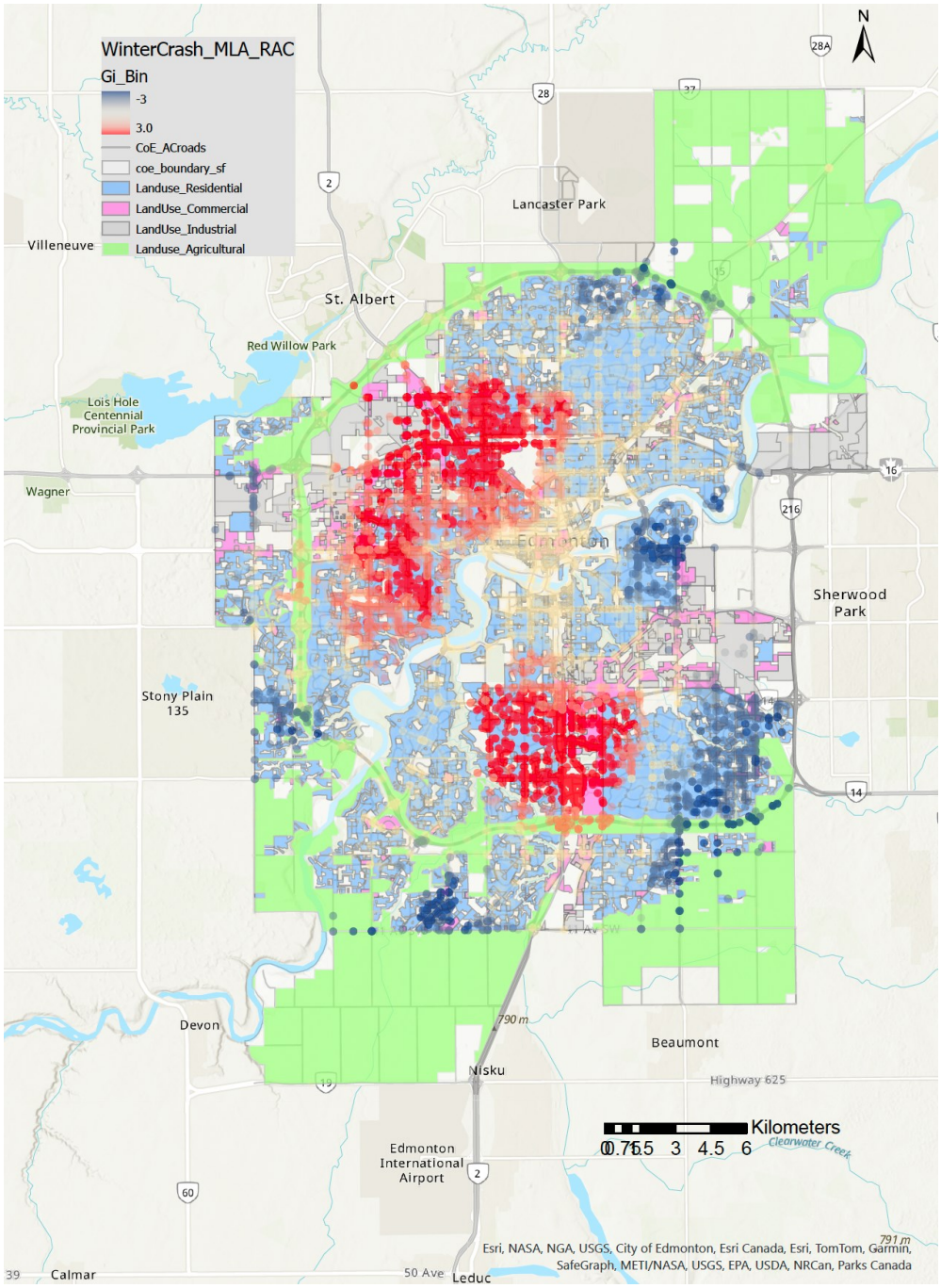


Figure 5-3 Land Use Layer on Hot and Cold Spots

In evaluating the road characteristic of speed limits (**Figure 5-4**), it is observed that in HS, there is an dense concentration of road networks with speed limits ranging from 50 to 60 km/h. Conversely, in CS, the road network appears to be less dense with lower speed limits ranging from 20 to 40 km/h.

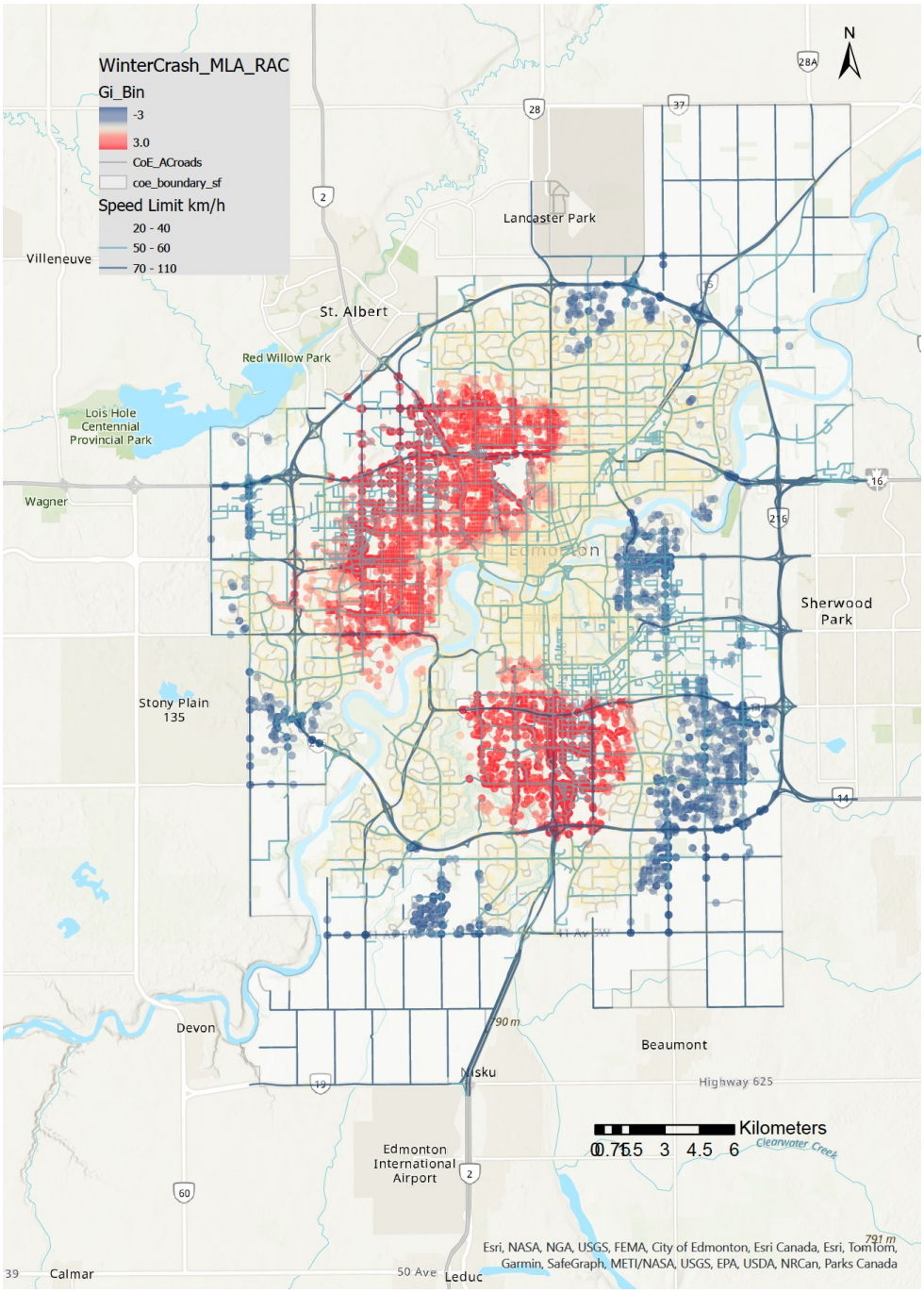


Figure 5-4 Speed Limit Layer on Hot and Cold Spots

Regarding snow and ice clearing priority routes (**Figure 5-5**), HS regions typically feature numerous intersections of high-priority snow clearing routes, along with densely packed areas of low clearing priority. Conversely, CS regions generally have fewer intersections of high clearing priority routes.

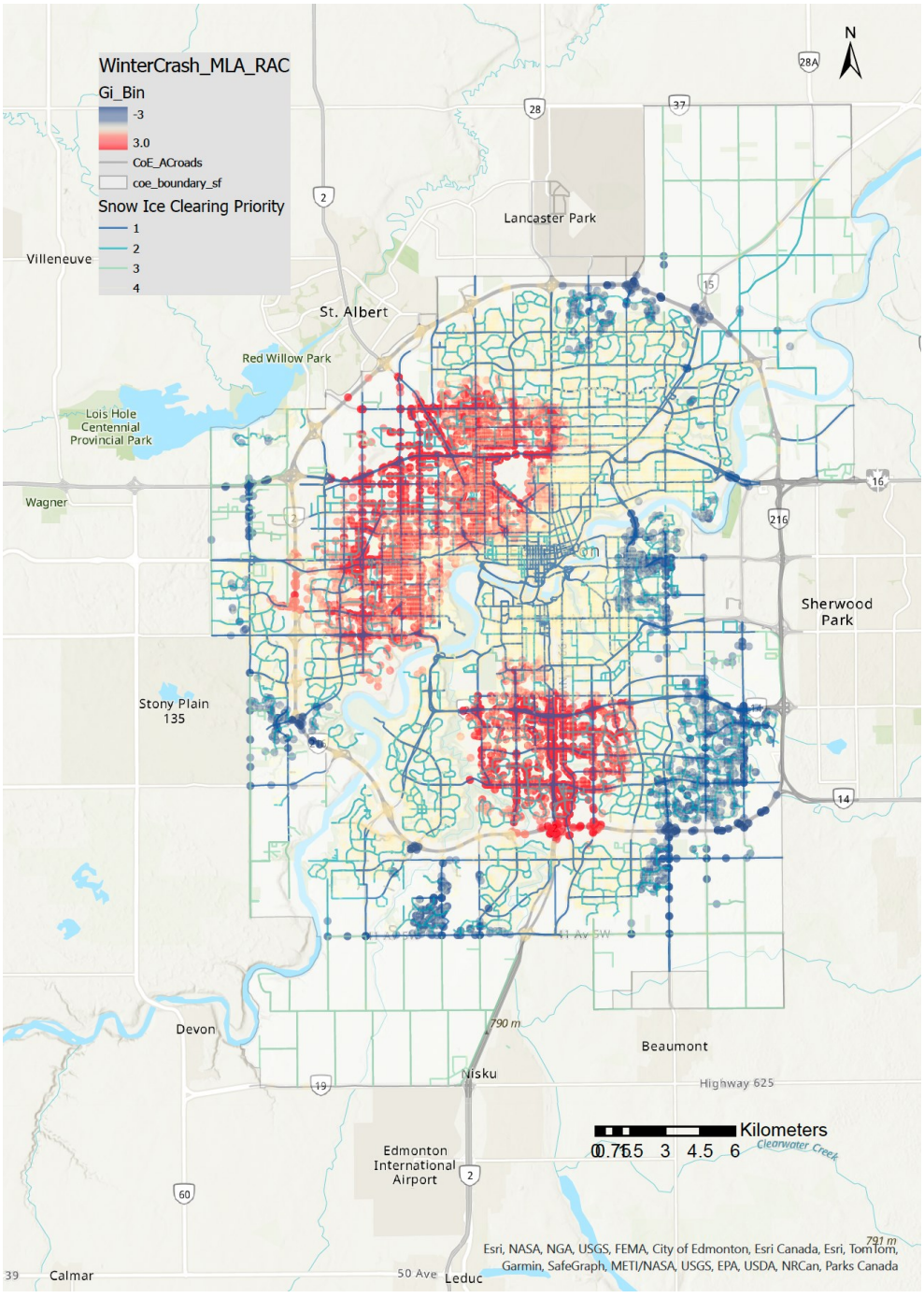


Figure 5-5 Snow Clearing Priority Route Layer on Hot and Cold Spots

Figure 5-6 illustrates the ATE placement in the city. It is observed that, compared to CS, HS have a higher number of ATE placements, indicating that the city has strategically placed ATEs to regulate and monitor high-collision locations.

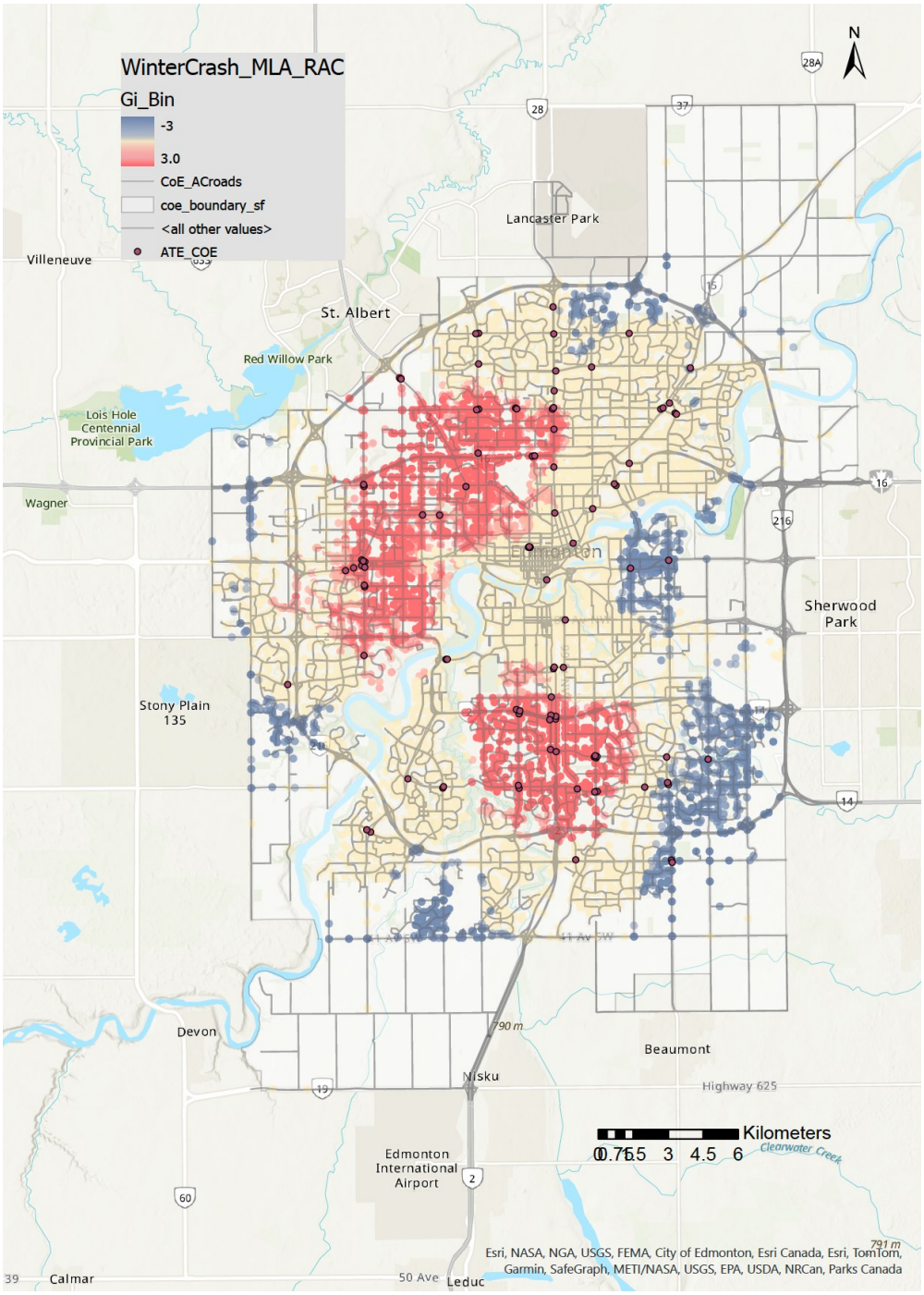


Figure 5-6 ATE Layer on Hot and Cold Spots

Figure 5-7 illustrates the road function class. It is acknowledged that higher function classes, often designed for arterial roads with high traffic demand, correspond to routes with high snow clearing priority. Therefore, a similar pattern to that observed in **Figure 5-5** can be seen.

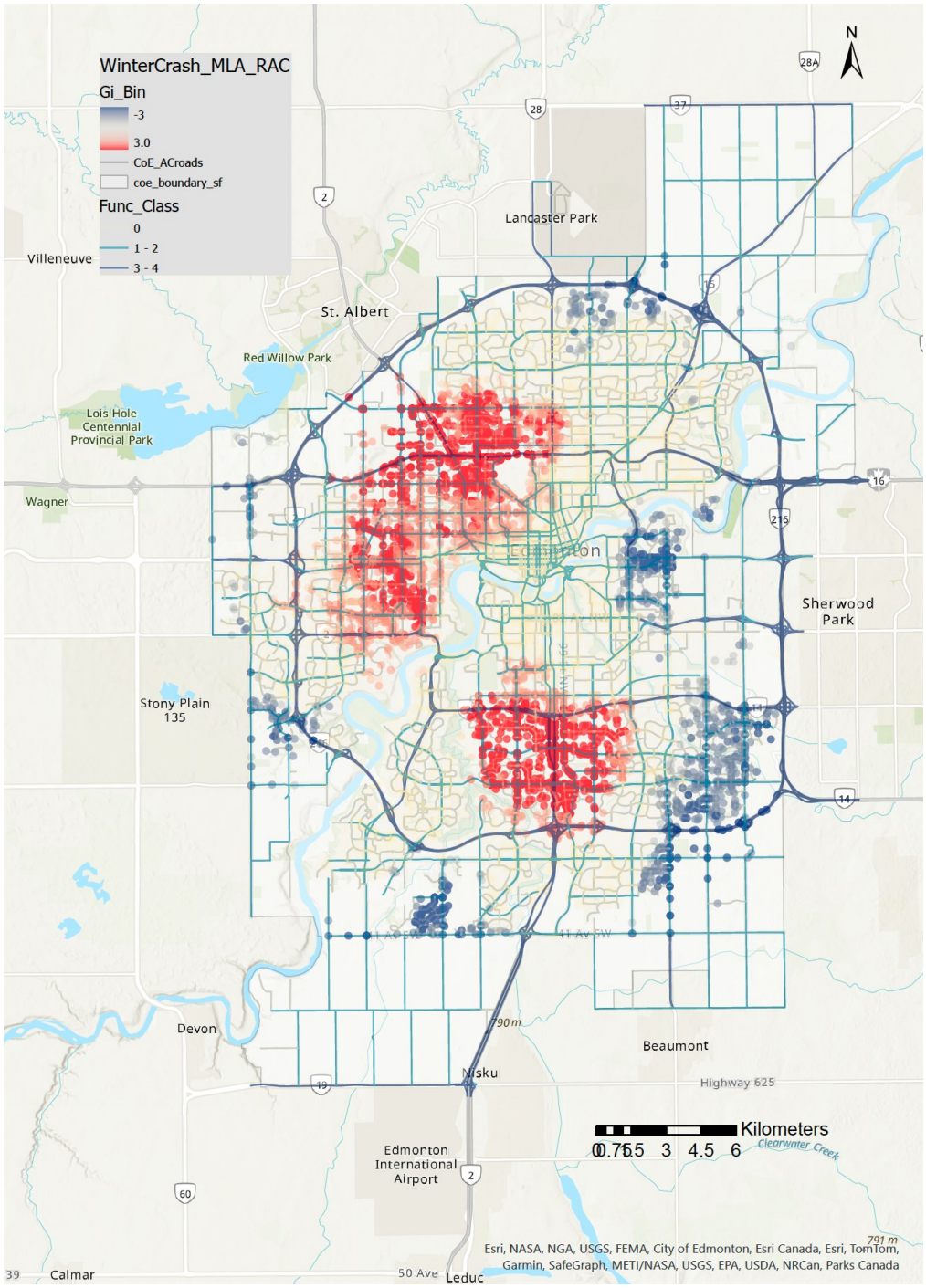


Figure 5-7 Road Function Class Layer on Hot and Cold Spots

Finally, the OK derived AAWDT traffic volume data is visually evaluated. To improve its visualization, the estimated AAWDT is converted into raster data using inverse distance weighted interpolation as shown in **Figure 5-8**. This figure highlights the traffic volume differences between HS and CS, where HS tend to have higher traffic volumes whereas CS do not.

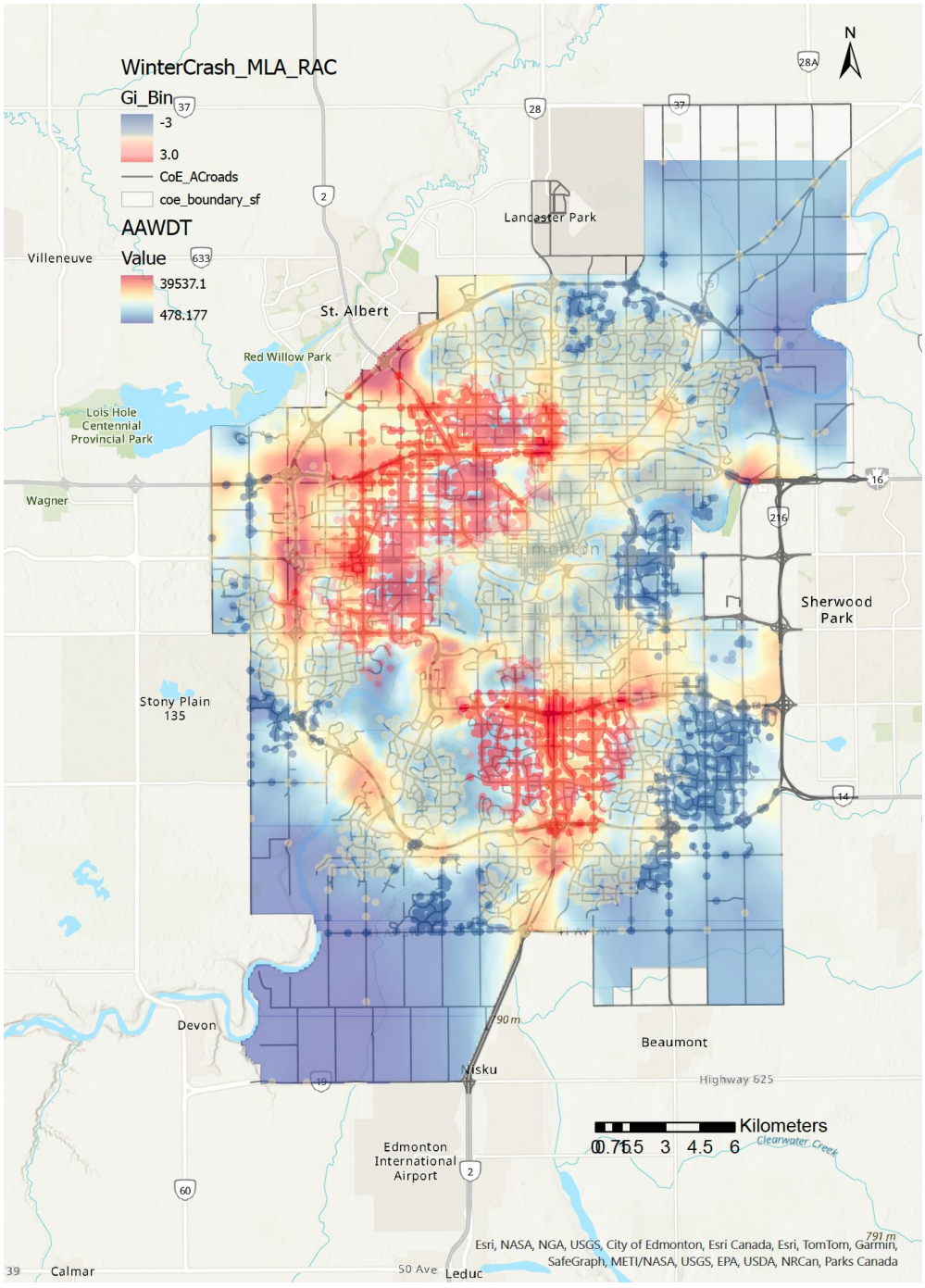


Figure 5-8 AAWDT Layer on Hot and Cold Spots

From this preliminary assessment using HSA, the feature interactions are revealed visually as demonstrated in the figures. This analysis helps in identifying the location of HS and CS regions within the city. By incorporating additional features juxtaposed on top of HS and CS, the interactions within these regions become intuitive and easier to interpret. This preliminary spatial analysis provides a fundamental understanding of the spatial patterns and interactions of various features within the clusters. These insights are critical as they form the basis for developing the subsequent ML crash frequency model, allowing for a more targeted approach in predicting and mitigating winter crashes.

5.3 Crash Frequency ML Model Development

Having visually confirmed the interactions between external features and crash distributions through spatial analysis, we now move on to the development of a crash frequency ML model. This phase involves leveraging ML techniques to predict crash frequencies based on the identified features and patterns. By incorporating these insights, we aim to build robust models that can accurately forecast crash occurrences and assist in proactive road safety management. The following sections detail the hyperparameter calibration of the ML models and compare their performance.

5.3.1 ML Model Hyperparameters Calibration

During the model training phase, the dataset was split such that 80% of the data was utilized to train the model, and the remaining 20% was used to validate the model's performance. To optimize performance, each machine learning model was fine-tuned with its hyperparameters using grid search. This approach allowed the models to learn the most from the training data. Upon obtaining the optimal hyperparameters, the models were then evaluated based on their percentage of retained variance and lowest error. The results presented below include the fine-tuned model parameters and descriptions of the hyperparameters.

For the XGBoost model, the learning rate was set to 0.05, controlling the step size the model takes towards minimizing the loss function. A lower rate results in slower convergence but can lead to better accuracy. The model max_depth was 10, helping the model capture complex data patterns, although higher value can result in overfitting. The n_estimators was 1000, indicating the number

of trees grown. More trees can improve the model's performance, but they also increase computation time. The min_child_weight was set to 3 to prevent overfitting by making the trees more robust to noise. Regularization alpha (reg_alpha) was 0.8, serving as the L1 regularization term on weights, which helps with feature selection. Regularization lambda (reg_lambda) was 1, acting as the L2 regularization term on weights, preventing overfitting by penalizing large weights.

As for the RF model, the fined tuned model shows n_estimators of 1000, and max_depth of 12. Regards to LightGBM model with num_leaves of 31, max_depth of -1, learning_rate of 0.05, n_estimators of 1000 give the prime model performance. For LightGBM, num_leaves control the maximum number of leaves in one tree. A max_depth of -1 meaning there was no limit to the depth of the trees.

5.3.2 A Comparison of ML Model Performance

The study conducted a thorough examination of three different crash frequency ML models. After fine-tuning the models, they were assessed using the R^2 value (shown in **Figure 5-9**) and error metrics (**Figure 5-10**) for both the training and testing sets across the three ML models used in our study: XGBoost, Random Forest (RF), and LightGBM. The R^2 value, also known as the coefficient of determination, measures the proportion of variance in the dependent variable that is predictable from the independent variables. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are examined using the prediction errors of the models, with lower values indicating better performance. Therefore, these key metrics are utilized to evaluate the performance of regression models.

It was observed that the XGBoost model exhibited the highest R^2 value among the three models, with a testing R^2 value of 92.67%. Following XGBoost, the RF model achieved an R^2 value of 90.96%, while LightGBM had the lowest R^2 value of 85.38%. The high R^2 value for XGBoost indicates that the model can explain a substantial portion of the variance in the complex dataset, making it a desirable choice for our predictive analysis. With respect to the MAE and RMSE, the XGBoost model also possessed the lowest error of 3.64 and 5.77, respectively. This comparison for both the R^2 value and error metrics underscores that the XGBoost emerging as the best performing model, followed by Random Forest, and then LightGBM.

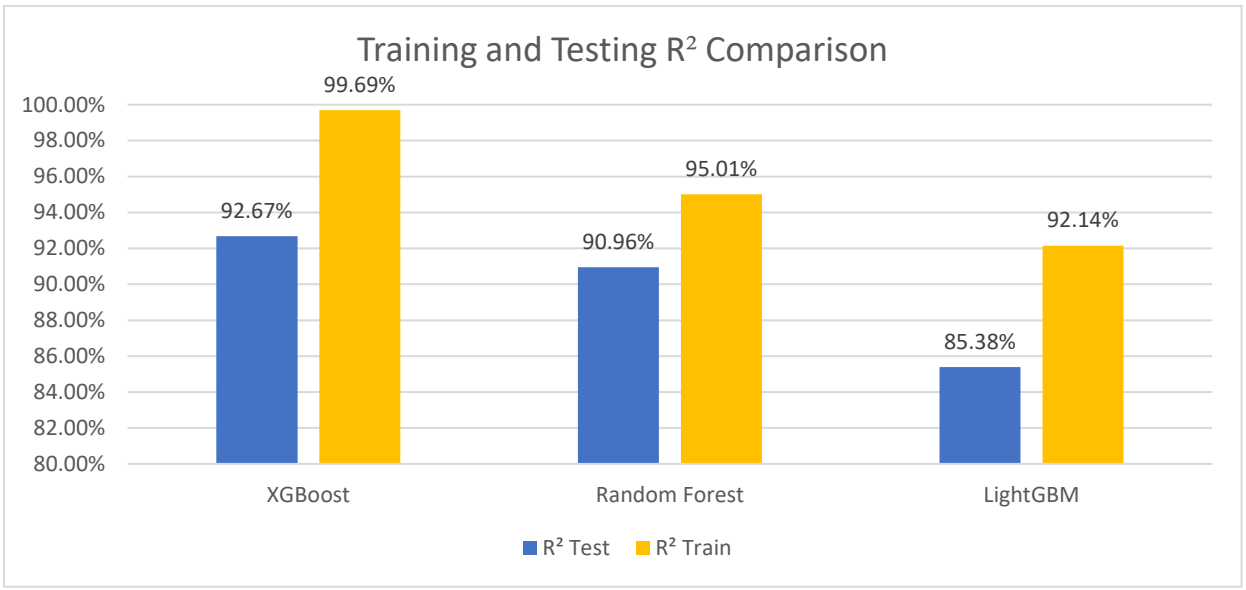


Figure 5-9 Model Training and Testing R^2 Value Comparison.

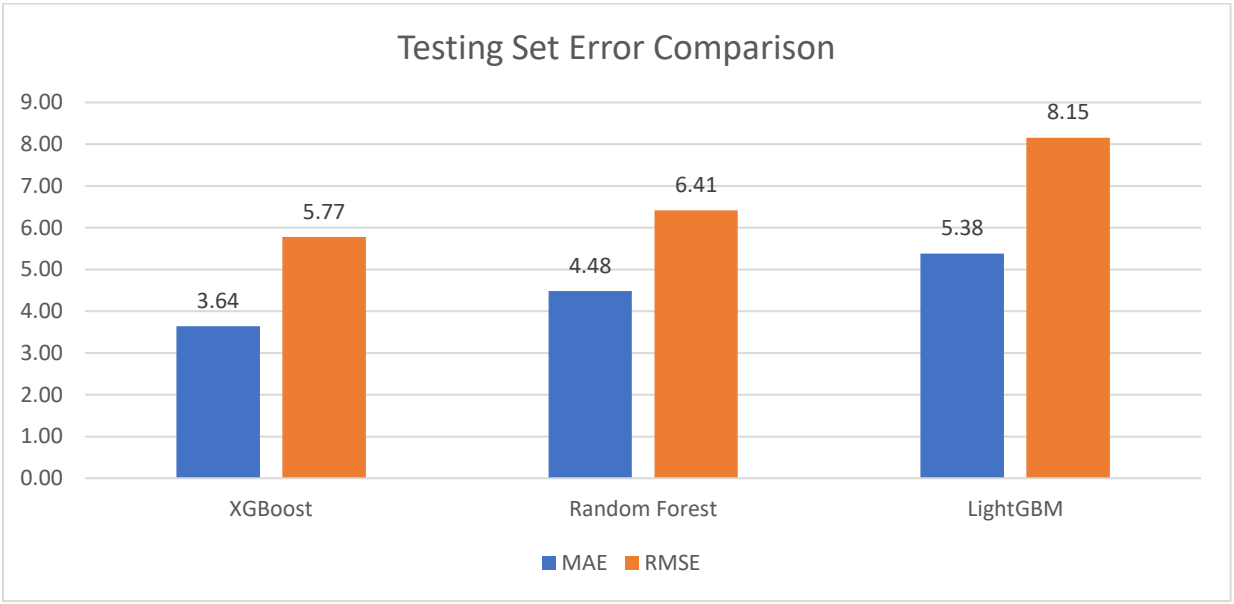


Figure 5-10 Model Testing Set MAE, RMSE Comparison

The disparity between the training and testing R^2 values suggests a potential risk of overfitting, particularly in the case of XGBoost, where the training set R^2 value reached 99.69%. Overfitting occurs when a model learns the training data too well, capturing noise along with the underlying patterns, which could lead to poorer performance on new data. While overfitting is a consideration

given that XGBoost sequentially builds models that could potentially incorporate noise from the training data, its ability to correct errors from previous iterations and capture complex patterns makes it highly effective for our dataset. The Random Forest model shows more subdued overfitting due to its method of building multiple decision trees from random data subsets, which generally improves model robustness. Despite these considerations, XGBoost, with its superior computation efficiency and accuracy, was chosen as the optimal model for subsequent SHAP analysis, balancing performance with interpretability.

Following the selection of XGBoost for further analysis due to its efficiency and accuracy, **Figures 5-11** and **5-12** provides a visual representation of the performance of the fine-tuned XGBoost model.

Figure 5-11 presents the actual crash frequency versus predicted crash frequency. In this plot, each point corresponds to a data instance from the testing set, where the actual value is plotted on x-axis and predicted value is plotted on y-axis. To aid in assessing the model's performance, two trendlines are included in the figure: a diagonal black dashed line that represents the ideal scenario where predictions perfectly align with actual values, and a red polynomial trend line that fits the scatter data, indicating the overall prediction trend. The proximity of these trend lines suggests that the model performs well.

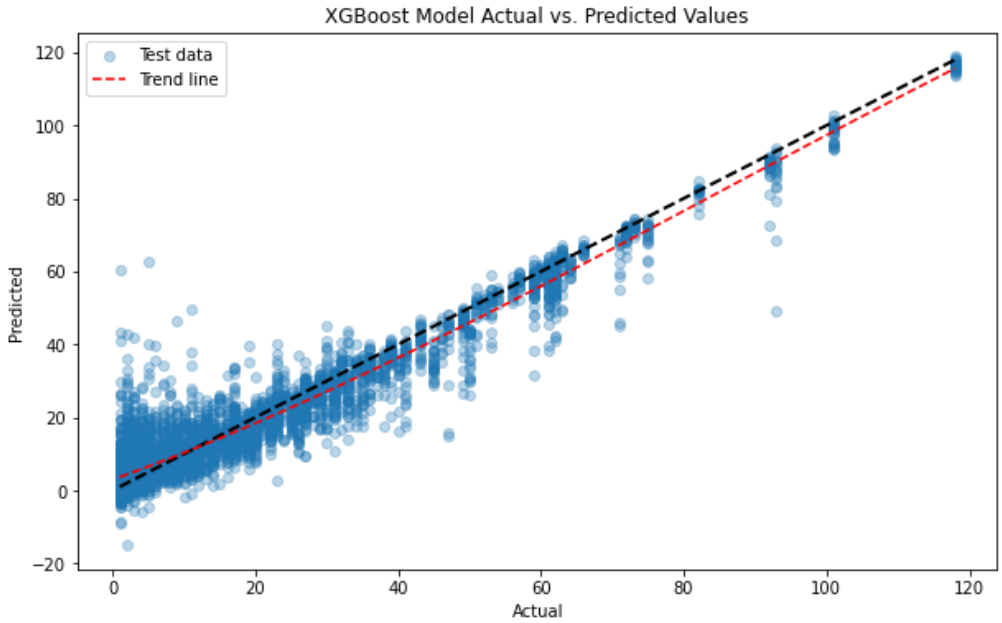


Figure 5-11 XGBoost Testing Set Visualization Actual vs. Predicted Values

Furthermore, **Figure 5-12** presents a histogram of residuals, which visualizes the distribution of prediction errors by calculating the difference between the actual and predicted values. The x-axis denotes the residual values while the y-axis shows their frequency, offering a quantitative insight into the model's prediction accuracy. Overlaid on the histogram is a kernel density estimate line, which provides a smoothed representation of the residuals' distribution. This line helps in identifying the density and spread of errors more clearly. Observations from this figure show that the residuals predominantly cluster around zero, which is indicative of a well-calibrated model. This clustering suggests that the model's predictions are generally accurate and well-aligned with the actual data, demonstrating no significant systematic error in terms of overestimation or underestimation.

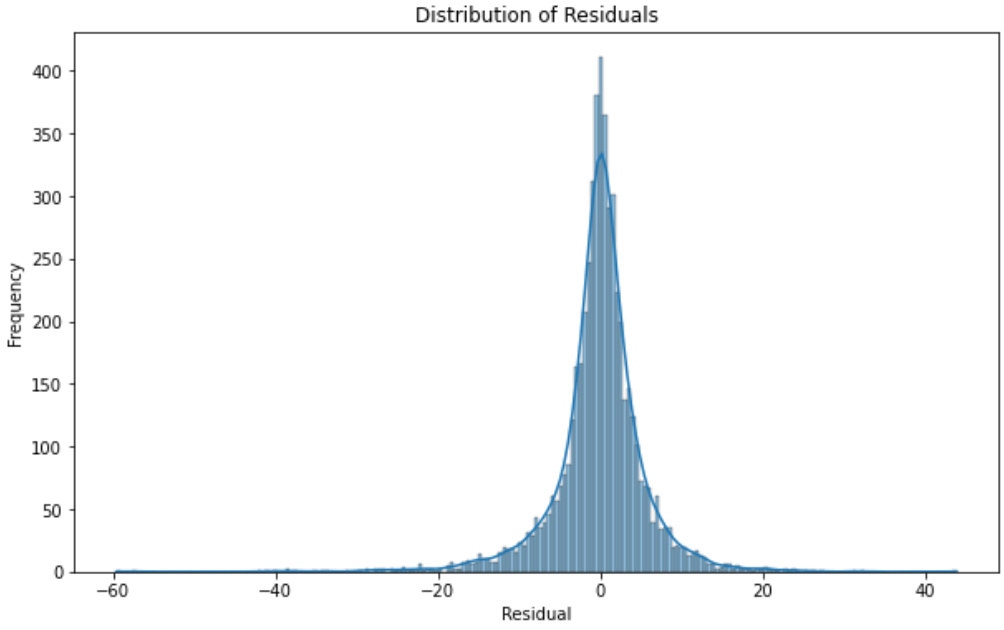


Figure 5-12 XGBoost Testing Set Distribution of Residuals

5.4 Global Interpretation of Winter Crash Frequency Model Variables

This section utilized the SHAP (SHapley Additive exPlanations) for the global analysis of the model, which helps in examining important feature rankings as shown in **Figure 5-13 (a)** and understanding how different feature values influence crash frequency, illustrated by the SHAP summary plot in **Figure 5-13 (b)**.

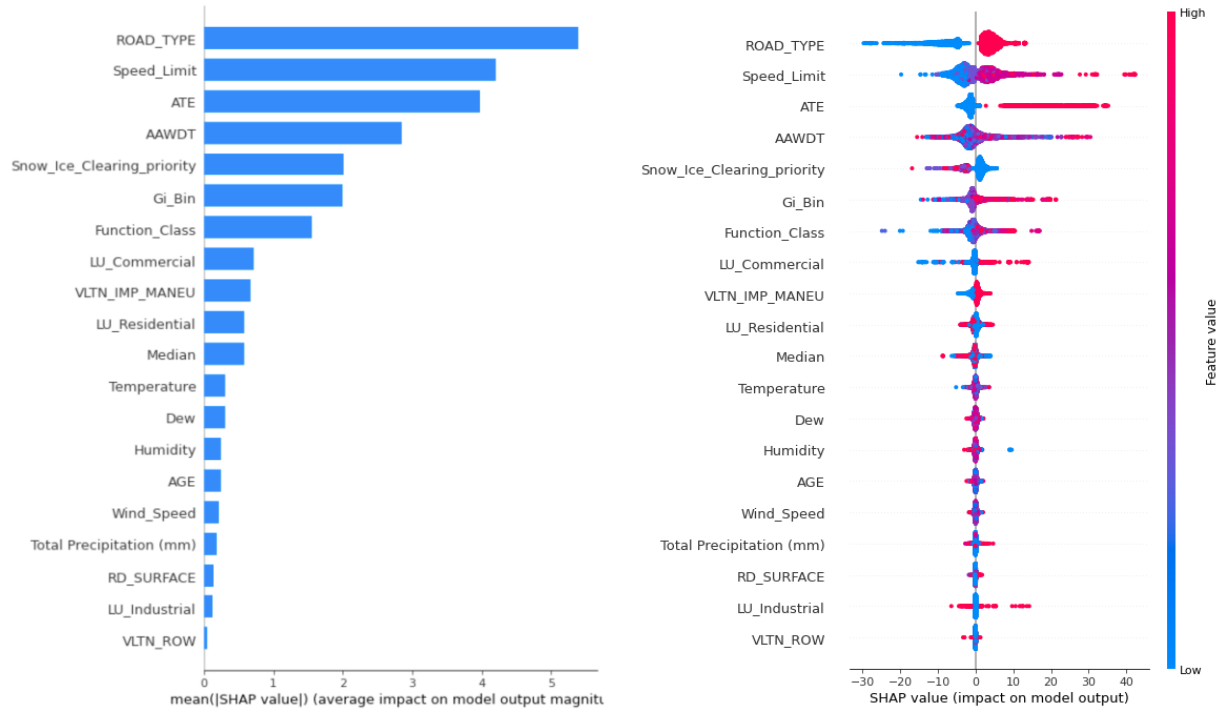


Figure 5-13 XGBoost Model (a) SHAP Feature Ranking Plot (b) SHAP Summary Plot

These SHAP global explanations offer insights into the top twenty significant features that contribute to building the crash frequency model. In **Figure 5-13 (a)**, it is observed that road type, speed limit and ATE are the top three most important features among the features listed. Existing research indicates that crashes tend to occur more frequently at intersections compared to midblocks. Intersections are critical points on roads where a variety of modes and traffic directions intersect. Intersections, being complex traffic points where different directions and modes intersect, become particularly hazardous in winter due to increased stopping distances and reduced visibility from adverse weather conditions [102]. Conversely, midblock sections are comparatively simpler than the intersections due to the consistent surrounding transportation infrastructures.

The speed limit is identified as the second most crucial feature. It has been identified in many literatures that a higher speed limit often results in more crashes, and the severeness of the crash is often higher [103]. With the added impact of adverse winter weather and road conditions, it can exacerbate the crash risks for road users. ATE ranks third as it acts as an important tool in modifying the driver behavior by imposing a deterrence effect.

The following are some additional significant features where the magnitude of the feature significance is about half of the first three forementioned key features. They are AAWDT, snow ice clearing priority route, Gi_Bin, and road function class, each sharing a commonality in traffic exposure. High traffic exposure often leads to more frequent vehicle interactions, which increases crash risks [104]. Gi_Bin is a spatial variable that areas with intense crash clustering, typically characterized by high traffic exposure [105].

Moderately significant features include land use variables, violation type, and road characteristics. Land use is often an important contributor, as it dictates to the layout and design of the road network; in turn the traffic demands are often different from one land use to another. Frequent traffic exposure under winter seasons induces greater danger to road users due to decreases in driving visibility. This reduced visibility hinders driving judgment, which evidently increases the frequency of faulty driving maneuvers. Hence, improper driving maneuvers becomes an important traffic violation type. The median is a preventive measure that separates opposing lanes of the traffic [106], organizing the flow of traffic, reducing the likelihood of confusion for drivers, while also acting as a visual cue to keep drivers aligned within their lanes, which is critical in low visibility conditions [107].

The remaining features have comparatively low significance to the model outcome. In descending order, the features are temperature, dew point temperature, humidity, drivers age, wind speed, precipitation, road surface condition, industrial land use, and right of way violation. Amongst the weather variables, the ambient temperature appears as the most influential weather feature.

Like the feature ranking diagram, the SHAP summary plot shown in **Figure 5-13 (b)** also provides the feature importance in descending order. The figure provides an indication in how feature value influences the SHAP values, hence providing insight as to how feature values contribute to crash frequency. Similar to the explanation provided in Chapter 4, the feature names are on the y-axis of the diagram, while the x-axis represents the corresponding SHAP values, with a black vertical line at a SHAP value of zero as a visual aid to separate the positive and negative SHAP values. Additionally, the high feature values are shown in red dots whereas the low feature values are in blue.

It appears that the three most significant features on the SHAP summary plot are road type, speed limit, and ATE, having a similar spread and placement of feature values. The three features all have a clear distinction between red and blue clusters, where the red dots tend to accumulate at positive SHAP values, while the blue dots accumulate at negative SHAP values. This gives a clear illustration that areas with an intersection road type, high speed limit, or the presence of ATE, will contribute towards increasing the crash frequency. Conversely, midblock road types, low speed limits, and the absence of ATE indicate areas with lower crash frequency.

For AAWDT, it is observed that high traffic volumes can have extremely high SHAP values, which increases crash risks. However, at a medium or low traffic volume, the contributions are mixed and concentrated around the SHAP value of zero, which is difficult to provide a definitive conclusion for. Such mixed conditions means that the contribution of AAWDT varies depending on the instance. Following is the snow ice clearing priority variable. According to the summary plot, routes with high clearing priority often have increased crash risks while routes with low clearing priorities tend to be safer. This is an indication that the city has appropriately classified high priority clearing routes correctly as those roads have a higher propensity for crashes.

Moving along are spatial analysis, function class, commercial land use, and improper driving maneuver variables. For these variables, high feature values tend to be distributed in the positive SHAP value regions, it concludes that spatial hot spots, high function class roads, commercial use land, and improper driving maneuvers often leads to higher accident frequency. In regards to residential land use, there is no clear pattern of blue and red dots, as red dots appear on both sides, and blue dots tend accumulate in the middle, which makes interpretation ambiguous. In terms of the median variable, it is observed the red dots (presence of median) have a dominantly negative SHAP value, thus illustrating the importance of median in reducing the winter crashes.

Moving forward to the features with lower significance to the model prediction, these features have all presented a mixture of red and blue dots close to the zero SHAP value. They are ambient temperature, dew point temperature, humidity, driver age, wind speed, precipitation, road surface condition, industrial land use, and right of way violation. Among those features, they are dominated by the weather variables. It suggests that the importance of these weather features varies depending on the specific characteristics of the instances. This could be due to the collected coarse

data obtained as daily average weather conditions, which is a limitation of this study. We acknowledged that finer hourly weather data could potentially enhance the model performance while also improving the models explainability.

5.5 Exploring Feature Dynamics at High and Low Crash Frequency Locations

5.5.1 Feature Interplay at High Crash Frequency Locations

The randomly generated high crash frequency waterfall plots are shown in **Figure 5-14** ordered by increasing predicted crash counts, ranging from 30.36 in **Figure 5-14 (a)** to 118.72 in **Figure 5-14 (d)**. This arrangement clearly illustrates how features behave differently as the predicted crash frequency increases and highlights the key findings derived from these plots.

The comparison table between the predicted crashes and the actual crashes for these four instances is summarized in **Table 5-4**, indicating that the predicted value is generally close to the actual value, with marginal percentage difference of about 2 to 3%.

Table 5-4 High Crash Frequency Model Prediction vs. Actual Frequency

Instance	Predicted	Actual	Absolute % Difference
(a) 8251	30.36	31	2.06
(b) 25325	39.50	41	3.66
(c) 4148	55.75	57	2.19
(d) 8184	118.72	118	0.61

To understand these waterfall plots, note that the y-axis in each instance ranks features by descending importance, with the most contributing features at the top and the least at the bottom. Feature values are displayed on the left side of each feature name, indicating their position within their respective categories, which can be referenced from the summary statistics (**Table 5-3**). Red bars indicate positive contributions to the model outcome, while blue bars represent negative contributions. The width of each bar visually indicates the significance of the feature's impact on the specific instance's prediction. The final prediction value, denoted as $f(x)$, is the cumulative result of all feature contributions, while the mean value $E(x)$=16.359, is shown below the x-axis for all cases.

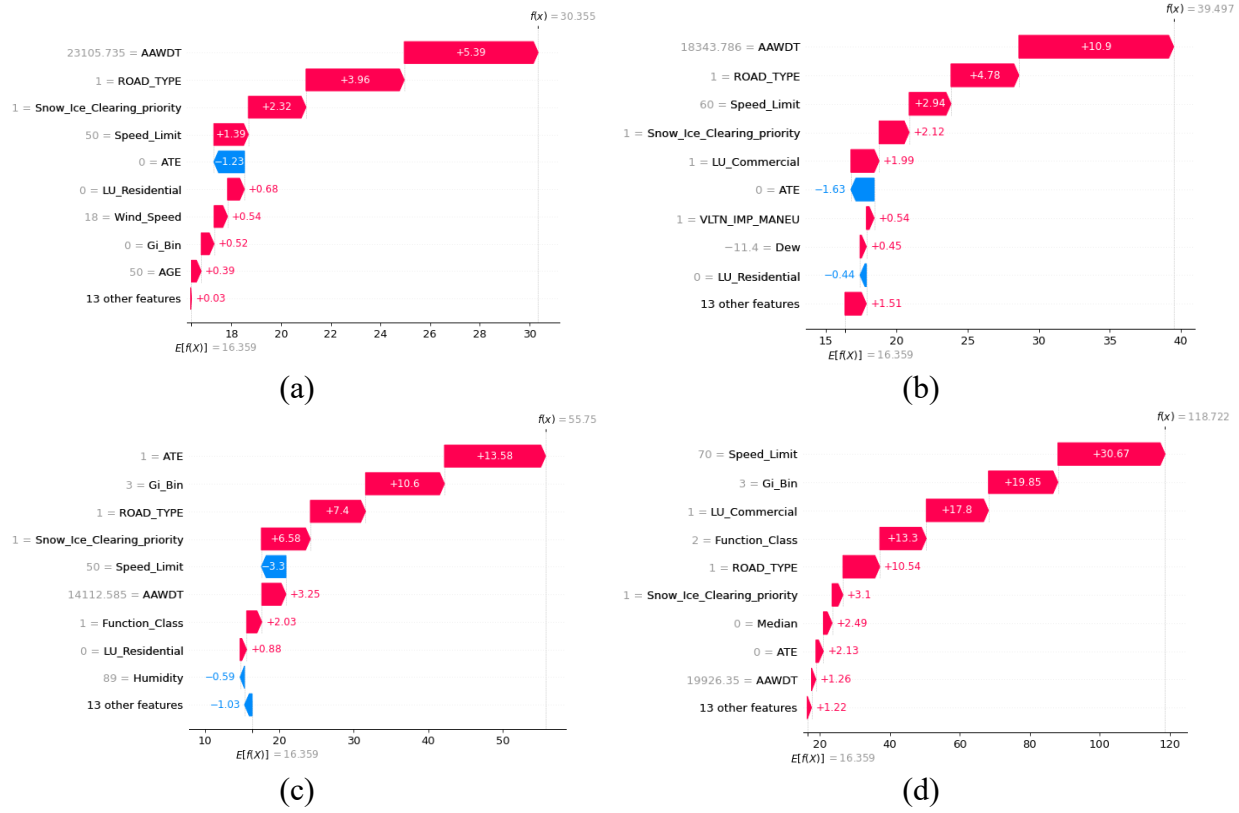


Figure 5-14 Waterfall Plots of High Winter Crash Frequency Instances in Ascending Order (a) Instance 8251, (b) Instance 25325, (c) Instance 4148, (d) Instance 8184

With a firm understanding of the basics of a waterfall plot, we can now investigate **Figure 5-14** to understand the feature interplay at high crash frequency locations. Across multiple instances, AAWDT consistently emerges as a highly influential factor. For instance, in **Figure 5-14 (a)**, AAWDT is 23105.735, significantly higher than the mean value of 13646.751 The substantial size of its respective bar underscores its critical role in predicting higher crash frequencies. Although AAWDT's contribution in **Figure 5-14 (d)** is less pronounced compared to other instances, despite being above the average traffic volume, the high-speed limit in a commercial zone significantly amplified the predicted crash frequency.

In addition, other critical factors are the road type feature and snow clearing priority route. Among all four instances under high crash frequency waterfall plots, intersections and snow clearing high priority routes have been recurrently highlighted as the top significant positive contributors. This finding reinforces the literature that shows intersections often have more complex traffic

movements, thereby increasing the potential for collisions, especially under diverse winter weather conditions. Furthermore, it emphasizes the importance of winter road maintenance on these high-priority routes. Similar to the high snow and ice clearing priority routes, the high importance of road function class also significantly influences the number of collisions positively. Snow clearing priority routes and function class rank are fundamentally regulated and ranked based on traffic volume demand, reflecting the level of traffic exposure.

Moreover, high speed limits also have an extremely high contribution to the model outcome. The magnitude of influence tends to increase as the predicted crash frequency rises. When the speed limit is identified as 50 km/h, as shown in **Figure 5-14 (a)** and **Figure 5-14 (c)**, it shows an unclear pattern to the influence on crashes as it can either increase or decrease. However, in **Figure 5-14 (b)** and **Figure 5-14 (d),** where the speed limit is 60 km/h and 70 km/h, respectively, the influence on the model prediction outcome becomes significantly higher, as indicated by the size of the red bar. This clearly shows the danger of high-speed limits, which magnify the impact on increasing crash frequency. Additionally, these two figures also identify that the locations are in commercial land use zoning, which further exacerbates the impact of increasing the number of crashes. Literature often cites that commercial zoning tends to have a more complex landscape and a variety of infrastructures, which demand much higher driver attention to the driving environment.

In **Figure 5-14 (c)**, the presence of ATE shows a significant positive influence on the prediction outcome. For the rest, in **Figure 5-14 (a) and (b)**, it shows that the absence of enforcement cameras lowers the chance of accidents. This aligns with the previous conclusion that the city has effectively installed traffic cameras at high-risk and crash-prone locations to alert drivers to their driving behavior. Moreover, the absence of preventive measures such as medians, as shown in **Figure 5-14 (d),** has increased the odds of collisions. However, a contradictory finding is shown in **Figure 5-14 (c),** where the high percentage of humidity lowers the accident rate. This discrepancy may be related to the limitations of the SHAP linear explanation, which can sometimes lead to inaccurate predictions as feature importance decreases.

When examining the behavior of the spatial variable, it is identified that **Figure 5-14 (c)** and **(d)** show a Gi_Bin of (+3), which spatially identifies 99% confident hot spots with the most intensely clustered high number of crashes. This indicates that spatially identified hot spots often tend to

result in a higher number of crashes compared to the normal spots shown in **Figure 5-14 (a) and (b),** which do not demonstrate a statistically significant spatial clustering pattern.

5.5.2 Feature Interplay at Low Crash Frequency Locations

Previously, **Figure 5-14** illustrated the behavior of high crash frequency risk factors, highlighting the significance of traffic volume, ATE, speed limit, and spatial characteristics. Here, **Figure 5-15** provides insights into the feature behaviors at low crash frequency locations. Similar to the previous section, the instances are randomly generated at first. Then, based on the predicted value, the figures are ranked and listed in order of increasing crash frequency, from a predicted crash frequency of 1.27 in **Figure 5-15 (a)** to 12.49 in **Figure 5-15 (d).** This arrangement helps reveal the trend of feature values, providing better insights and capturing consistent patterns in feature behaviors.

As in the previous section, the comparison between the predicted number of crashes and actual crashes are also applied to the low crash frequency instances shown in **Table 5-5**. It is observed that the prediction for the low crash instances tend to have a larger percentage difference compared to their actual crash frequency than the high crash frequency instance, which is reasonable, as low crash locations tend to be more spatially scattered than the relatively condensed high crash frequency locations.

Table 5-5 Low Crash Frequency Model Prediction vs. Actual Frequency

Instance	Predicted	Actual	Absolute % Difference
(a) 18547	1.27	1	27.00
(b) 15231	2.48	4	38.00
(c) 7752	4.32	4	8.00
(d) 8634	12.49	11	13.55

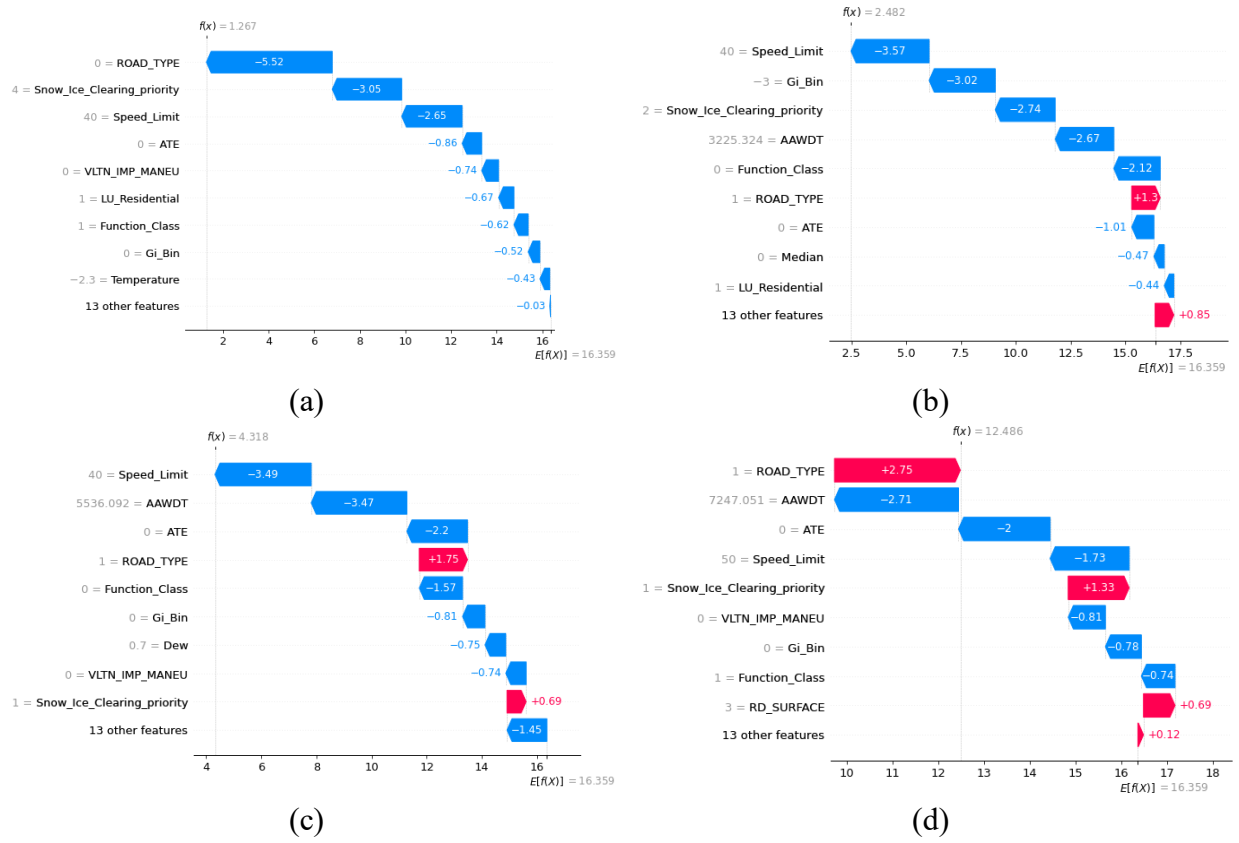


Figure 5-15 Waterfall Plots of Low Winter Crash Frequency Instances in Ascending Order. (a) Instance 18547, (b) Instance 15231, (c) Instance 7752, (d) Instance 8634

At first glance, it is clear that some key features exhibit drastically different behaviors. These include road type, snow clearing priority, and speed limit. Regarding road type, **Figure 5-15 (a)** suggests that midblock shows a negative influence in blue bar that reduces the crash frequency whereas the intersection in red bar increases. Hence confirming that midblock sections are safer, contributing to fewer crashes compared to intersections, as shown in **Figure 5-15 (b), (c), and (d).** Additionally, comparing the snow clearing priority routes, we see an increasing trend from low to high priority from **Figure 5-15 (a) to (d),** with a corresponding increase in crashes. This indicates that high clearing priority routes often have more traffic exposure, highlighting the importance of active clearing on these routes under harsh weather conditions. Speed limit is another prominent feature consistently ranked at the top. For these low crash frequency locations, the speed limit typically falls at 40 km/h or at 50km/h, reducing the crashes shown in all low crash frequencies instances in **Figure 5-15**.

An interesting phenomenon that can be observed is the reduced significance of traffic volume at low crash frequency locations. According to the descriptions from **Figure 5-15 (b), (c), and (d)** the traffic volume tends to be far less than the average traffic volume which negatively contribute to the crash risks. However, taking a closer look of these three plots, it is also observed that with the increase of AAWDT, it influences the prediction outcome towards increasing. This finding aligns with research, indicating that lower traffic exposure results in less congestion, fewer vehicle interactions, and thus fewer collisions. Among all low crash frequency locations, the presence of ATE is uncommon, often due to the lower traffic demand in these areas.

These low-accident frequency locations typically occur in residential with lower speed limits. Unlike commercial zones, which are mostly identified in high crash locations with higher traffic volumes, residential areas tend to have lower traffic volumes. Additionally, the road function class at low crash frequency locations differs significantly from high crash frequency locations. Crashes at low frequency locations often occur on collector roads, as shown in **Figure 5-15 (b) and (c).** These local roads are designed to connect neighborhoods and have lower traffic capacity. Typically, these regions are well-equipped with preventive measures and narrower road widths to limit vehicle speeds, resulting in safer driving maneuvers, as illustrated in **Figure 5-15 (a), (c), and (d).**

When spatially describing these low crash frequency locations, they often occur in normal spots and occasionally in cold spots, as shown in **Figure 5-15 (b),** within residential neighborhoods with lower speed limits. This is markedly different from the behaviors observed in high crash frequency instances, where crashes are mostly located in hot spot regions. This identification effectively conveys that areas with intense clustering of low crash values often have lower speed limits and less traffic volume, resulting in fewer vehicle interactions and lower crash occurrences.

Chapter 6: CONCLUSIONS

This chapter provides the final remarks of the thesis. It starts with an overview of the thesis, restating the objective and motivations that have driven this study. It then gives a highlight of the key findings and contributions. Additionally, the thesis also acknowledges the limitations of this study, and discusses future research and recommendations.

6.1 Overview of the Thesis

One of the primary motivations behind this thesis is the enhancement of traffic safety under harsh winter conditions. It is critically important for maintenance personnel to conduct effective and accurate executions of winter road maintenance (WRM) tasks to ensure the safety of road users in the winter. The thesis first explores the classification of crash-inducing snowstorm events by utilizing weather data, traffic exposure data, road condition data, and most importantly, maintenance data. To address the high dimensional dataset with a limited 231 events collected from two winter seasons, Support Vector Machine (SVM) was utilized due to its proficiency in dealing with high dimensional and non-linear datasets. This approach helps understand the risk factors that contribute the most to crash-inducing snowstorms and the interplay of underlying features from specific prediction instances. By applying explainable artificial intelligence (XAI), the transparency of the model decisions is revealed, uncovering the decision-making process within the model.

Additionally, efforts were made to develop a granular crash frequency model. The goal was to establish a model that can accurately predict the number of crashes, considering both micro and macro-level variables. Micro-level variables include collision records, weather variables, road characteristics, and traffic volume, while macro-level variables incorporate spatial analysis and land use for a more comprehensive analysis. Due to traffic volume data coverage issues during the data collection phase, an Ordinary Kriging (OK) method was applied to interpolate the traffic volume for crash locations lacking data, enabling a spatially continuous traffic volume estimation for the study period. In terms of spatial analysis, hot spots (HS) and cold spots (CS) for winter collisions were successfully identified using the Hot Spot Analysis (HSA) tool in ArcGIS. By

layering additional features on top of the hot spot map, we could visually inspect feature behaviors and perform a preliminary assessment before building the model.

Upon evaluating the features, the focus shifted towards training and establishing an optimal model by comparing three widely adopted tree-based machine learning (ML) models: XGBoost, Random Forest (RF), and LightGBM. With XGBoost demonstrating the best performance, the model was further examined using the SHAP XAI technique to determine feature behaviors at high and low crash frequency locations.

Overall, the thesis follows a structure of building an optimal high predictive power ML model, followed by an application of XAI to explain the model outcomes and bring transparency to the model predictions. Achieving both high performance and credible explainability ensures that these models are practical and applicable in real-world scenarios.

6.2 Key Findings of the Thesis

The key findings of this thesis consist of two parts. The first part involved identifying crash-occurred snowstorm events using an SVM classification approach. The model considered risk factors from a microscopic perspective, focusing on weather variables, road conditions, traffic exposures, and maintenance record variables. In the second part, a ML model was built to predict the crash frequency under the influence of both micro level and macro level variables. The included micro-level variables are similar to that of part one, which are the weather variables, road characteristics, traffic exposure, and collision records. To enhance the predictive power of this granular prediction model for the regression task, macro-level variables including land use and spatial analysis were incorporated. With such a holistic analysis from both micro and macro level, the model gives a high prediction accuracy in predicting the number of crashes. For both parts of our study, the model's working mechanisms are revealed using SHAP XAI technique. By utilizing this additional analysis coupled with high-performance ML models, we were able to examine feature contributions to the model from both global and local perspectives.

6.2.1 Classification of Crash-Inducing Snowstorm Events and SHAP explanations

To classify crash-inducing snowstorm events, the study applied SVM modeling due to its ability to handle non-linear and high dimensional datasets. The model with the best performing kernel,

the radial basis function (RBF), achieved an accuracy rate of 87.2% and a recall rate of 80%. The model was further evaluated using SHAP to uncover the models internal decision-making structure, revealing that duration, road length, and precipitation were the most significant factors. Besides, SHAP summary plot not only demonstrated the feature importance to the model outcome, but also showed the impact of different feature values. Findings suggest that longer snowstorm duration, longer road length, and higher precipitations are the top three leading factors associated with higher risk of accidents. Furthermore, it highlights the importance of maintenance activities in winter cities like Edmonton, illustrating that the absence of plowing operations and anti-icing increases the risk of accidents.

This study also acknowledges issues with counterintuitive feature behaviors observed using the SHAP summary plot interpretation, particularly with precipitation intensity and ADT. This prompted the use of SHAP local explanations to investigate representative instances, ultimately assisting in explaining these counterintuitive behaviours. In addition, it is important to reiterate that SHAP uses a linear explanation model to provide local estimates, which can disregard feature dependence. This sometimes results in unintuitive explanations for certain features. Despite generally sound explanations for many features, maintenance personnel are encouraged to critically evaluate these outcomes using their industry expertise.

6.2.2. Winter Crash Frequency Model and SHAP explanation

Beyond the classification model in the first part of the analysis, the study also performed an in-depth analysis of winter crash frequency, encompassing 26,970 winter crashes. This model provided a more granular analysis, facilitating the discovery of broader winter crash patterns, specifically for high and low crash frequency locations. Unlike the previous analysis, this part included both micro and macro-level variables. With the help of spatial analysis, HS and CS regions were then identified. Overlaying additional features revealed that HS were often associated with high traffic volume, high function class, higher snow and ice clearing priority routes, commercial zoning, and intersections with higher speed limit routes. Conversely, CS were typically located in areas with lower traffic volumes, lower function class, and lower speed limits, mostly condensed in residential zones.

After conducting the preliminary assessment of feature interactions to HS and CS regions, the crash frequency ML model was developed with the addition of this spatial variable. The study first evaluated three tree-based models, XGBoost, RF, and LightGBM due to their high performance and computational efficiency. After fine-tuning via hyperparameter optimization, XGBoost was selected as the best-performing model with a testing R^2 value of 92.67%, a MAE of 3.64, and a RMSE of 5.77 which had the highest R^2 value and the lowest error among the three models. Using the identified optimal model, SHAP analysis was implemented to demonstrate both the global and local interpretations of the model's predictions, revealing feature behaviors at high and low crash frequency locations.

From SHAP global interpretation, the road type, speed limit, and the presence of ATE cameras were identified as the top three significant features contributing to winter crashes. A SHAP local explanation was then implemented to further demonstrate how features interacted differently at high and low crash frequency locations. For high crash frequency locations, the top contributing features were high traffic volume, presence of ATE, intersection, high speed limits, and high-priority snow clearing routes. For the global features, high frequency regions were mostly categorized as commercial zones, which often fell into the identified HS. Low crash frequency locations were typically associated with midblock sections, absence of ATE, low priority clearing routes, lower speed limits, and lower traffic volumes. Regarding macro level features, these locations were commonly identified as residential zones at cold spots or normal spots, indicating low vehicle interactions, ultimately reducing crash likelihoods.

Compared to the previous crash-inducing snowstorm analysis, this crash frequency model was developed on a much larger dataset. This made the model more stable, and the explained features more intuitive. This supports the commonly accepted notion that larger datasets would allow for more consistent behaviors in the model predictions and reduce the chance of counterintuitive outcomes.

6.3 Contributions of the Thesis

Throughout the entire thesis, substantial methodological and practical contributions have been made as summarized below.

6.3.1 Methodological Contributions

- **Enhanced the understanding of the significant risk factors in crash-inducing snowstorm events.** Applied weather data, maintenance record data, road friction, traffic exposure data to predict crash-inducing snowstorm events and justified the prediction with reasoning based on feature contributions.
- **Conducted spatial analysis to identify clusters of high and low crash regions in the city.** Screened the crash-prone locations and visualized how various features behaved differently at hot and cold spots.
- **Improved the crash frequency model's predictive power by combining both micro and macro-level variables**. The combination of micro and macro-level variables improved the model performance and enhances the model comprehensiveness.
- **Conducted comprehensive comparisons of high-performance machine-learning algorithms for classification and regression tasks.** The SVM RBF kernel function worked the best in evaluating the crash-inducing snowstorm events, while the XGBoost model was the most accurate algorithm in predicting crash frequency.
- **Improved transparency of model predictions using the SHAP XAI technique.** Although ML algorithms show high predictive power in application, model transparency is a critical concern. To understand the rationale behind the predictions, SHAP analysis was deployed to provide both global and local analysis of prediction outcomes.

6.3.2 Practical Contributions

- **Provided comprehensive analysis of ML prediction outcomes to assist winter road maintenance personnel in decision-making.** The application of XAI revealed the reasonings behind ML predictions, assisting WRM personnel in making informed decisions, ultimately enhancing the effectiveness of resource allocation.
- **Identified and evaluated crash-inducing snowstorm events to improve winter road safety.** Deploying a model that accurately predicts the crash-inducing snowstorm events along with insights on the prediction outcomes can improve the effectiveness of WRM decisions-making.

- **Locate crash prone locations and understand their characteristics.** Through detection and evaluation, the characteristics of high or low crash locations provided insights which can guide planners in design and infrastructure development towards improving winter road safety.

6.4 Limitations and Future Work

This section discusses the limitations encountered during the study and outlines the potential areas for further research as listed below.

- The study faced challenges in collecting granular traffic volume data from the city, particularly in acquiring hourly traffic volume datasets to increase the ML predictive power and interpretability. In the first part of the study, ADT was used for snowstorm events, while the second part involved the use of AAWDT for traffic volume estimation. The traffic volume data was obtained using coarser and interpolated traffic volume estimates, which may have affected the accuracy of the predictions. This limitation underscores the need for more detailed and accurate traffic volume data in future studies.
- This thesis focused on developing and comparing ML model algorithms for predictive modelling for winter road safety, providing an intensive analysis of data-driven approaches. However, traditional statistical models remain unexplored. Future research could include a comparative analysis between non-parametric models and conventional parametric models to better understand the strengths and weaknesses of each approach.
- One major limitation of this study is the data size. Larger datasets tend to increase model stability and overall reliability. Addressing this limitation could significantly enhance the effectiveness of the predictive models used in this research.
- While SHAP enhances ML model transparency, it faces several limitations, including its linear assumption, issues with global interpretation, feature independence, and computational efficiency. In particular, SHAP's additive nature assumes linear feature contributions, which may fail to capture non-linear interactions in complex ML models like SVM. Moreover, while global interpretability offers insights into overall trends, and local interpretations can reveal potential counterintuitive results, these may not always align with human intuition. Additionally, the computational intensity of SHAP calculations can limit

its usage in real-time applications. Future work could focus on exploring more efficient algorithms or approximations to mitigate these computational demands.

- In addition to SHAP, several other XAI techniques can be utilized to improve model interpretability such as LIME (Local Interpretable Model-agnostic Explanations) and PDP (Partial Dependence Plots). Like SHAP, LIME also provides explanations for individual predictions by generating perturbed samples around the instance of interest to capture the model's local behavior. PDP helps visualize the effects of a feature on the predicted outcome by marginalizing other features. While SHAP offers high interpretations at the cost of high computation expense, LIME and PDP can be less complex approaches that increase both the computation efficiency and model transparency.

BIBLIOGRAPHY

[1] "How Do Weather Events Impact Roads?," [Online]. Available: https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm. [Accessed 10 July 2023].

[2] "Just the Facts – Winter driving," Royal Canadian Mounted Police, 17 December 2019. [Online]. Available: https://www.rcmp-grc.gc.ca/en/gazette/just-the-facts-winter-driving. [Accessed 1 July 2023].

[3] D. Eisenberg and K. E. Warner, "Effects of Snowfalls on Motor Vehicle Collisions, Injuries, and Fatalities," *American journal of public health,* no. 1, pp. 120-124, 2005.

[4] B. Mills, J. Andrey, B. Doberstein, S. Doherty and J. Yessis, "Changing patterns of motor vehicle collision risk during winter storms: A new look at a pervasive problem," *Accident Analysis & Prevention,* vol. 127, pp. 186-197, 2019.

[5] A. Andersson and L. Chapman, "The use of a temporal analogue to predict future traffic accidents and winter road conditions in Sweden," *Meteorological Applications,* vol. 18, no. 2, pp. 125-136, 2011.

[6] A. Abohassan, K. El-Basyouny and T. J. Kwon, "Exploring the associations between winter maintenance operations, weather variables, surface condition, and road safety: A path analysis approach," *Accident Analysis & Prevention,* vol. 163, no. 106448, 2021.

[7] H. Perrin, P. T. Martin and B. G. Hansen, "Modifying Signal Timing During Inclement Weather," *Transportation Research Record,* vol. 1748, no. 1, pp. 66-71, 2001.

[8] L. C. Goodwin and P. A. Pisano, "Weather-responsive traffic signal control," *Institute of Transportation Engineers. ITE Journal,* vol. 74, no. 6, p. 28, 2004.

[9] R. B. Mitchell and C. Rapkin, Urban traffic: A function of land use, Columbia University Press, 1954.

[10] M. Boarnet and R. Crane, "The influence of land use on travel behavior: specification and estimation strategies.," *Transportation Research Part A: Policy and Practice,* vol. 35, no. 9, pp. 823-845, 2001.

[11] R. V. D. Horst and S. d. Ridder, "Influence of roadside infrastructure on driving behavior: driving simulator study.," *Transportation Research Record,* vol. 1, pp. 36-44, 2018.

[12] P. P. Jovanis and H.-L. Chang, "Modeling the Relationship of Accidents to Miles Traveled," *Transportation Research Record,* vol. 1068, no. 42-51, pp. 57-58, 1986.

[13] D. M. Levinson, F. Xie and S. Zhu, "The co-evolution of land use and road networks.," *Transportation and traffic theory,* pp. 839-859, 2007.

[14] A. Soltani and S. Askari, "Exploring spatial autocorrelation of traffic crashes based on severity," *Injury,* vol. 48, no. 3, pp. 637-647, 2017.

[15] C. A. Blazquez, B. Picarte, J. F. Calderón and F. Losada, "Spatial autocorrelation analysis of cargo trucks on highway crashes in Chile," *Accident Analysis & Prevention,* vol. 120, pp. 195-210, 2018.

[16] X. Wen, Y. Xie, L. Jiang, Z. Pu and T. Ge, "Applications of machine learning methods in traffic crash severity modelling: current status and future directions," *Transport reviews,* vol. 41, no. 6, pp. 855-879, 2021.

[17] S. Lakshmi, I. Srikanth and M. Arockiasamy, "Identification of traffic accident hotspots using geographical information system (GIS)," *International journal of engineering and advanced technology,* vol. 9, no. 2, pp. 4429-4438, 2019.

[18] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives.," *Transportation research part A: policy and practice,* vol. 44, no. 5, pp. 291-305, 2010.

[19] D. Bzdok, M. Krzywinski and N. Altman, "Machine learning: A primer," *Nature methods,* vol. 14, no. 12, p. 1119, 2017.

[20] C. Ley, R. K. Martin, A. Pareek, A. Groll, R. Seil and T. Tischer, "Machine learning and conventional statistics: making sense of the differences.," *Knee Surgery, Sports Traumatology, Arthroscopy,* vol. 30, no. 3, pp. 753-757, 2022.

[21] Y. Ali, F. Hussain and M. M. Haque, "Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review.," *Accident Analysis & Prevention,* vol. 194, no. 107378, 2024.

[22] G. Baryannis, S. Dani and G. Antoniou, "Predicting supply chain risks using machine learning: The trade-off between performance and interpretability," *Future Generation Computer Systems,* vol. 101, pp. 993-1004, 2019.

[23] X. Wen, Y. Xie, L. Jiang, Y. Li and T. Ge, "On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development," *Accident Analysis & Prevention,* vol. 168, no. 106617, 2022.

[24] C. Arteaga, A. Paz and J. Park, "Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach," *Safety Science,* vol. 132, no. 104988, 2020.

[25] L. Budach, M. Feuerpfeil, N. Ihde, A. Nathansen, N. Noack, H. Patzlaff, F. Naumann and H. Harmouch, "The effects of data quality on machine learning performance," *arXiv preprint arXiv:2207.14529,* 2022.

[26] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing,* vol. 97, p. 105524, 2020.

[27] P. Barbiero, G. Squillero and A. Tonda, "Modeling generalization in machine learning: A methodological and computational study.," *arXiv preprint arXiv:2006.15680,* 2020.

[28] A. Abdulhafedh, "Crash frequency analysis," *Journal of Transportation Technologies,* p. 169, 2016.

[29] T. Bhowmik, S. Yasmin and N. Eluru, "A new econometric approach for modeling several count variables: a case study of crash frequency analysis by crash type and severity," *Transportation research part B: methodological,* vol. 153, pp. 172-203, 2021.

[30] Y. Zhang, Y. Xie and L. Li, "Crash frequency analysis of different types of urban roadway segments using generalized additive model," *Journal of Safety Research,* vol. 43, no. 2, pp. 107-114, 2012.

[31] "Canadian Climate Normals 1981-2010 Station Data," Goverment of Canada, 2024. [Online]. Available: https://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?searchType=stnProx&txtRadius=25&optProxType=city&selCity=53%7C33%7C113%7C30%7CEdmonton&selPark=&txtCentralLatDeg=&txtCentralLatMin=0&txtCentralLatSec=0&txtCentralLongDeg=&txtCentralLon. [Accessed 14 July 2024].

[32] L. Wang, M. Abdel-Aty, X. Wang and R. Yu, "Analysis and comparison of safety models using average daily, average hourly, and microscopic traffic," *Accident Analysis & Prevention,* vol. 111, pp. 271-279, 2018.

[33] M. Abdel-Aty, R. Pemmanaboina and L. Hsia, "Assessing crash occurrence on urban freeways by applying a system of interrelated equations," *Transportation research record,* vol. 1953, no. 1, pp. 1-9, 2006.

[34] Y. Peng, M. Abdel-Aty, Q. Shi and R. Yu, "Assessing the impact of reduced visibility on traffic crash risk using microscopic data and surrogate safety measures," *Transportation research part C: emerging technologies,* vol. 74, pp. 295-305, 2017.

[35] S. Jung, K. Jang, Y. Yoon and S. Kang, "Contributing factors to vehicle to vehicle crash frequency and severity under rainfall," *Journal of safety research,* vol. 50, pp. 1-10, 2014.

[36] J. Seeherman and Y. Liu, "Effects of extraordinary snowfall on traffic safety," *Accident Analysis & Prevention,* vol. 81, pp. 194-203, 2015.

[37] E. Hermans, T. Brijs, T. Stiers and C. Offermans, "The impact of weather conditions on road safety investigated on an hourly basis," in *Washington, DC: Proceedings of the 85thAnnual meeting of the Transportation Research Board.*, Washington, DC, 2006.

[38] S. Chand, Z. Li, A. Alsultan and V. V. Dixit, "Comparing and contrasting the impacts of macro-level factors on crash duration and frequency," *International Journal of Environmental Research and Public Health,* vol. 19, no. 9, p. 5726, 2022.

[39] H. Huang, B. Song, P. Xu, Q. Zeng, J. Lee and M. Abdel-Aty, "Macro and micro models for zonal crash prediction with application in hot zones identification," *Journal of transport geography,* vol. 54, pp. 248-256, 2016.

[40] S. Pervaz, T. Bhowmik and N. Eluru, "Integrating macro and micro level crash frequency models considering spatial heterogeneity and random effects," *Analytic methods in accident research,* vol. 36, no. 100238, 2022.

[41] J. Wang, H. Huang and Q. Zeng, "The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian," *Accident Analysis & Prevention,* vol. 98, pp. 223-231, 2017.

[42] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," *Transportation research part A: policy and practice,* vol. 44, no. 5, pp. 291-305, 2010.

[43] B. Jones, L. Janssen and F. Mannering, "Analysis of the frequency and duration of freeway accidents in Seattle," *Accident Analysis & Prevention,* vol. 23, no. 4, pp. 239-255, 1991.

[44] F. Wei and G. Lovegrove, "An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression," *Accident Analysis & Prevention,* vol. 61, pp. 129-137, 2013.

[45] H. C. Chin and M. Abdul Quddus, "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections," *Accident Analysis & Prevention,* vol. 35, no. 2, pp. 253-259, 2003.

[46] A. Abohassan, K. El-Basyouny and T. J. Kwon, "Exploring the associations between winter maintenance operations, weather variables, surface condition, and road safety: A path analysis approach," *Accident Analysis & Prevention,* vol. 163, p. 106448, 2021.

[47] A. Ghasemzadeh and M. M. Ahmed, "Complementary parametric probit regression and nonparametric classification tree modeling approaches to analyze factors affecting severity of

work zone weather-related crashes," *Journal of Modern Transportation,* vol. 27, pp. 129-140, 2019.

[48] J. Wang and A. Boukerche, "Non-parametric models with optimized training strategy for vehicles traffic flow prediction," *Computer Networks,* vol. 187, 2021.

[49] P. Li and M. Abdel-Aty, "Real-time crash likelihood prediction using temporal attention-based deep learning and trajectory fusion," *Journal of transportation engineering, Part A: Systems,* vol. 148, no. 7, p. 04022043, 2022.

[50] X. Li, D. Lord, Y. Zhang and Y. Xie, "Predicting motor vehicle crashes using support vector machine models," *Accident Analysis & Prevention,* vol. 40, no. 4, pp. 1611-1618, 2008.

[51] M. Zarei, B. Hellinga and P. Izadpanah, "Application of Conditional Deep Generative Networks (CGAN) in empirical bayes estimation of road crash risk and identifying crash hotspots," *International journal of transportation science and technology,* vol. 13, pp. 258-269, 2024.

[52] J. Hu, M.-C. Huang and X. Yu, "Efficient mapping of crash risk at intersections with connected vehicle data and deep learning models," *Accident Analysis & Prevention,* vol. 144, p. 105665, 2020.

[53] L.-Y. Chang and W.-C. Chen, "Data mining of tree-based models to analyze freeway accident frequency," *Journal of safety research,* vol. 36, no. 4, pp. 365-375, 2005.

[54] Z. Zhang, Q. Nie, J. Liu, A. Hainen, N. Islam and C. Yang, "Machine learning based real-time prediction of freeway crash risk using crowdsourced probe vehicle data," *Journal of Intelligent Transportation Systems,* vol. 28, no. 1, pp. 84-102, 2024.

[55] A. Goswamy, M. Abdel-Aty and Z. Islam, "Factors affecting injury severity at pedestrian crossing locations with Rectangular RAPID Flashing Beacons (RRFB) using XGBoost and random parameters discrete outcome models," *Accident Analysis & Prevention,* vol. 181, p. 106937, 2023.

[56] C. Ley, R. K. Martin, A. Pareek, A. Groll, R. Seil and T. Tischer, "Machine learning and conventional statistics: making sense of the differences," *Knee Surgery, Sports Traumatology, Arthroscopy,* vol. 30, no. 3, pp. 753-757, 2022.

[57] P. Linardatos, V. Papastefanopoulos and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy,* vol. 1, no. 1, p. 18, 2020.

[58] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science,* vol. 349, no. 6245, pp. 255-260, 2015.

[59] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," *arXiv preprint arXiv:1702.08608.,* 2017.

[60] R. Wexler, "When a Computer Program Keeps You in Jail," *The New York Times,* p. 1, 13 June 2017.

[61] D. V. Carvalho, E. M. Pereira and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics,* vol. 8, no. 8, p. 832, 2019.

[62] A. Weller, "Challenges for transparency," *arXiv: 1708.01870 v1,* 2017.

[63] C. Molnar, "Interpretable Machine Learning," 26 May 2024. [Online]. Available: https://christophm.github.io/interpretable-ml-book/. [Accessed 30 May 2024].

[64] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems,* vol. 30, 2017.

[65] P. A. Moreno-Sanchez, "Development of an Explainable Prediction Model of Heart Failure Survival by Using Ensemble Trees," in *2020 IEEE International Conference on Big Data (Big Data)*.

[66] M. Athanasiou, K. Sfrintzeri, K. Zarkogianni, A. C. Thanopoulou and K. S. Nikita, " An explainable XGBoost–based approach towards assessing the risk of cardiovascular disease in patients with Type 2 Diabetes Mellitus," in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*.

[67] K. Lin and Y. Gao, "Model interpretability of financial fraud detection by group SHAP," *Expert Systems with Applications,* vol. 210, p. 118354, 2022.

[68] Y. Meng, N. Yang, Z. Qian and G. Zhang, "What Makes an Online Review More Helpful: An Interpretation Framework Using XGBoost and SHAP Values," *Journal of Theoretical and Applied Electronic Commerce Research,* vol. 16, no. 3, pp. 466-490, 2020.

[69] "How Hot Spot Analysis (Getis-Ord Gi*) works," esri, 2024. [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm. [Accessed 16 May 2024].

[70] H. Wackernagel, Multivariate geostatistics: an introduction with applications, Springer Science & Business Media, 2003.

[71] R. A. Olea, The Semivariogram. In: Geostatistics for Engineers and Earth Scientists., Springer, Boston, MA, 1999.

[72] H. Bhavsar and A. Ganatra, "A comparative study of training algorithms for supervised machine learning.," *International Journal of Soft Computing and Engineering (IJSCE),* pp. 2231-2307, 2012.

[73] A. H. Fielding, Cluster and classification techniques for the biosciences, Cambridge University Press, 2006.

[74] V. Vapnik, The nature of statistical learning theory, Springer science & business media, 2013.

[75] A. Zohair and L. Mahmoud, "Prediction of Student's performance by modelling small dataset size," *International Journal of Educational Technology in Higher Education,* vol. 16, no. 1, pp. 1-18, 2019.

[76] K. Pasupa and W. Sunhem, "A comparison between shallow and deep architecture classifiers on small dataset.," in *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2016.

[77] W. S. Noble, "What is a support vector machine?," *Nature biotechnology,* vol. 24, no. 12, pp. 1565-1567, 2006.

[78] A. Patle and D. S. Chouhan, "SVM kernel functions for classification," in *2013 International conference on advances in technology and engineering (ICATE)*, 2013.

[79] S.-i. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Networks,* vol. 12, no. 6, pp. 783-789, 1999.

[80] "1.4. Support Vector Machines," scikit-learn, [Online]. Available: https://scikit-learn.org/stable/modules/svm.html. [Accessed 10 July 2023].

[81] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016.

[82] "XGBoost Tutorial," xgboost developers, 2022. [Online]. Available: https://xgboost.readthedocs.io/en/stable/tutorials/model.html. [Accessed 27 May 2024].

[83] L. Breiman, "Random Forests," *Machine learning,* vol. 45, pp. 5-32, 2001.

[84] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in neural information processing systems 30*, 2017.

[85] H. Shi, "Best-first decision tree learning," Doctoral dissertation, The University of Waikato, 2007.

[86] "LightGBM," Microsoft Corporation, 2024. [Online]. Available: https://lightgbm.readthedocs.io/en/latest/Features.html. [Accessed 28 May 2024].

[87] S. Fadel, "Explainable Machine Learning, Game Theory, and Shapley Values: A technical review," Statistics Canada, 2022. [Online]. Available: https://www.statcan.gc.ca/en/data-science/network/explainable-learning. [Accessed 20 July 2023].

[88] A. Abohassan, "Assessing the Effects of Snow and Ice Control Operations: The Interdependency between Weather Variables, Maintenance Operations, Pavement Friction, and Collisions," 2021. [Online]. Available: https://era.library.ualberta.ca/items/11932dfb-202a-4387-ba79-b78b34a39e4d. [Accessed 14 12 2023].

[89] L. Demidova, E. Nikulchev and Y. Sokolova, "The SVM classifier based on the modified particle swarm optimization," *arXiv preprint arXiv,* vol. 3, no. 08296, pp. 16-24, 2016.

[90] T. Usman, L. Fu and L. F. Miranda-Moreno, "A disaggregate model for quantifying the safety effects of winter road maintenance activities at an operational level.," *Accident Analysis & Prevention ,* no. 48, pp. 368-378, 2012.

[91] X. Qin, D. A. Noyce, C. Lee and J. R. Kinar, "Snowstorm event–based crash analysis," *Transportation research record,* no. 1, pp. 135-141, 1948.

[92] L. Qiu and W. A. Nixon, "Effects of adverse weather on traffic crashes: systematic review and meta-analysis.," *Transportation Research Record,* vol. 1, no. 2055, pp. 139-146, 2008.

[93] F. Malin, I. Norros and S. Innamaa, "Accident risk of road and weather conditions on different road types," *Accident Analysis & Prevention,* vol. 122, pp. 181-188, 2019.

[94] J. Cervantes, F. Garcia-Lamont, L. Rodriguez, A. López, J. R. Castilla and A. Trueba, "PSO-based method for SVM classification on skewed data sets," *Neurocomputing,* vol. 228, pp. 187-197, 2017.

[95] H. Chen, I. C. Covert, S. M. Lundberg and S.-I. Lee, "Algorithms to estimate Shapley value feature attributions," *Nature Machine Intelligence,* pp. 1-12, 2023.

[96] "Winter in Edmonton," University of Alberta, 2024. [Online]. Available: https://www.ualberta.ca/international/international-student-services/life-in-canada/weather-in-edmonton/winter-in-edmonton.html#:~:text=Winter%20temperatures%20range%20from%20%2B5,in%20winter%2C%20so%20dress%20appropriately.. [Accessed 20 May 2024].

[97] Q. Cai, M. Abdel-Aty, J. Lee and H. Huang, "Integrating macro-and micro-level safety analyses: a Bayesian approach incorporating spatial interaction," *Transportmetrica A: transport science,* pp. 285-306, 2019.

[98] M. Wu, T. J. Kwon and K. EI-Basyouny, "A Hybrid Geostatistical Method for Estimating Citywide Traffic Volumes–A Case Study of Edmonton, Canada," *Journal of Geographical Research,* vol. 5, no. 2, pp. 52-68, 2022.

[99] S. L. Skszek, "State-of-the-art report on non-traditional traffic counting methods," Arizona. Dept. of Transportation, 2001.

[100] "Collect Events (Spatial Statistics)," Esri, [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/collect-events.htm. [Accessed 14 July 2024].

[101] "Hot Spot Analysis (Getis-Ord Gi*) (Spatial Statistics)," esri, 2024. [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/hot-spot-analysis.htm. [Accessed 2 June 2024].

[102] R. T. Tan, "Visibility in bad weather from a single image," in *2008 IEEE conference on computer vision and pattern recognition*.

[103] H. Renski, A. J. Khattak and F. M. Council, "Effect of speed limit increases on crash injury severity: analysis of single-vehicle crashes on North Carolina interstate highways.," *Transportation research record,* vol. 1665, no. 1, pp. 100-108, 1999.

[104] P. P. Jovanis and J. Delleur, "Exposure-Based Analysis of Motor Vehicle Accidents," *Transportation Research Record,* vol. 910, pp. 1-7, 1983.

[105] C. Plug, J. (. Xia and C. Caulfield, "Spatial and temporal visualisation techniques for crash analysis," *Accident Analysis & Prevention,* vol. 43, no. 6, pp. 1937-1946, 2011.

[106] M. L. Sirega, H. R. Agah and F. A. Arifin, "Median-type adjustment factor for road capacity calculation," *International Journal of Technology,* vol. 6, no. 5, pp. 762-769, 2015.

[107] J. C. Yu, "Median Visibility Improvements: Needs, Methods, and Trends," *Highway Research Record,* vol. 366, p. 92, 1971.

[108] T. Usman, L. Fu and L. F. Miranda-Moreno, "Accident prediction models for winter road safety: Does temporal aggregation of data matter?," *Transportation research record ,* no. 1, pp. 144-151, 2011.

[109] D. Santos, J. Saias, P. Quaresma and V. B. Nogueira, "Machine learning approaches to traffic accident analysis and hotspot prediction," *Computers,* vol. 10, no. 12, p. 157, 2021.

[110] R. Rodríguez-Pérez and J. Bajorath, "Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values," *Journal of Medicinal Chemistry,* vol. 63, no. 16, pp. 8761-8777, 2019.

[111] L. Auria and R. A. Moro, "Support vector machines (SVM) as a technique for solvency analysis," 2008.

[112] L. Wang, Support vector machines: theory and applications, vol. 177, Springer Science & Business Media, 2005.

[113] A. Ziakopoulos and G. Yannis, "A review of spatial approaches in road safety," *Accident Analysis & Prevention,* vol. 135, p. 105323, 2020.

[114] T. Zhang, L. Sun, L. Yao and J. Rong, "Impact analysis of land use on traffic congestion using real-time traffic and POI," *Journal of Advanced Transportation,* 2017.

[115] L.-Y. Chang and H.-W. Wang, "Analysis of traffic injury severity: an application of non-parametric classification tree techniques," *Accident Analysis & Prevention,* vol. 38, no. 5, pp. 1019-1027, 2006.

[116] K. Aas, M. Jullum and A. Løland, "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values," *Artificial Intelligence,* vol. 298, p. 103502, 2021.