Estimating Winter Road Friction Coefficients: An Integrated Approach using Machine Learning, Explainable AI, and Model Transferability

by

Xueru Ding

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

Ensuring road safety and efficient traffic movement through Winter Road Maintenance (WRM) operations is a pressing concern, particularly during harsh weather conditions. The challenge of accurately monitoring road friction coefficients, which play a crucial role in WRM, often leads to impractical and expensive solutions. To tackle such challenges, we introduce in this thesis a two-phase methodological framework that focuses on the development and optimization of machine learning (ML) models for road friction coefficient estimations, thereby bridging the gap between theoretical research and real-world application.

In the first phase, we concentrate on the precise estimation of road friction coefficients. Utilizing meteorological and geographic data from the Road Weather Information Systems (RWIS), we developed a Regression Tree model that achieved a high accuracy of 93.3%. To ensure spatially continuous friction estimations, we employed Ordinary Kriging interpolation to handle missing weather data. By categorizing road friction coefficients into distinct risk levels, we were able to provide critical insights on road surface conditions, achieving nearly 90% accuracy.

The second phase of our research emphasizes the refinement and augmentation of our models. We conducted a comparative analysis of ML algorithms, including Regression Tree, Support Vector Regression (SVR), Random Forest, and Extreme Gradient Boosting (XGBoost). We found a positive correlation between complexity and accuracy with XGBoost emerging as the most reliable model. To provide deeper insights into these models' inner workings, we leveraged SHAP explainable artificial intelligence (AI). We also examined the transferability of the XGBoost model, and after calibrating it for a new dataset, the updated model exhibited a significant improvement in accuracy on the new data, thereby affirming its robustness, adaptability, and transferability.

The framework presented in this thesis goes beyond theoretical modeling to offer tangible and innovative solutions that can be readily applied in the field of WRM. By utilizing meteorological and geographic data in tandem with advanced ML models, this research creates a pathway for more effective estimation and interpretation of road friction. The synergy of explainable AI, accurate estimation methods, and the successful transfer of models to new data sets demonstrates the potential of modern technology to transform traditional practices, ultimately for improved mobility and safety of winter travelling public.

PREFACE

The work presented in this thesis is either published or is under review for publication.

Journal Paper

Ding, X., and Kwon, T.J. Winter Road Friction Estimations via Multi-Source Road Weather Data—A Case Study of Alberta, Canada. Accepted for Publication in *Future Transportation* 2, no. 4: 970-987. December 2022.

Conference Proceedings

Ding, X., and Kwon, T.J. Road Friction Estimations Using Multi-Source Road Weather Data. *Presented at the 102nd Transportation Research Board Annual Conference,* Washington, D.C., United States, January 2023.

Journal paper (Submitted)

Ding, X., and Kwon, T.J. Enhancing Winter Road Maintenance with Explainable AI: SHAP Analysis for Interpreting Machine-Learning Models in Road Friction Estimation. Under review in the *Cold Regions Science and Technology*.

ACKNOWLEDGEMENT

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 invaluable, 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. Qipei Mei, and Dr. Jeff Boisvert, for being members of my defense committee and for providing valuable insights and feedback on my research results.

I would also like to sincerely thank each member of our research group for their joint efforts and help. 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 for their unwavering support and understanding throughout the research process. Their encouragement has been a constant source of motivation for me.

CONTENT

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURES

AI	Artificial Intelligence
ANN	Artificial Neural Network
AT	air temperature
CAM	Class Activation Mapping
CNN	Convolutional Neural Networks
COPD	Chronic Obstructive Pulmonary Disease
ESS	Environmental Sensor Stations
FR	False Risky
gcForest	Deep Forest
ICU	Intensive Care Unit
IDW	Inverse Distance Weighting
k-NN	K-Nearest Neighbor
LightGBM	Light Gradient Boosting Machine
LIME	Local Interpretable Model-agnostic Explanations
MAE	mean absolute error
ML	Machine Learning
mRWIS	Mobile Road Weather Information System
MSE	mean squared error
NB	negative binomial
OK	Ordinary Kriging
RH	relative humidity
RMSE	root mean square error
RSC	Road Surface Condition
RWIS	Road Weather Information System
SHAP	Shapley Additive Explanations
SHAP	SHapley Additive exPlanations
sRWIS	Stationary Road Weather Information System
ST	surface temperature
SVM	Support Vector Machine

SVR	Support Vector Regression
V2X	vehicle to everything
WRM	Winter Road Maintenance
XGBoot	Extreme Gradient Boosting

1 INTRODUCTION

1.1 Background

Throughout human history, transportation has been essential, serving vital functions such as commuting, shopping, socializing, or traveling. As vehicle speeds have increased and road infrastructure has improved, road safety has also become a growing concern. Each year in Canada, traffic collisions result in about 1,800 deaths and 150,000 injuries, including almost 10,000 serious injuries, costing society $40.7 billion annually [1]. These statistics highlight the urgency of addressing traffic safety concerns that require a concerted effort.

The urgency is especially pronounced in regions with lengthy winter seasons. Adverse weather conditions, such as snowfall, icing, and snow accumulation, reduce road friction, which in turn leads to dangerous travel and increased accident risk. During the winter months, vehicle crashes typically increase by 3.1-4.7% [2]. According to the National Collision Database, over 27 percent of collisions took place on wet, snowy, or icy roads in 2020, leading to 38,368 injuries and 508 fatalities [3]. Research also shows that accident rates are higher in low friction conditions due to poor speed adaptation, thus increasing risks on slippery roads [4]. Furthermore, traffic efficiency can be negatively impacted by adverse weather conditions. Studies have shown that driving on snowy or slushy roads can reduce vehicle speeds by 30 to 40 percent [5] and can delay travel by up to 50 percent [6], causing congestion and reduced travel speeds. These cascading effects emphasize the criticality of managing traffic mobility and road safety during winter.

Transportation agencies have recognized this challenge and have responded by implementing proactive measures such as timely mobilization of anti-icing operations, enhanced traffic monitoring, and improved early warning systems [7, 8]. These strategies are part of a broader focus on winter road maintenance (WRM), underlined by research that highlights its significance. Studies have demonstrated not only the safety benefits of proper road surface conditions (RSC) [9] but also the positive impact of WRM on friction, speed-stopping distance, and overall traffic flow [10]. Furthermore, these efforts are not merely about safety and efficiency; economic benefits of WRM have been shown to save road users significant costs on highway driving, thereby offsetting the cost involved in maintaining bare pavement conditions [11].

However, WRM's substantial benefits come with significant investment requirements. In 2021, more than 20 percent of Alberta government maintenance budgets, or $54.3 million were spent on snow and ice control operations [12]. This investment needs to be carefully weighed against maintenance benefits to ensure proper allocation. In this regard, accurate RSC information becomes vital for decision-making.

Currently, most jurisdictions use descriptor measures like wet, snow-covered, and icy roads to monitor winter conditions [13, 14]. However, there is no uniform standard for these descriptions, which can easily lead to ambiguity in the information provided. In contrast, using friction as an RSC indicator offers an objective and accurate metric for WRM operations, facilitating the identification of high-risk road sections. But to fully capitalize on this advantage, specific tools and methods are needed.

Among different types of road monitoring systems, Road Weather Information Systems (RWIS) have emerged as a potential solution to this challenge. There are two main types of RWIS: stationary RWIS (sRWIS) and mobile RWIS (mRWIS). sRWIS gathers data from environmental sensor stations (ESS) installed along the road network but is limited spatially while mRWIS continuously collects road weather data but is limited temporally. These systems offer a practical means of implementing the friction-based approach through each comes with its own set of limitations.

The strategies and statistics detailed above underscore the complexity and importance of maintaining safe and efficient road conditions during winter. Traditional monitoring systems, while valuable, have revealed areas for improvement, such as the reliance on ambiguous descriptor measures. This illustrates the necessity for a more accurate and unified approach; namely, a friction-based RSC monitoring and estimation method. By offering an objective and accurate metric for WRM operations, it holds the potential to significantly enhance efficiency and safety for both maintenance authorities and road users.

1.2 Problem Statement and Motivation

As illustrated in the previous section, accurately understanding road surface conditions (RSC) is crucial for an effective winter road maintenance (WRM) operation. The friction coefficient,

recognized as one of the most objective measures of road conditions, underscores the need for an in-depth study on accurate friction estimation.

Road conditions are largely influenced by weather factors. Despite this, current research on utilizing weather information for RSC estimation remains insufficient. Existing studies predominantly focus on model construction with less emphasis on practical implementation and efficient application. This leaves a gap in understanding how to leverage road weather and surface conditions information effectively and how to integrate both mobile and stationary Road Weather Information Systems (mRWIS and sRWIS) to optimize estimation accuracy.

Machine learning (ML) methods present promising avenues for modeling road surface friction with studies employing techniques like support vector machine (SVM) and Neural Networks [15, 16]. However, the diverse performance of ML models, which may vary depending on the datasets used for training [17] along with a lack of comprehensive discussion on continuous friction estimation models highlights key obstacles. In addition, ML models have limitations in terms of generalization; they can only make accurate predictions based on the data they were trained on. Encountering unfamiliar datasets could cause their performance to suffer. Consequently, strategies such as transfer learning might then be essential to broaden the models' usability, applicability, and adaptability.

Another challenge lies in the complex nature of ML models. As complexity increases, interpretability decreases [18, 19, 20], obscuring the understanding of their inner workings. This complexity introduces considerable risks if the models are to be applied in real-world scenarios. To mitigate these risks, a rigorous examination of their reliability and the transparency of decision-making systems may compel the incorporation of explainable artificial intelligence (AI).

These challenges and the current gaps in research serve as the driving force behind this thesis. The multifaceted problem landscape — marked by the complexity of accurately estimating RSC friction, the shortcomings of current methods, the potential and pitfalls of ML approaches, and the pressing need for pragmatic solutions — calls for an in-depth exploration. A thorough investigation of these issues holds the key to enhancing existing RSC monitoring and estimations, potentially leading to breakthroughs in WRM operations, winter traffic safety, and efficiency.

1.3 Objectives

Building on the challenges and gaps identified in the previous sections, the accurate estimation of road friction coefficients emerges as a pivotal concern for ensuring efficient real-time winter road maintenance (WRM) operations. The intricate interplay between weather factors, machine learning (ML) models, and practical application strategies reveals a complex landscape that existing research has yet to fully explore.

Therefore, the primary objective of this thesis is to develop a robust friction estimation model, incorporating weather and geographic data, for practical use in winter road maintenance. The research is divided into two main phases. In phase one, a friction estimation model that outputs real-time, spatially continuous road friction estimation is constructed using weather and geographic information. In phase two, the focus shifts to comparing model performance, exploring the internal workings of the models through explainable artificial intelligence (AI), and demonstrating the transfer learning's ability to broaden the models' applicability. To achieve these overarching goals, the research has the following specific tasks:

1) Develop a robust friction estimation model that incorporates both weather and geographic information and apply it to practice with geostatistical interpolation techniques to obtain spatially continuous road friction estimates;
2) Generate a road risk map based on road friction estimates to assist WRM personnel in making timely and effective decisions and provide information to road users;
3) Conduct a comprehensive comparative analysis of well-adopted ML algorithms to determine the most suitable one for accurate friction estimation;
4) Leverage explainable AI to gain insight into the inner logic of complex ML models to improve model transparency and interpretability; and
5) Enhance the transferability of the friction estimation model by employing transfer learning techniques to expand its application scope.

By methodically pursuing these goals, this thesis aims to deliver a method for estimating continuous friction values that are accurate and intuitive. Through these concerted efforts, this thesis provides practical insights that could improve decision-making in WRM, ultimately leading to enhanced safety and efficient resource allocation.

1.4 Thesis Structure

This thesis is split into six chapters. Chapter 1 introduces the motivation behind the work involved and the research objectives.

Chapter 2 is the literature review section outlining existing research on road condition estimation methods, the factors affecting road friction coefficient, and the use of explainable artificial intelligence (AI) and transfer learning in machine learning (ML) models. This review serves as the theoretical and methodological foundation for the subsequent chapters.

Chapter 3 introduces the principles behind the methods employed in this study, providing a detailed overview of ML algorithms, interpolation methods, SHapley Additive exPlanations (SHAP) explainable AI techniques, and the two-stage boosting algorithm designed for robust prediction.

Chapter 4 presents a comprehensive description of the development process of the friction estimation model, including the generation of the data set, training method for the model, incorporation of geostatistical interpolation techniques, model application and performance results. Additionally, a road risk classification system is constructed to generate road friction risk maps.

Chapter 5 delves into three following aspects to further enhance the efficiency of the proposed framework. Firstly, a thorough comparative analysis of various ML algorithms, including Regression Tree, Random Forest, Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost), is conducted to select the optimal model for road friction estimation. Secondly, the intrinsic logic of the models is explored in depth using SHAP explainable AI to enhance its interpretability. Lastly, transfer learning techniques are applied to expand the application scope of the model.

Chapter 6 concludes the thesis with a summary and discussion of the research outcomes and contributions. Furthermore, the chapter discusses the challenges and limitations encountered during the study and proposes suggestions for future research directions.

2 LITERATURE REVIEW

In this chapter, we will examine previous studies on road friction estimation and the effectiveness of machine learning (ML) algorithms. The chapter is divided into four sections. The first section will cover the current methods used for estimating road friction. The second section will explore how weather factors affect road conditions. In the third section, we will investigate how to improve the interpretability of ML methods. Finally, we will provide a summary and note any limitations found within the reviewed literature.

2.1 Methods for Road Conditions Monitoring and Estimation

Monitoring and assessing road surface conditions (RSC) is crucial for improving traffic flow and enhancing road safety. In terms of the performance measure used to perform this operation, there exist several metrics: visual indicators, time to normal (or bare pavement regain time), traffic speed regain time, and road friction levels [21], of which the visual indicator and road friction are the most commonly used in monitoring RSC.

The Finnish Road Agency [22] classified road conditions based on factors such as road friction level, RSC, and slipperiness classification. They separated road conditions into six categories: wet ice, icy, packed snow, rough ice/ packed snow, clear and wet, and clear and dry. Each category was discovered to correspond to a specific friction range, indicating a negative correlation between road condition and friction coefficient. For example, wet ice corresponds to the lowest friction range, while clear and dry corresponds to the highest friction range.

Ilkka et al. [13] proposed a method based on statistical equations to model road surface friction. By studying the correlation between RSCs and various influencing factors, they found that the predictor behavior depended on the RSC itself. Consequently, models would have to be developed separately for each condition type. The authors did not develop models for every condition but instead focused on two, snow- and ice-covered roads and wet roads. They used three linear regression functions to estimate the effects of snow- and ice-covered roads, water-covered roads, and existing snow, ice, and water on friction, respectively. These models were validated and showed R-squared values of more than 0.8, indicating that the models performed well and were able to predict road friction.

Similarly, the RoadSurf simulation model proposed by Kangas et al. [23] estimated road friction separately depending on the road conditions. The road friction estimation model was divided into three sub-models: snowy and/or icy roads, wet road surfaces, and dry road surfaces. On dry road surfaces, friction was assumed to be constant (0.82), while for wet road surfaces, a model was built with water layer thickness being the only predictor. Lastly, the model for snowy and icy roads incorporated road temperature, snow layer thickness, and ice layer thickness as factors influencing friction.

In recent years, due to advancements in machine learning (ML), ML has become the dominant method for building road friction estimation models. The advantage of this approach is that instead of building multiple sub-models for different situations, a comprehensive model that deals with a large range of conditions can be trained, which significantly simplifies the model development process.

Takasaki et al. [24] developed a RSC estimation model using Random Forest techniques. They classified the road surface into four categories (dry, wet, slushy, and snowy) and investigated the relationship between road conditions and weather, traffic, and tire noise. To determine whether tire noise significantly affects road conditions, researchers constructed two models for comparison, one with and one without tire noise. The result showed that the accuracy of the two models was very close, 94.8% and 94.7%, respectively, indicating that tire noise had a negligible effect on road conditions compared to weather and traffic factors.

Take and Duran [25] constructed a k-nearest neighbor (k-NN) model to estimate RSCs based on weather data. The conditions considered in this study were dry, wet, salty-wet, and icy. To assess the performance of the k-NN model, they compared it with other classification models such as logistic regression, support vector machine (SVM), naïve Bayes, decision tree classifier, and Random Forest classifier. The result indicated that the k-NN model performed the best, with an impressive accuracy of 99.71%.

Minge [16] employed an artificial neural network (ANN) algorithm to estimate the classification of road friction coefficients that were classified into six categories and two categories, respectively. The model used both vehicle sensor data and weather data as influencing factors. To verify the significant effect of weather data on road friction, he built two models, one using only vehicle sensor data and the other with both vehicle and weather data. The results showed that adding

weather variables significantly improved the model's performance; the estimation accuracy increased by 3.7%, from 45.27% to 49.03% in the six-classification problem. And in the two-classification problem, the model performance improved by 21.5 percent, from 54.47% to 75.97%. This result demonstrated the positive impact of weather data on estimating road friction.

Linton and Fu [26] proposed a road condition monitoring system by combining vehicle imagery with road weather data to determine if the road condition is bare, partially snow-covered, or fully snow-covered. The estimation process involves using the images to get an initial estimate, followed by the involvement of weather data to obtain more accurate results. ML algorithms explored in this study are ANNs, random trees, and Random Forest. The results of the study showed that the accuracy of the classification results was improved by an average of 18% through the usage of this two-step estimation approach. Moreover, among the three models, the Random Forest model showed the best performance.

As mentioned above, most of the existing studies treat road state estimation as a classification problem. The advantage of this approach is that it provides an intuitive interpretation of the RSC. However, it may lead to inconsistent measurement results due to the subjective nature of descriptive road condition labels. In comparison, friction coefficient is a much better measure because it is a device-measured value that does not require user interpretation in the collection process. Nevertheless, there are only a few studies that attempt to estimate friction coefficients.

Kim et al. [27] performed one of the few studies that attempted to estimate friction. They developed an estimation model using an ANN with rainfall intensity, water film thickness, and road temperature as inputs and friction coefficient as output. To provide a simplified way of interpreting the outputted friction values, they labeled the samples into three categories based on their predicted friction value. These three classes include hydroplaning, wet, and moist road conditions, corresponding to 0.67-1.00, 0.34-0.66, and 0-0.33, respectively. By evaluating the confusion matrix classification for wet roads, the results showed an accuracy of 92% for the hydroplaning condition, 100% for moist, 78% for wet, and an overall accuracy of 94%. Such high accuracy indicated that the model is able to accurately predict the RSCs during rainy weather. The results of these studies demonstrate the feasibility of estimating road friction coefficients with weather information.

2.2 Factors Affecting Road Friction

Road friction is influenced by a variety of factors, including intrinsic factors such as tires and pavement material, as well as external factors like precipitation that influence the amount of contaminants present on the roadway [28]. The independent effects of these factors will be explored in detail to provide a clear explanation of their influence on road surface friction.

Weather events such as rainfall and snowfall can directly change the wetness and friction characteristics of the road surface. Wet road surfaces reduce the contact between the tires and the road surface, thus reducing friction. Do et al. [29] investigated the effect of thin water films (< 1 mm) on road friction that occurs after rainfalls or during drizzles. The results showed that friction decreased significantly before the water film thickness reached 0.4 mm, indicating that thin water films can significantly alter the available road friction by reducing contact between the tire and the road. Snow-covered roads and icy roads can reduce adhesion to the surface, which further reduces low friction, and vehicles are more likely to slip under these road conditions. According to the findings of Ichihara and Mizoguchi [30], the skid resistance coefficient on ice ranges from 0.1 to 0.2, while on fresh snow it ranges from 0.2 to 0.25, denoting that low friction is linked to snow and ice presence.

Air humidity is another important factor that affects road friction. In a high-humidity environment, the moisture in the air interacts with the road surface to form a thin film of water, leading to a decrease in the friction coefficient of the road surface. On the contrary, under low humidity conditions, there is less moisture in the air, and the road surface is relatively dry, thereby increasing the friction coefficient of the road surface. In general, changes in air humidity directly affect the friction characteristics of the road.

As for temperature, it mainly affects road friction through thermal effects [31]. Under high-temperature conditions, heat is transferred to the road, promoting the melting of snow and ice accumulation and accelerating the evaporation of water. However, under low-temperature conditions, the temperature affects the properties of the road material itself, which leads to a decrease in the coefficient of friction.

While geographic conditions are generally constant when compared to ever-changing weather conditions, they still have a significant impact on road conditions. Among the geographical factors,

latitude and altitude have a significant contribution to road friction, mainly through an indirect effect on road temperature.

The influence of latitude on road temperature is due to the different amounts of solar radiation received at different latitudes. Mildrexler et al. [32] utilized satellite data to create maps depicting the annual maximum temperature of the Earth; it can be observed that the temperature reaches its highest value near the equator and shows that temperature decreases when latitude increases. At high latitudes, the angle of direct sunlight is lower and solar radiation is relatively weaker, resulting in lower road temperatures. Conversely, at low latitudes, the angle of direct sunlight is higher and solar radiation is stronger, resulting in higher road temperatures. Therefore, the variation of latitude leads to the difference in road temperature in different regions, which in turn affects the magnitude of road friction.

In addition, Altitude also has a negative correlation with temperature. As altitude increases, both atmospheric pressure and air temperature decrease. For every 100 m increase in altitude, the average temperature drops by about 0.7°C [33]. Since the drop in temperature affects road surface temperature, altitude also has an effect on road friction. In general, the lower temperatures at higher altitudes make the road temperature lower as well, thus increasing the friction of the road surface. Conversely, higher temperatures at lower altitudes result in the reduction of road friction.

Furthermore, traffic factors have a significant impact on road friction as well. Generally, high traffic load and speed can cause a decrease in road friction. Rasol et al. [34] conducted a literature review and found that road friction tends to decrease as vehicle speeds increase on wet road surface surfaces. Additionally, traffic can impact road friction through its effect on the condition of road coverings, such as snow and ice. Ichihara and Mizoguchi [30] observed a notable decrease in the coefficient of friction for roads exposed to traffic. Specifically, the coefficient of friction on newly compacted snow layers by traffic can be as low as 0.10 to 0.15. On frequently encountered compacted snow or ice surfaces, the coefficients typically range from 0.2 to 0.3.

2.3 Improving Machine Learning Interpretability and Transferability

This section addresses two essential aspects of machine learning (ML): enhancing interpretability through explainable artificial intelligence (AI) and improving transferability through transfer

learning. These considerations are vital in creating models that are not only transparent in their decision-making but also adaptable to different data scenarios, aligning with the thesis's aim to develop robust and reliable applications.

Improving Interpretability by Explainable AI

ML models have gained popularity in recent years because of their superior performance over traditional statistical model techniques. However, the lack of clarity on their decision-making process makes them difficult to incorporate due to reliability and trustworthiness concerns.

Traditionally, the evaluation of ML models relied heavily on accuracy as a metric. However, Doshi-Velez and Kim [35] pointed out that solely relying on accuracy was not sufficient for evaluating the models. Even though some models may perform well in terms of accuracy, their underlying logic may not be consistent with reality. Carvalho et al. [36] indicated in their study that ML models could have serious consequences when supporting high-stakes decisions. Examples include the incorrect release of a potentially dangerous criminal [37] and misclassifying a dangerous situation as safe in a pollution model [38]. Because ML models lack transparency and accountability, the reasons behind these errors are not clearly understood.

A notable example provided by Caruana et al. [39], referred to as “Pneumonia – Asthma”, illustrates the consequences and risks associated with incorrect model predictions. A ML model that predicts pneumonia risk incorrectly learned a rule that individuals with asthma and heart disease had a significantly lower risk of pneumonia-related death compared to healthy individuals, completely contradicting the truth. Such erroneous result was caused by the fact that pneumonia patients with a history of asthma were usually admitted directly to the Intensive Care Unit (ICU) and received effective treatment, which lowered their risk of death. Because the prognosis of these patients was better than average, it resulted in the model formulating the wrong relationship.

Given the potential high risk associated with the usage of ML models, improving the interpretability of ML becomes essential. Doshi-Velez and Kim [35] emphasized the significant gap between the complexity of ML models and the practical need for interpretability. To counteract this problem, a systematic framework was proposed to guide the interpretation methods of ML models. By making the model transparent and providing reliable information, we can ensure the

reliability of decisions and reduce the possibility of misleading and wrong judgments, thus improving the reliability and acceptability of ML model applications.

Model interpretability can be categorized into intrinsic and post hoc interpretability [18]. Intrinsic interpretability refers to models that have a relatively simple structure and can be directly interpreted, while post hoc interpretability involves applying specific methods or techniques post-model development. Given the poor interpretability of complex ML models, there is considerable interest in the study of post hoc explainable models. Currently, SHapley Additive exPlanations (SHAP) has emerged as a powerful post hoc model that has been employed in several fields. By applying game theory concepts, SHAP calculates Shapley values to quantify the contribution of each feature to the model prediction and represent each feature's average marginal impact when considering all possible combinations of features [40, 41]. In addition, SHAP enables users to distinguish the relative contributions of different features by combining the computed Shapley values using a linear additive structure. The strong theoretical foundation of the SHAP model enables it to produce highly credible explanations of ML models.

Currently, SHAP has been widely used in the medical field. Liu et al. [42] combined SHAP models with ML models to solve the problem of high dimensional feature datasets in Parkinson's disease diagnosis that make the models difficult to interpret. They combined four models, deep forest (gcForest), Extreme Gradient Boosting (XGBoost), light gradient boosting machine (LightGBM), and Random Forest with SHAP, then evaluated and compared their performance. The results showed that SHAP-gcForest performs the best, with a classification accuracy of 91.78% and an F1 score of 0.945. However, the computation is time-consuming because of the use of kernel methods. Considering both model performance and computational efficiency, SHAP-LightGBM was identified as the most suitable option for practical applications. Kor et al. [43] utilized the SHAP model to provide a visual interpretation of individual risk prediction in a chronic obstructive pulmonary disease (COPD) model. This approach proved to be beneficial for clinicians as it facilitated understanding of the contribution of each feature in the model and enhanced comprehension of the decision-making process of the model. Similarly, Zhang et al. [44] integrated SHAP into a risk model for acute kidney injury after liver transplantation, allowing for the interpretation of ML models at the individual level and assisting clinicians in making informed decisions.

In an effort to help the emergency department to identify patients properly, Duckworth et al. [45] combined the SHAP model with a machine-learning model to monitor data drift. The study's results illustrated that the AUROC of the model remained above 80% for two different periods, indicating that the model is both accurate and reliable. However, significant changes in the importance of features were observed through the SHAP model results. By tracking changes in SHAP values relative to global importance, the researchers identified a complementary measure of data drift, which highlights the need to retrain a predictive model. Furthermore, by observing relative changes in feature importance, emergency health risks were identified.

Farzaneh et al. [46] proposed a framework combining ML, SHAP, and expert validation to aid in predicting traumatic brain injury prognosis risk. Initially, they developed a ML predictive model and then used SHAP to calculate the global contribution of each feature to the prediction. Subsequently, a human expert team interpreted the SHAP results based on their knowledge. This process ensured the credibility of the entire prediction process. The results showed that the machine-learning model achieved an accuracy of 0.7536. Then, with the help of SHAP, the non-robust features that had a minimal contribution to the predictions were excluded, resulting in a new model with an accuracy of 0.7440. After that, via expert knowledge, 3 out of the 21 significant contributing features to the SHAP results were identified as unintuitive and eliminated, leading to a final model accuracy of 0.7488. This variation in model accuracy demonstrates the importance of selecting appropriate input features when using data-driven algorithms.

In addition, SHAP has found applications in the financial field. Roa et al. [47] employed TreeSHAP to facilitate the interpretation of credit risk models with stochastic gradient boosting algorithms. Lin and Gao [48] proposed a group SHAP method to reveal the common characteristics of companies engaged in fraudulent financial activities by measuring the combined effect of a set of features. Analyzing the results generated by the group SHAP method is crucial for customer monitoring and investment management, as it assists financial institutions in better identifying potential risks and safeguarding customer interests. Xia et al. [49] conducted a comparative study of seven feature selection techniques to generate reliable model predictions. The study results revealed that the method incorporating the SHAP model outperformed the other models, and the generated SHAP values aligned more closely with human intuition. This suggests that the SHAP

model is particularly suitable for feature selection and can provide more reliable and explanatory model predictions.

In some existing studies, the SHAP model has been compared with another widely used model of post hoc explainable AI, known as the Local Interpretable Model-agnostic Explanations (LIME) model. Different from SHAP, the LIME model generates only local explanations by perturbing data points and using simple linear regression to fit the relationship between features and the local prediction [50]. Gramegna and Giudici [51] evaluated the discriminative power of credit risk by using the SHAP and LIME methods. By comparing the bidimensional plots generated by the two methods, it was evident that the SHAP method was more effective in dividing the two clusters in space. In terms of AUC values, the mean value of the SHAP method after 50 repetitions was 0.864, while the mean value of the LIME method was slightly lower at 0. 839. After careful consideration, the researchers concluded that the SHAP was superior to the LIME in this task.

Although the LIME model may have outperformed the SHAP model in some studies [52], overall, SHAP produces more stable explanations. The study by Hailemariam et al. [53] focused on the robustness and security of explainable AI models. They evaluated the performance of LIME and SHAP on two different types of datasets, tabular and image datasets. This process was done using Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), two deep learning methods. Evaluation metrics, including interpretation invariance, identity, separability, and stability, were employed. The results showed that the SHAP model performed better than the LIME model on both tabular and image datasets in a security-sensitive domain. Another study by Nguyen et al. [54] evaluated three methods, SHAP, LIME and Class Activation Mapping (CAM), for image classification problems. The results demonstrated that the SHAP method outperformed the other two methods, followed by the LIME method. However, it should be noted that these methods came at the expense of calculation efficiency.

Improving Transferability by Transfer Learning

Models based on data-driven ML typically work best when the data they analyze is similar to the data used to train them. However, if there are changes in the characteristics of the dataset, such as a different location for data collection, the model's performance may suffer. This may require retraining the model from scratch, which can be time-consuming and resource-intensive. To speed up this transition, transfer learning [55] is a useful technique. This involves adding a small amount

of new data to an existing model and allowing the new data to modify a portion of the model's learned relationships.

The benefits of transferring learning were demonstrated by Dai et al. [56], who proposed a boosting-based algorithm called TrAdaBoost. TrAdaBoost updates an existing model by combining source data with a small amount of new target data. The method was evaluated on the support vector machine (SVM) classification problem and compared with the SVM baseline models. The results indicated that the accuracy of TrAdaBoost was improved by at least 5% relative to the SVM baseline model, highlighting the potential of TrAdaBoost in effectively dealing with problems involving a change in dataset characteristics.

The study by He et al. [57] applied TrAdaBoost for mobile LiDAR classification. They proposed a structure that combined VoxNet and Multiclass TrAdaBoost. Evaluation results show that this approach not only improved accuracy but also achieved more balanced performance in each category. Similarly, Tang et al. [58] conducted a study comparing the effectiveness of TrAdaBoost, AdaBoost, and traditional negative binomial (NB) models for traffic crash prediction using four different datasets. The idea is to select the most valuable instances as new data for transfer learning. The results of the study showed that the transfer learning approach significantly outperformed the traditional NB model, with TrAdaBoost exhibiting the highest accuracy. This demonstrates the strong performance of TrAdaBoost in regression problems and its potential for addressing challenges arising from changing dataset characteristics.

2.4 Summary

This chapter provides an overview of models used to estimate road friction coefficients and explores the factors affecting road friction coefficients. It also discusses methods for enhancing the efficiency of machine-learning approaches.

Traditionally, road condition estimation models have usually been treated as a classification problem, where visual indicators or risk levels are used to categorize friction coefficients. However, this classification approach has limitations in accurately accounting for situations that fall within the ambiguous boundaries of these categories. Therefore, there is a need to develop models that can estimate continuous road friction values more effectively.

To develop a suitable model with appropriate inputs, we delve into a detailed discussion of the factors that affect road friction coefficients, which serves as a foundation for our modeling efforts. In addition, as discussed, machine learning (ML) faces two main challenges: difficulties in interpreting model decisions and the lack of adaptability to new datasets with slightly different characteristics such as location. To cope with these challenges, we introduce the SHapley Additive exPlanations (SHAP) explainable artificial intelligence (AI) model to improve the interpretability of the model, and we adopt the two-stage TrAdaBoost.R2 method to improve the transferability of the model. Through this process, the ML models we developed can improve their reliability and efficiency.

3 METHODOLOGY

The previous chapter highlighted the shortcomings of road friction estimation models in terms of their construction, application, and optimization. To address these gaps in current research, this thesis aims to establish a machine learning (ML)-based framework for accurate estimation of road friction coefficients (as depicted in **Figure 3-1**). The framework consists of two main phases. In the first phase, we develop a friction model and propose ways in which it can be practically applied. The second phase focuses on model optimization by comparing the performance of different ML algorithms, leveraging explainable artificial intelligence (AI) to improve model transparency and interpretability, and employing transfer learning to enhance model transferability. This comprehensive workflow allows us to further understand the potential of ML models in providing accurate road friction estimates and assess their practical feasibility. Within this chapter, we delve into the algorithms employed in this framework and the underlying principles behind them.

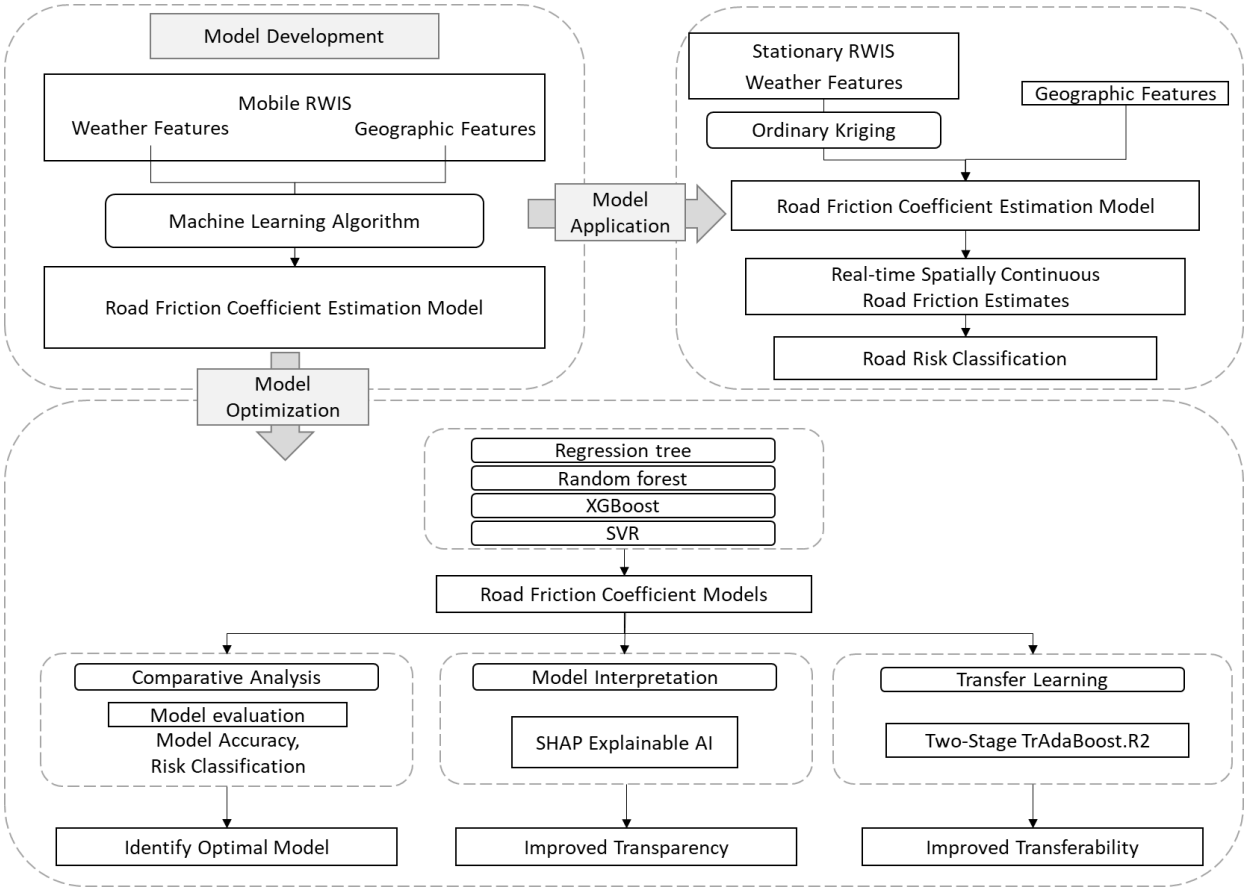


Figure 3-1 Methodological Framework

3.1 Machine Learning Algorithms

Machine learning (ML) algorithms serve as the foundation of the estimation model in this study. Among the many methods available, we select Regression Tree, Random Forest, Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) to develop our friction estimation model. Excluding SVR, the remaining three algorithms are all tree-based, each with unique characteristics and advantages. Herein, phase one will focus solely on the Regression Tree model, and in phase two, Random Forest, XBGoost, and SVR will be introduced for comparison purposes. Details of each algorithm are given below.

3.1.1 Tree-Structured Models

3.1.1.1 Regression Tree

Regression Tree [59, 60], as the name suggests, models the relationship between inputs and outputs similar to tree growth, as shown in **Figure 3-2**. The dataset is separated into two subsets at the root node according to different conditions. Data that satisfies the corresponding condition flows to the following internal node, where another condition exists to split the data further. Eventually, the tree will stop growing because of growth restrictions. The main advantages of using a Regression Tree are that it is intuitive, non-parametric, and the decision criteria are visible and easy to interpret.

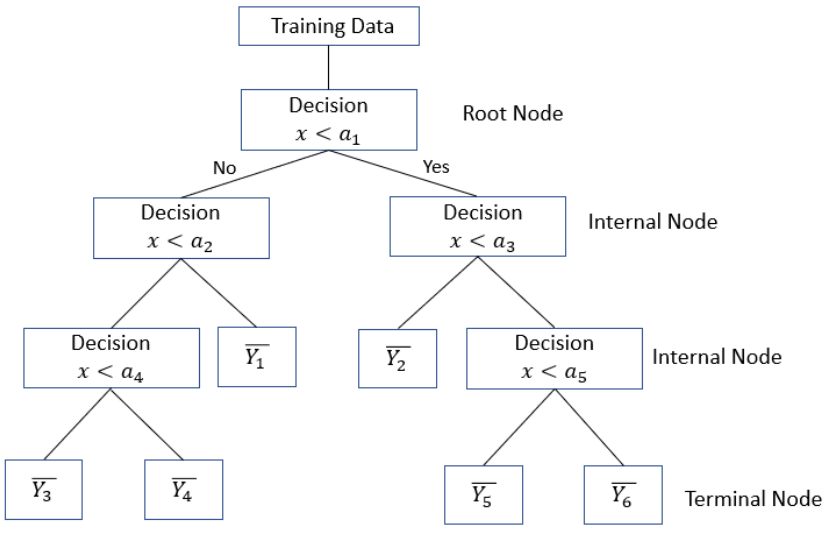


Figure 3-2 Regression Tree Visualization

The splits of Regression Tree are determined by a traverse approach with the least square error criterion. The decision condition at each node, including an input feature and a split value, is selected by calculating the minimum residual sum of squares between observations and the mean. This process is expressed as **Equations (3.1)** and **(3.2)**.

$$\boldsymbol{min_{j,s}[min_{c1}} \sum_{\boldsymbol{x_i \in R_1(j,s)}} \boldsymbol{(y_i - c_1)^2 + min_{c2}} \sum_{\boldsymbol{x_i \in R_2(j,s)}} \boldsymbol{(y_i - c_2)^2]} \quad \textbf{(3.1)}$$

$$\boldsymbol{c_m = ave(y_i | x_i \in R_m)} \quad \textbf{(3.2)}$$

Where j is the split feature, s is the split value, R_1 and R_2 are subsets divided by condition (j, s), c_m is the mean of subset R_m, x_i and y_i are observations.

Despite its many advantages, the Regression Tree has an inherent flaw; it is prone to overfitting. As such, model parameters should be tuned carefully to avoid this issue, which are maximum depth, minimum samples split, and minimum samples leaf. Maximum depth refers to the model complexity. The deeper the tree, the more information it captures. Minimum sample split determines the minimum number of samples that are required to split an internal node; a split will not happen if there are less than a certain number of records specified by this parameter. Finally, the minimum samples leaf determines the minimum number of samples at the leaf node. To determine reasonable values for these parameters, the grid search algorithm was applied, which is an exhaustive search method that iterates through all possible parameter candidates to find the most optimal configuration. In this study, we used r squared as the performance selection criteria based on prior studies.

3.1.1.2 Random Forest

Random Forest [61, 62] is an ensemble learning algorithm that utilizes Regression Trees as its base learners. It randomly selects subsets of the training data and input features to train multiple individual Regression Trees. Each tree in the Random Forest independently generates estimations, and the final output is obtained by aggregating the estimates from all the trees, typically through a weighted average.

Compared to the Regression Tree algorithm, Random Forest exhibits improved robustness and accuracy because of its ability to mitigate overfitting by limiting the data and features used for each tree. However, it is important to note that the improved performance of Random Forest comes

at the expense of reduced interpretability. With a large number of trees and random feature selection, the resulting model becomes more complex and challenging to interpret compared to a single Regression Tree. Therefore, although Random Forest improves predictive power, understanding its underlying decision-making process may be more difficult due to its complex structure.

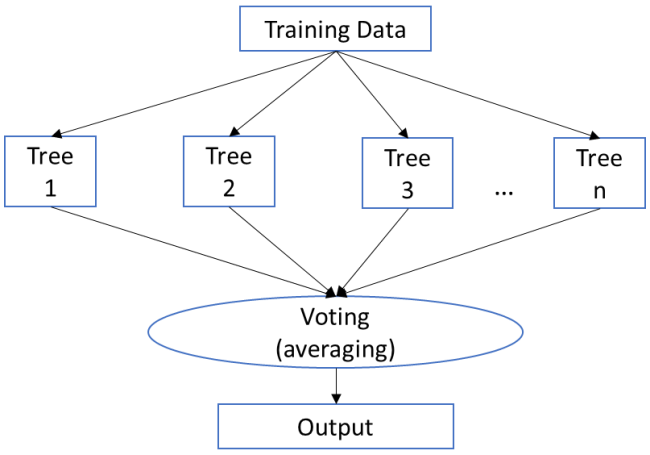


Figure 3-3 Random Forest Visualization

3.1.1.3 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) [63] is an algorithm designed for high execution speed and high accuracy. It belongs to the ensemble learning family and leverages the gradient boosting technique to sequentially combine multiple weak estimators.

The XGBoost process begins with the initialization of a simple Regression Tree. It then calculates the residuals between the predicted values and the true values for each sample. A new tree is then added to model the residuals generated from the previous iteration, after which the model recalculates the residuals to account for the newly added tree. This cycle repeats until a stopping condition is met. To optimize the model, XGBoost employs the gradient boosting algorithm to minimize the objective function and adjust the model parameters. The objective function consists of two components, the loss function and the regularization. Where the loss function measures the prediction error of the model, while the regularization controls the complexity of the model to prevent overfitting, the specific formula is as **Equation (3.3)**:

$$\boldsymbol{Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i{}^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + constant} \quad \textbf{(3.3)}$$

Where l is the loss function, Ω is the regularization, f_t is a Regression Tree, y_i is the true value, and $\hat{y}_i$ is the predicted value.

The final XGBoost model surpasses the accuracy of individual Regression Trees in sequence.

3.1.2 Support Vector Regression Model

Support Vector Regression (SVR) [64, 65] works differently from the aforementioned tree-based models. By mapping the dataset to a high-dimensional space, SVR aims to find a hyperplane that maximizes the margin distance between the hyperplane and the nearest data point and minimizes the prediction error at the same time. The objective function is shown as **Equation (3.4)**:

$$\begin{aligned} \boldsymbol{minimize} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i{}^*) \\ \boldsymbol{subject\ to} \quad & \begin{cases} y_i - \langle w, x_i\rangle - b \le \varepsilon \\ \langle w, x_i\rangle + b - y_i \le \varepsilon \end{cases} \end{aligned} \quad \textbf{(3.4)}$$

Where $\|w\|^2$ is the squared norm of the weight vector w, used for regularization, C is the regularization parameter, and ξ_i and $\xi_i{}^*$ are slack variables.

The principle of SVR gives it an advantage when dealing with high-dimensional data and nonlinear relationships. However, the mechanism by which the model constructs a hyperplane leads to difficulties in explaining the logic behind its output predictions.

As discussed above, given the varying levels of complexity, accuracy, and interpretability in these four models, this study's challenge lies in navigating the trade-offs among these aspects. The Regression Tree offers transparency at the potential expense of accuracy, while Random Forest, SVR, and XGBoost progressively increase in complexity and predictive power but decrease in interpretability. Hence, in the initial phase, the Regression Tree is chosen due to its elevated interpretability, providing a reassuringly intuitive understanding of the accuracy of our model's logic. And in the second phase, we introduce two additional tree models of different complexity to better explore and compare the trade-off between accuracy and interpretability. We also present

a SVR model to examine how algorithms based on different underlying mechanisms perform in the context of road friction estimation using road weather data. Through this approach, the study aims to provide insights that can guide the selection and utilization of machine learning (ML) models in transportation applications, taking into account both the predictive needs and the necessity for transparent and understandable models.

3.2 Ordinary Kriging

The goal of spatial interpolation is to estimate values at unsampled points using observed data points. Among the many spatial interpolation methods available, kriging has been shown to produce the most accurate estimates [14, 66, 67]. The kriging estimates consist of deterministic trends and residuals. Ordinary Kriging (OK) [68] is one variant of kriging that is considered the most widely employed among all the Kriging methods due to its simplicity and high accuracy. For this reason, OK has been selected in this study. OK assumes an unknown but constant mean over the area. This model can be expressed as **Equation (3.5)**.

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

Where μ is the unknown constant, and $\varepsilon(s)$ is the residual. Similar to Inverse Distance Weighting (IDW) method, weights are needed in kriging to estimate values at unknown points. However, the difference is that the primary assumption of kriging is spatial autocorrelation, implying an internal spatial relationship between sample points. Kriging weights therefore depend on both the distance between observations and prediction location and the overall spatial arrangement of the observations. Hence, the objective of OK interpolation is to determine the optimal kriging weights that minimize the estimation variance. The OK estimate is calculated by **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 $Z(s_i)$.

Semivariance represents the reciprocal of the spatial autocorrelation, which is calculated using **Equation (3.7)**.

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

where $\gamma(h)$ is the semivariance between sample points x_i and x_j in h distance, and z is the feature value. An empirical semivariogram is generated using Equation (3.7) to explore the spatial autocorrelation pattern of the observations. Afterward, several theoretical models are considered to fit the empirical semivariogram model. These include circular, spherical, exponential, gaussian, and linear models. The fitted semivariogram model provides three spatial parameters: sill, range, and nugget [67]. The sill is the semivariance at which the model begins to plateau, and the range is the lag distance where the semivariance reaches the sill, beyond which the spatial autocorrelation is considered non-existent. Finally, the nugget is the spatial variability at a distance smaller than the shortest distance between observations, often termed measurement error.

3.3 SHapley Additive ExPlanations Explainable Artificial Intelligence

SHapley Additive exPlanations (SHAP) [50] is a method for explaining machine learning (ML) predictions by quantifying the contribution of each feature. It is based on the concept of Shapley values from cooperative game theory. The rationale behind SHAP is to compute the importance of each feature by evaluating its contribution to the prediction, thus providing explanations for the black-box model.

The shapely value of feature i is determined as the function defined in **Equation (3.8)**:

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S)) \quad (3.8)$$

Where S is the subset of the features, n is the number of features, and v is a characteristic function.

One of the main strengths of SHAP is its ability to provide not only global feature importance but also local explanations for individual instances. By interpreting specific instances, we can determine how each feature value affects the estimation. The SHAP explanation of each instance can be represented by a linear function, shown in **Equation (3.9)**:

$$g(z') = \varphi_0 + \sum_{j=1}^{M} \varphi_j z'_j \quad (3.9)$$

Where φ_0 is a constant, φ_j is the Shapely value of feature j, M is the number of features, $z' \in [0,1]^M$, when equals to 1 represents a feature is observed, otherwise is 0.

Overall, SHAP is a powerful tool for explaining complex ML models by quantifying the contribution of each feature to estimation and providing insights at both the global and local levels. In the context of this study, where interpretability and accuracy must be carefully balanced, SHAP provides a valuable means of exploration. By quantifying the contribution of individual features in the complex models employed for road friction estimation, SHAP enables to peer into the 'black box' of these models. Its application in this research is not only anticipated to offer precise insights into the workings of each model but also to contribute to the broader goal of fostering transparency and robustness in ML applications for transportation in general and winter road maintenance (WRM) in particular. This aligns with the study's commitment to understanding not only what the models are predicting, but also why they are making these predictions, thereby enhancing both the utility and credibility of the findings.

3.4 Transfer Learning

The two-stage TrAdaBoost.R2 [58, 69] algorithm is an extension of the Adaboost model that incorporates the concept of domain adaptation for transfer learning. Adaboost [70] assigns weights to the training dataset and iteratively generates new weak regressors while updating the weights. The weights are adjusted such that samples with larger residuals are assigned higher weights, making them more influential in subsequent iterations, while samples with smaller residuals have their weights decreased. TrAdaBoost, on the other hand, assigns different weights to the source training set and the target training set during the iterations. It increases the weights of target instances that are consistent with the source data and decreases the weights of target instances that are inconsistent. However, TrAdaBoost faces two challenges: the weights of source instances can sometimes be reduced to zero, and outlier instances are assigned high weights.

To address these challenges, a two-stage TrAdaBoost.R2 has been proposed. In the first stage, the weights of source instances are gradually adjusted downwards until they reach a certain threshold. This allows for a balance between the source and target data. In the second stage, the weights of all source instances are frozen while the weights of target instances are updated as normal in AdaBoost.R2. The weights are updated as **Equation (3.10)**.

$$w_{k+1,i} = \begin{cases} w_{k,i}\beta_k{}^{e_{k,i}}/Z_k, & for\ T_{source} \\ w_{k,i}/Z_k, & for\ T_{target} \end{cases} \quad (3.10)$$

Where Z_k is the normalizing constant, and β_k is designed such that the total weight of T_{target} is $\frac{m}{n+m} + \frac{k}{(S-1)}(1 - \frac{m}{n+m})$.

The two-stage TrAdaBoost.R2 ensures a more accurate and robust transfer learning process, allowing the model to adapt to new target domains while leveraging knowledge from the source domain.

3.5 Summary

This chapter provides an overview of the algorithms used in this thesis. The machine learning (ML) algorithm section introduces the principles behind the Regression Tree, Random Forest, Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) algorithms. By explaining how they function, we highlight these algorithms' differences and discuss their strengths and limitations. It is also noted that among the tree-structured algorithms, Random Forest and XGBoost are considered much more complex than the Regression Tree algorithm. The benefit of increasing model complexity is that it often leads to higher performance, but at the cost of decreased interpretability. In addition to ML methods, this section also introduced Ordinary Kriging (OK), which is a powerful method for data interpolation.

In an effort to understand the decision-making process behind the ML models, we then present the principles of explainable artificial intelligence (AI), specifically focusing on the game theory-based SHapley Additive exPlanations (SHAP) method. Lastly, we introduce the two-stage TrAdaBoost.R2 algorithm for transfer learning.

In the following two chapters, we will demonstrate the application of our proposed framework through case studies involving road friction estimation, followed by further improvement of our models. This will allow us to gain valuable insights into the practical implementation of the framework and contribute to enhanced accuracy in friction estimation.

4 MODELING AND CLASSIFYING ROAD FRICTION

In this chapter, a case study in Alberta is used to demonstrate how road friction can be estimated. The chapter is divided into four parts for a comprehensive explanation. The first section provides an overview of the study area and how the data was prepared. The second section focuses on training the model using mobile road weather information system (mRWIS) data. The third section applies the model to stationary road weather information system (sRWIS) data and incorporates geographic interpolation techniques for practical use. The final section classifies the estimated friction values based on risk level, making it easy to understand and interpret road risk information.

4.1 Study Area and Data Preparation

4.1.1 Study Area

The case study takes place in Alberta, Canada, a province located between 49° and 60° north latitude and 110° and 120° west longitude, covering a total area of 661,848 km^2. Alberta's northerly location makes it prone to Arctic weather systems, leading to extreme winter conditions. The movement of fronts between air masses leads to rapidly changing arctic air masses that can produce extreme temperatures as low as −54 °C.

To conduct the study, RWIS data was collected from multiple highways in Alberta province (depicted in **Figure 4-1(left)**). All available mobile road weather information system (mRWIS) data was carefully utilized for model training. However, due to differences in time and space between the collected mRWIS and stationary road weather information system (sRWIS) data, a specific segment of the highway was chosen for model application. This particular segment allows for the simultaneous collection of both mRWIS and sRWIS data, ensuring that the model can be trained and applied using consistent and comparable information.

This particular segment spans a distance of 127 km, running in a north-south direction from Valleyview to Whitecourt. Within this segment, there are four evenly deployed RWIS stations: AB_DOT_43-08, AB_DOT_43-10a, AB_DOT_3-06, and AB_DOT_43-14, as illustrated in **Figure 4-1(right)** below. This particular highway segment was selected as the study area for its high geographic variation (i.e., vegetation, terrain types, and altitude) and data distribution along the road segment.

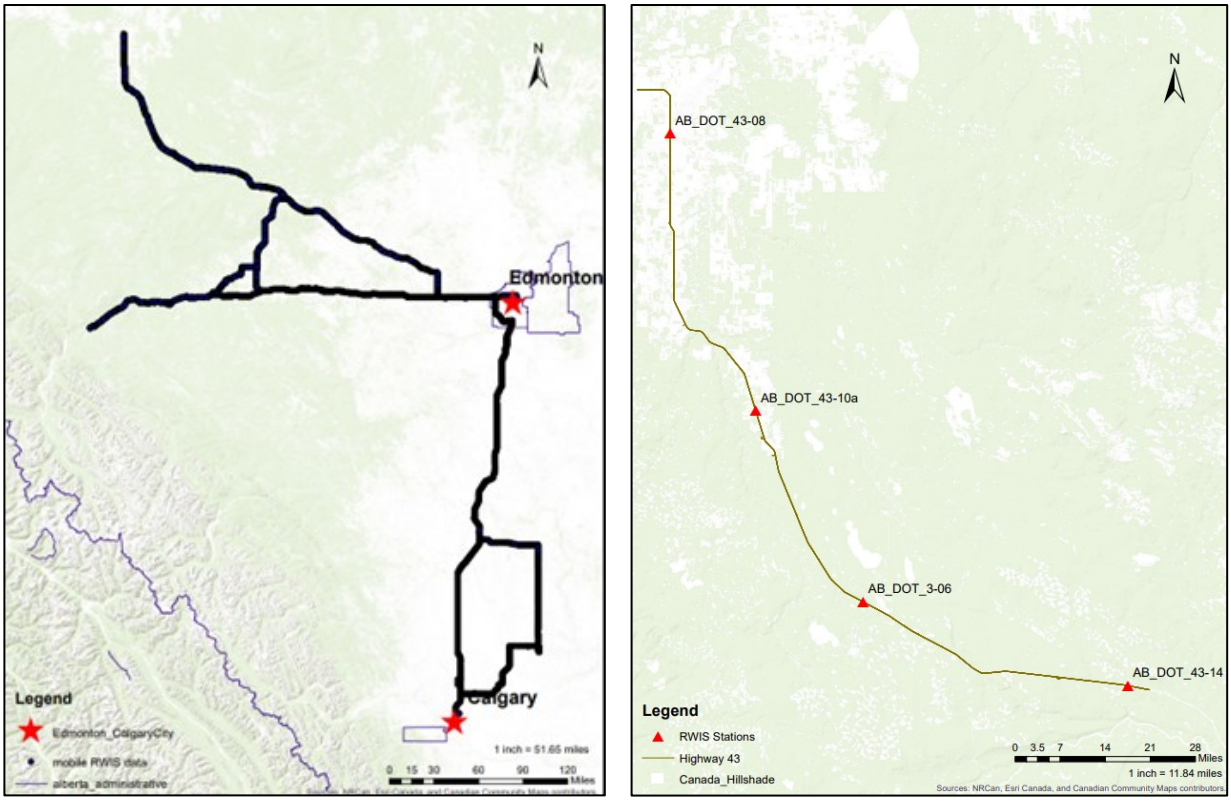


Figure 4-1 Highway Segments with Mobile RWIS Data Collection (left) and Selected Road Segment with Fixed RWIS Deployment for Model Application (right)

4.1.2 Data Preparation

The data used in this study was collected from both mobile and stationary road weather information systems (mRWIS and sRWIS). The mRWIS data was gathered using a non-invasive Vaisala Condition Patrol DSP 310 installed in the testing vehicle. This device captured various parameters such as air temperature, relative humidity, dewpoint, surface temperature, water depth, snow depth, ice depth, and friction at 3-second intervals, along with corresponding time and location tags. Road surface states (i.e., layer thickness and friction) in particular, were measured by the remote sensor DSC111 that uses the spectroscopic measuring principle to collect road condition information. Since it is costly and laborious to collect mRWIS data, the data used in this paper were collected in one direction only during inclement weather events that occurred between November 2014 and February 2015, with 93,101 data points in total.

On the other hand, sRWIS recorded its environmental data every 20 min at their respective locations. Environmental data collected includes air temperature, average wind speed, average

wind direction, dewpoint temperature, relative humidity, 1 h precipitation, 3 h precipitation, surface temperature, water depth, freeze point, and maximum and minimum temperatures as well as its station ID and data recording time. A total of 7344 sRWIS data points were collected and utilized in this study. Because sRWIS data are used for friction estimation, to compare with actual values, the data extracted from the sRWIS stations matched the date, time, and location of the mRWIS passing by; therefore, sRWIS data between 5 and 7 p.m. on 22 November 2014 were selected for model application.

The two RWIS data sources were used for separate purposes. The mRWIS data was used to develop and calibrate the friction estimation model by using the collected environmental data with the corresponding friction outcome. The sRWIS data was then employed to evaluate the model's applicability by providing weather information to spatial interpolate friction values. It is important to note that there are some subtle differences in the measurement values between mRWIS and sRWIS due to differences in measurement technique and sensor mounting locations. For example, sRWIS are slower in measuring changes in surface temperatures as their sensors are embedded into the ground, whereas mRWIS sensors are installed on the vehicle above the road surface. This difference results in a ±1 °C difference in recorded temperature values. Therefore, care should be taken to ensure data consistency when comparing mobile and stationary RWIS data. Additionally, only matching variables recorded by both RWIS formats were used as input model parameters. These variables are latitude, longitude, altitude, air temperature, relative humidity, dewpoint, surface temperature, and water depth. The friction value from mRWIS is used as the target variable in the modeling process.

4.2 Model Development with Mobile Road Weather Information Systems

Before calibrating the model, it is essential to select a predictor with a strong relationship to friction. Thus, a correlation matrix was generated with all the potential predictors to find the relationships between each variable pairing (see **Figure 4-2**).

	Latitude	*Longitude*	*Altitude*	*Air [C]*	*Humidity [%]*	*Dewpoint [C]*	*Surface [C]*	*Water [mm]*	*Snow [mm]*	*Ice [mm]*	*Friction*
Latitude	1.000										
Longitude	−0.784	1.000									
Altitude	−0.594	0.070	1.000								
Air [C]	0.153	−0.214	0.035	1.000							
Humidity [%]	0.524	−0.550	−0.124	0.280	1.000						
Dewpoint [C]	0.280	−0.336	−0.010	0.966	0.517	1.000					
Surface [C]	0.108	−0.174	0.041	0.928	0.342	0.919	1.000				
Water [mm]	0.015	−0.049	0.020	0.022	0.050	0.033	0.048	1.000			
Snow [mm]	0.080	−0.200	0.163	−0.030	0.230	0.030	−0.017	−0.029	1.000		
Ice [mm]	0.050	−0.029	−0.020	0.027	0.090	0.047	0.048	0.009	0.006	1.000	
Friction	−0.219	0.317	−0.079	0.185	−0.281	0.095	0.175	−0.028	−0.726	−0.172	1.000

Figure 4-2 Correlation Matrix of the Potential Predictors.

From the correlation matrix, snow depth has the strongest correlation with friction at −0.73, which intuitively makes sense as road surface slipperiness increases as snow layers get thicker. However, stationary road weather information system (sRWIS) in this study area does not collect snow depth values, so it cannot be implemented in the model. Longitude has the second highest correlation at 0.32, suggesting that the friction values generally increase with longitude. Relative humidity and latitude also have a strong but negative correlation with friction, meaning that the increase in these values lead to a reduction in friction, which is logically consistent. Air and surface temperature have a similar positive correlation with friction. Other than the factors previously mentioned, there are some factors with weak correlations less than 0.10: dewpoint temperature, altitude, and water depth. These highly correlated variables are then inputted in a stepwise manner based on the correlation matrix to calibrate the Regression Tree model. Since variables with low correlation (less than 0.1) contribute little to improving model accuracy, dewpoint temperature and water depth were omitted. Nevertheless, we kept altitude as it is a geographic predictor. To summarize,

the model has road friction as the target variable and longitude, latitude, altitude, air temperature, relative humidity, and surface temperature as predictor variables.

For model calibration, the mobile RWIS (mRWIS) data was randomly split into a training set with 74,480 data points and a testing set containing the remaining data. After finetuning, the following parameters were chosen: max depth of 26, minimum samples leaf of 2, and minimum samples split of 7. The model's accuracy reached 93.3%, suggesting that the model's performance was more than adequate. The validity of the model was then tested by making friction estimates using sRWIS measured predictor values and then comparing the predictions to the mRWIS measured friction values. This validation process requires using friction measurements close to the sRWIS station. In addition, the collected data needs to satisfy temporal consistency. Otherwise, the predictor variable values from sRWIS may differ significantly from the mRWIS due to potential spatial heterogeneity. **Figure 4-3** illustrates the friction estimates from the four sRWIS stations along with its actual measurement from the nearby mRWIS. As shown, the estimations are fairly close to the measured values, especially for station AB_DOT_3-06, where the estimation error is only 0.005 with station AB_DOT_43-08's estimation error being relatively higher at 0.14 error.

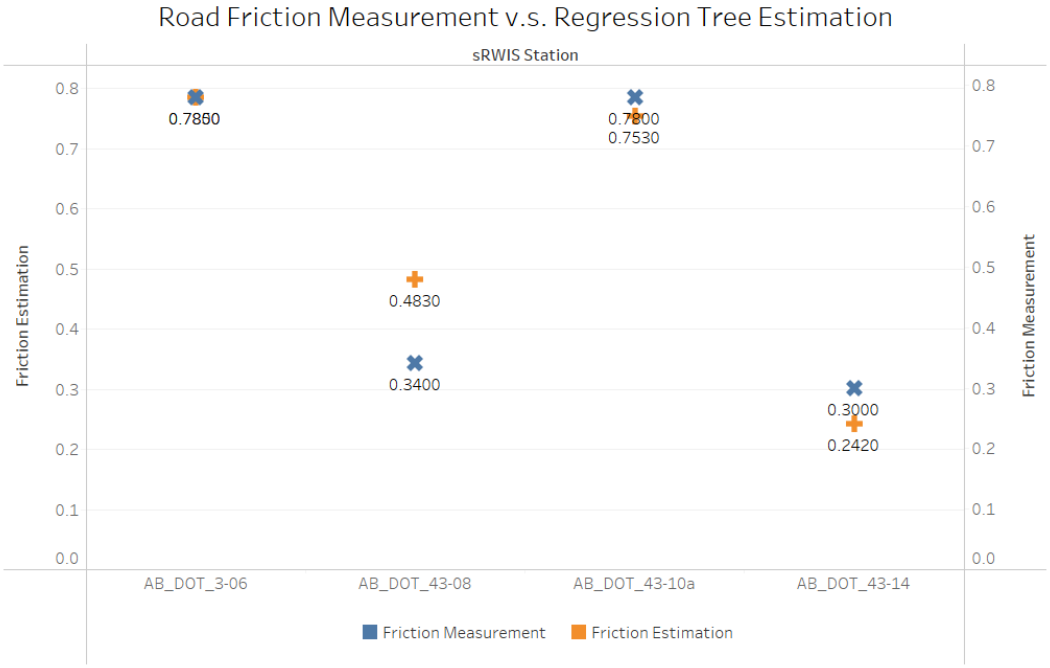


Figure 4-3 Comparison Between Road Friction Measurement and Model Estimation.

4.3 Model Application and Kriging Interpolation

4.3.1 Spatial Interpolation

This step is the generation of spatially continuous friction estimates to fill in the gaps between stations, providing greater details on the distribution of road surface friction values. In particular, the proposed method takes friction values measured at a stationary road weather information system (sRWIS) location and interpolate friction records between each pair of sRWIS using previously developed tree-based friction model. In this study, spatial interpolation using Ordinary Kriging (OK) was performed to estimate the weather information in-between sRWIS stations.

As mentioned earlier, the estimation results between sRWIS and mobile RWIS (mRWIS) are not identical. To properly evaluate the model performance and measure the estimation error of the model, the impact of the data inconsistency should be eliminated or at least minimized. This requires us to create a surrogate sRWIS by averaging the data collected from an mRWIS taken within a pre-defined buffer zone road surface (i.e., 400 m in our case upon reviewing the similarity of measured values) around an actual sRWIS. Doing so will remove the inherent measurement inconsistencies between the two RWIS types, allowing for more accurate and fair model validation. A total of 39 mobile data points were selected as testing points to evaluate the model and interpolation process. OK was then applied to air temperature (AT), surface temperature (ST), and relative humidity (RH) using Esri's ArcGIS [71, 72, 73], which can tune the semivariogram parameters automatically via its built-in model optimization function. The semivariogram parameters from the optimized models for each weather factor are listed below (see **Figure 4-4**).

The semivariogram values for each of the three environmental factors provide insight into their spatial autocorrelation. For AT, the nugget is zero, indicating that there is no inherent measurement error found, a rare but ideal outcome. The range and sill indicate that autocorrelation exists until 43 km and the maximum semivariance is 0.604. RH also has a small nugget that is approximately 0, a long range of almost 30 km, and a large sill of 12. As for the ST, it has a relatively large nugget value, a small range, and a small sill at 0.226, 240 m, and 0.433, respectively. The parameter values for ST make sense as they are significantly influenced by the local ground radiation, surrounding vegetation, solar exposure, and the amount of traffic driving over that location. The differences between the semivariograms are also intuitive as the atmospheric variables (air temperature and relative humidity) tend to be more stable over larger expanses than ground-based variables. The

constructed semivariogram model values were then used to estimate each weather factor along the highway segment. The interpolated estimates were then compared to the observed values from the mRWIS and plotted in **Figure 4-5** and their associated descriptive statics are shown in **Table 4-1**.

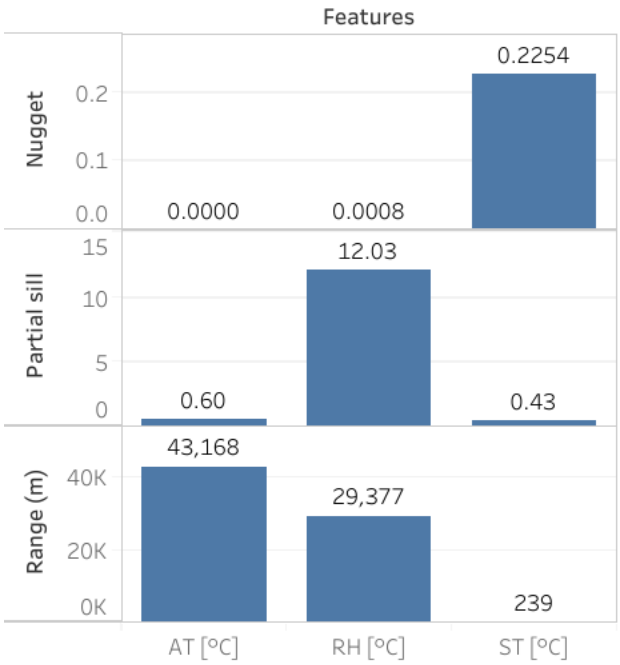
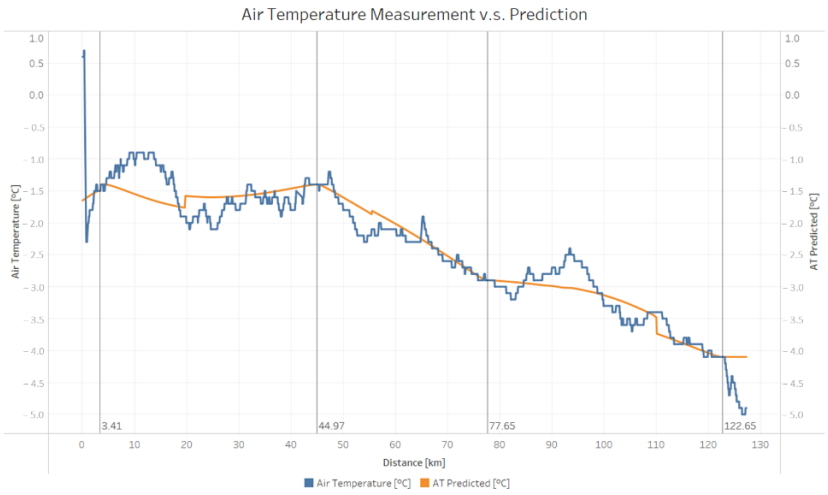


Figure 4-4 Optimized Semivariogram Parameters

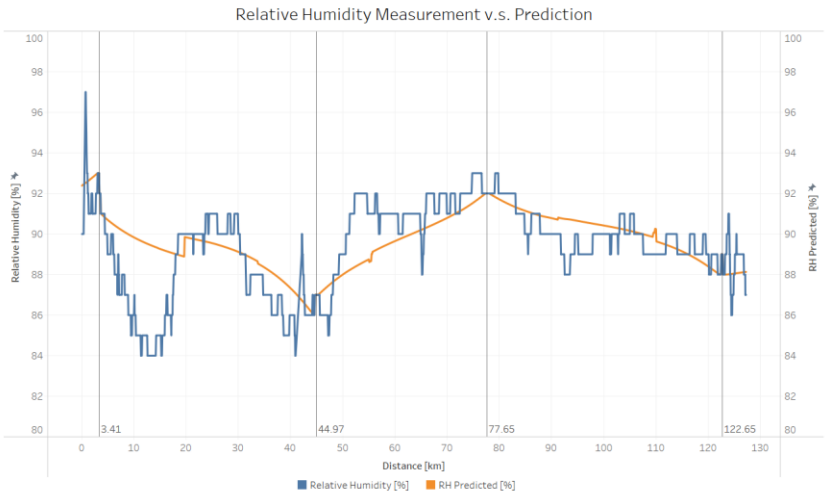
Table 4-1 Descriptive Statistics of Weather Parameters

	Air Temperature		Relative Humidity		Surface Temperature	
	Measured	**Predicted**	**Measured**	**Predicted**	**Measured**	**Predicted**
Mean	−2.49	−2.47	89.37	89.65	−3.45	−3.38
Std. Dev.	1.05	0.93	2.10	1.33	1.36	1.18
Min	−5.00	−4.10	84.00	86.00	−9.30	−7.57
Max	0.70	−1.40	97.00	93.00	−1.50	−2.20
Obs.	1588		1588		1588	

(a)



(b)



(c)

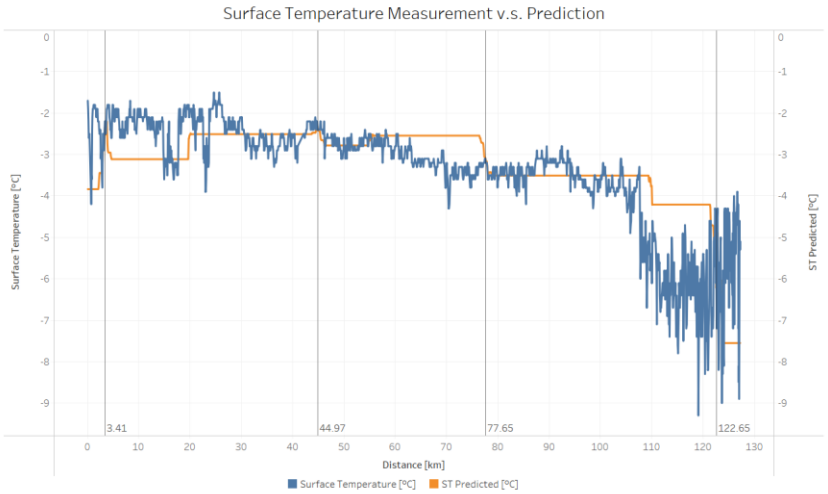


Figure 4-5 Interpolated (Orange) and Observed (Blue) Values of Weather Information. (a) Air Temperature; (b) Relative Humidity; (c) Surface Temperature

Based on the interpolated vs. observed plots shown above, it can be asserted that Ordinary Kriging (OK) was able to capture general variations over large distances (20 km) with some sections having less accurate predictions, possibly due to inherent local variations. An example of this limitation is seen between 3 km (where AD_DOT_43-14 locates) and 20 km. Actual AT and ST values fluctuate drastically with peaks, and RH trends opposite, while their predicted values follow a similar trend but are much smoother. This is because kriging produces smooth estimates by taking an average to offset the variation between high and low values.

4.3.2 Road Friction Estimation with Stationary Road Weather Information System

To estimate the road friction throughout the area, the spatially interpolated weather factors and geographic features generated in the previous step were inputted into the Regression Tree model to produce friction estimates for the entire study area. Estimations were made at 20 m resolution. However even at high accuracy the model can hardly predict exact the same value as measured, so the estimates were aggregated every 1.0 km intervals on the road to reduce variance and make the results more realistic when plotted. A comparison plot between the estimated and actual friction is shown in **Figure 4-6**.

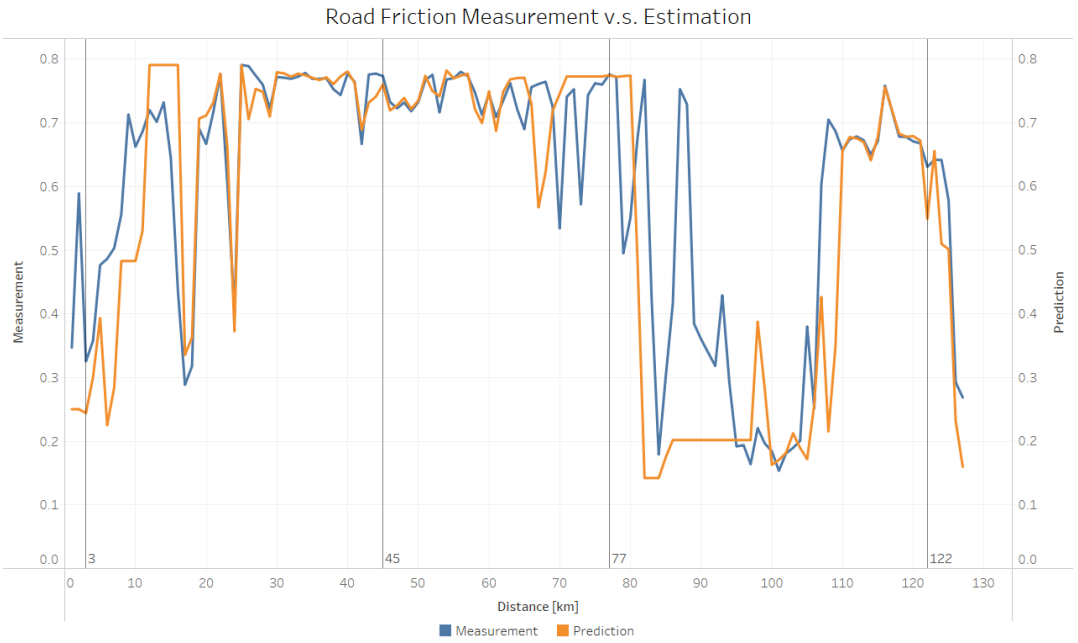


Figure 4-6 Interpolated (Orange) and Observed (Blue) Values of Friction Over the Study Area

Figure 4-6 suggests that the estimated and measured friction curves closely resemble one another, especially at the 16 km to 61 km and the 111 km to 121 km segments. However, there are also regions with moderate differences, such as the 0 km to 11 km stretch. Minor discrepancies like this are likely because the input variables themselves are not measured values but estimates with error. The predicted air and surface temperature were lower than the actual measurements, while the estimated relative humidity was higher than the measured values. Since friction is positively correlated with temperature and negatively correlated with humidity, the estimated friction is lower than the actual value. The differences seen at 70 km to 76 km can be similarly explained except in the opposite direction.

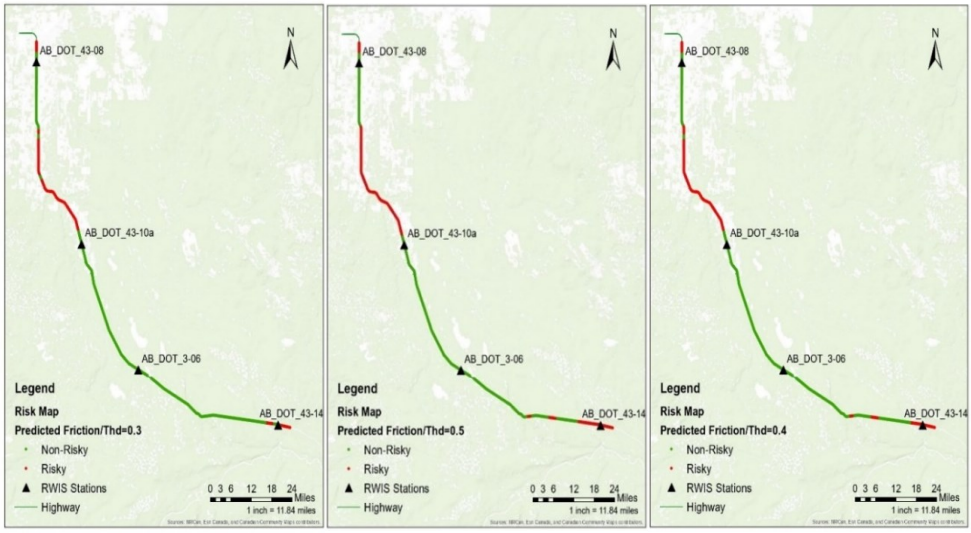
While most discrepancies can be explained as a product of interpolation error propagating from the previous step, for locations between 85 km and 94 km, the difference is quite significant and cannot be attributed to predictor interpolation error. Deeper scrutiny revealed that the measured snow depth drastically changes within this region, from 1.2 mm to 0 mm. This 1.2 mm difference represents a change from being snow-covered to bare pavement due to snow depth information being not accounted for in the modeling process as stationary road weather information system (sRWIS) does not collect snow depth data. The same logic can be applied to 89 km, where the snow depth increased by 1 mm, resulting in a sharp friction value drop. Likewise, at the 91 km to 94 km section, where snow depth fluctuates between 0.4 mm and 1 mm, a similar friction fluctuation between 0.2 and 0.6 was observed. Based on these observations, snow depth accumulation may be a strong indicator of friction values. Nonetheless, the ordinary Kriging (OK) interpolation model developed has proven to capture the general road surface conditions (RSCs) patterns with help of just four sRWIS data points—the first of its kind in existing literature.

4.4 Road Risk Identification

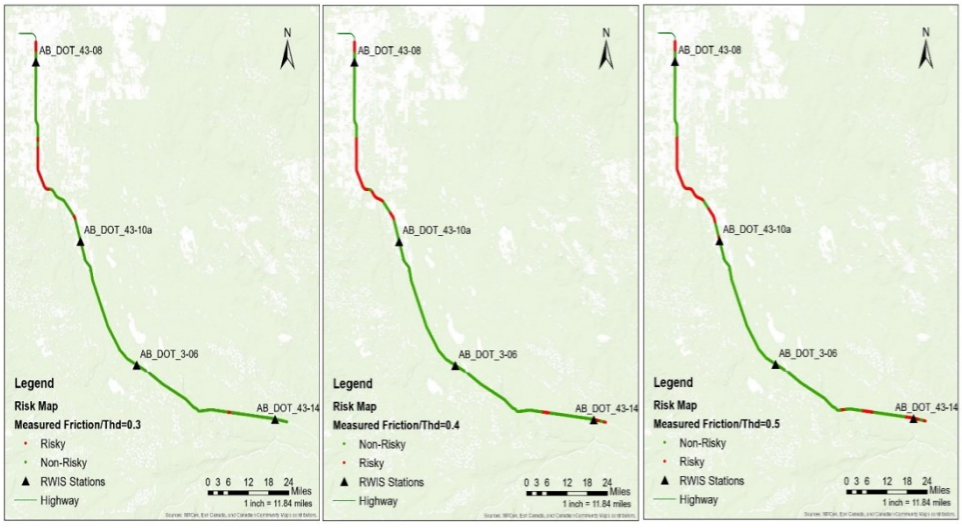
4.4.1 Binary System

Providing only friction values has little meaning to the general public as it is challenging for them to grasp the safety implications between the friction values. For this reason, we simplify our friction outputs by classifying them as either risky or non-risky for the purpose of creating a color-coded risk map. Previous studies have suggested several cutoff values for when friction values are considered unsafe, which can range from 0.3 to 0.5 [74, 75, 76, 77]. In this study, we consider

three cutoffs within this interval: 0.3, 0.4, and 0.5 to categorize respective driving risks. The performance of the classification was then evaluated based on the confusion matrix with accuracy selected as the metric of focus. This choice was made because we were primarily concerned with global sample prediction, specifically the proportion of samples correctly classified. The results show that at $Threshold = 0.3, 0.4, and\ 0.5$, the accuracies for road risk estimation are $84.25\%, 89.76\%, and\ 88.98\%$, respectively. Regardless of the selected threshold, the accuracy of road risk estimation is near 90%; hence, our developed model is highly accurate.



(a)



(b)

Figure 4-7 Road Risk Maps at Threshold = 0.3, 0.4, and 0.5. (a) Risk Maps Based on Predicted Friction; (b) Risk Maps Based on Measured Friction

4.4.2 Three-Category System

The benefit of using a binary classification system is that it is easy to understand, but this benefit comes at the cost of reducing the amount of information presented. Although general road users may appreciate its simplicity, maintenance personnel may desire more detailed information. Hence, we increased the risk categories to three to give a more comprehensive view of Road Surface Condition (RSC) along the road network. The adoption of the three-category system is not only significant for providing a detailed understanding suitable for various stakeholders but also aligns with standards that are generally accepted and utilized by many transportation agencies including Alberta Transportation [78, 79, 80]. The three-category system is as follows: friction below 0.3 is high risk, medium risk is between 0.3 and 0.5, and low risk is above 0.5—intervals that have been generally accepted and thus used in existing literature.

Figure 4-8 illustrates the risk warning map of measured and predicted friction under this classification system, with red, amber, and green representing high, medium, and low risk, respectively. The accuracy based on this classification is 77.95%, which was lower than the binary system, but the model still had acceptable performance. Most of the prediction errors (18.90%) are False Risky (FR), where non-risky road sections are classified as risky, implying that our model is conservative in its predictions, which does not compromise safety.

The color-coded map provides a risk level along the select highway segments, which can be highly informative in practical applications. For instance, by providing real-time road condition estimations via online platforms (e.g., 511), trip makers will be able to adjust their travel plans and divert accordingly to avoid potential accidents and traffic delays. Likewise, this tool, once refined and tested with more data, can also be integrated into vehicle to everything (V2X) technologies [21, 81, 82] to provide real-time RSCs information for improved safety before, during, and after inclement weather events. Such information can also be utilized by winter road maintenance personnel as the maps would provide information pertaining to dangerous road sections so that maintenance vehicles could be dispatched in a more efficient and timely manner to perform location-specific and targeted maintenance operations to ensure the safety and mobility of the winter travelling public.

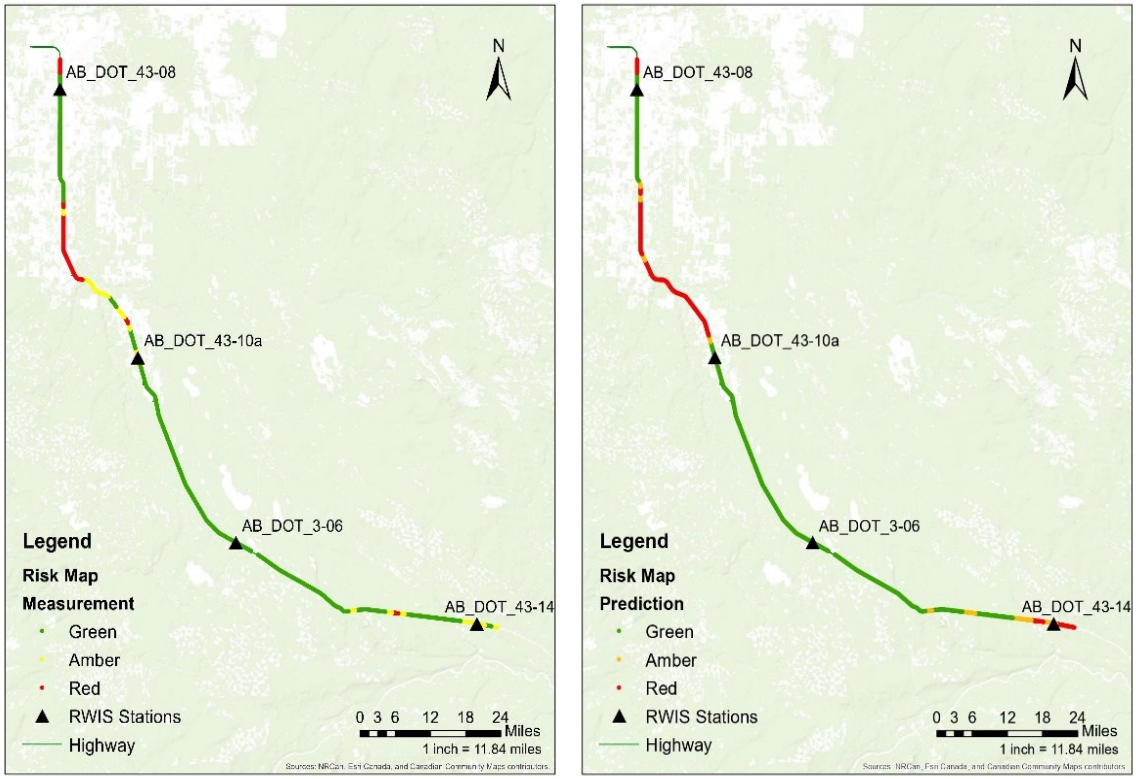


Figure 4-8 Road Risk Warning Map with Measured (Left) and Predicted Friction (Right)

5 ENHANCING MODEL ACCURACY, EXPLAINABILITY AND TRANSFERABILITY

The previous chapter provided a detailed overview of how winter road friction coefficients can be estimated. In this chapter, we will further explore the model developed by comparing its performance with more complex algorithms in terms of model accuracy and intuitiveness, as well as evaluate the best-performing model in a new dataset to explore its transferability. This chapter is divided into four sections. The first section provides a description of the data utilized. In the second section, we compare the performance of the decision model developed in Chapter 4 with more complex models, including Random Forest, Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR). The third section involves using SHapley Additive exPlanations (SHAP) to evaluate the decision-making process behind each model, which serves as a crucial basis for selecting the best model. In the last section, we apply transfer learning

techniques to the best-performing model (determined based on accuracy and intuitiveness) to enable it to adapt to an entirely new dataset.

5.1 Data Processing

The study presented in this chapter focused on using mobile road weather information system (mRWIS) data. As mentioned earlier, the amount of data is enormous due to the high-frequency data acquisition of mRWIS every 3 seconds. However, the SHapley Additive exPlanations (SHAP) model is computationally inefficient. Generating model interpretations using SHAP becomes computationally expensive and time-consuming when dealing with such large datasets. To address this, an additional preprocessing step is implemented to aggregate the data at a one-minute interval, reducing computational complexity. This additional step also helps to handle outliers. In order to maintain consistency with the models discussed in the previous chapter, only weather and geographic features that are deemed important for predicting friction are considered as model inputs, that is, air temperature (AT), relative humidity (RH), surface temperature (ST), latitude, longitude, and altitude.

We utilized two sets of mRWIS data for our analysis. The first set, introduced in Chapter 4, was used for model development. It spanned 21 days from 2014 to 2016 and comprised 4,704 samples. For training and testing the models, we used 70% and 30% of the samples, respectively. The second dataset was specifically collected for transfer learning purposes. It covered a 10-day period between January and March of 2023 on Highway 2, between Edmonton and Calgary (as depicted in **Figure 5-1**) and contained a total of 980 samples. Descriptive statistics for both datasets are provided in **Table 5-1**. By leveraging these datasets, we were able to successfully establish, evaluate, and optimize the road friction estimation models.

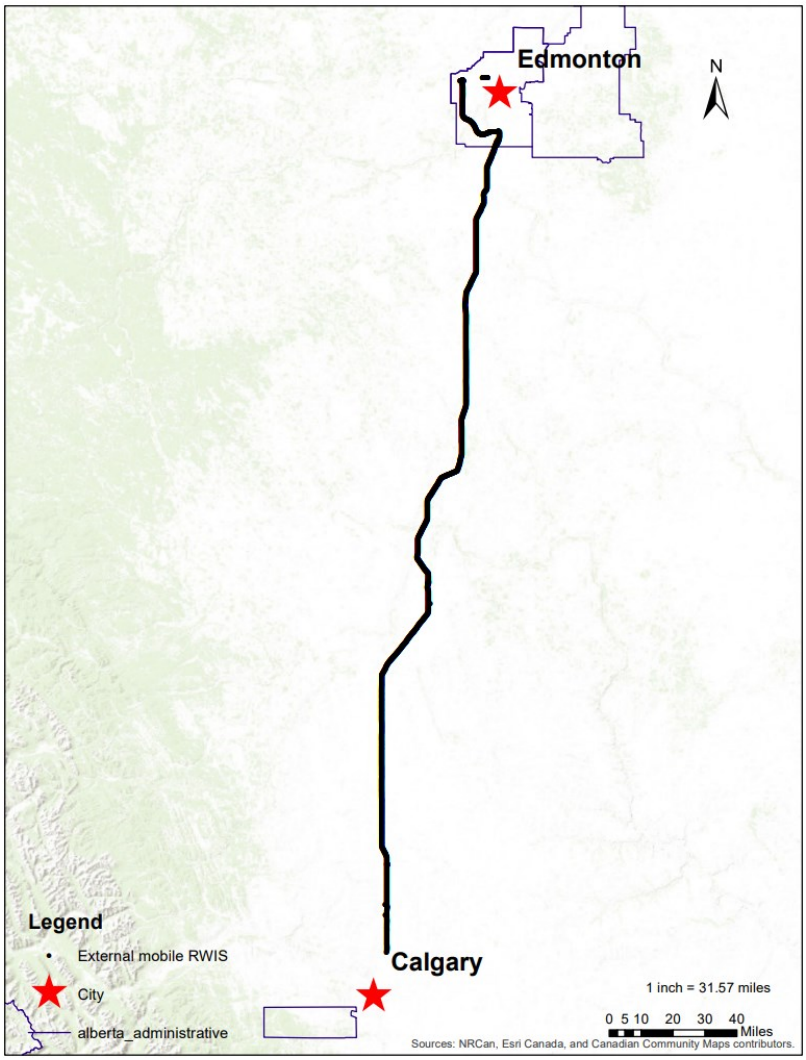


Figure 5-1 Road Segments with External mRWIS Data Collection

Table 5-1 Descriptive Statistics

		Mean	Std. Dev.	Min.	Max.	Obs.
Calibration Dataset	AT [C]	-4.526	6.293	-28.440	7.210	4704
	RH [%]	84.027	10.706	40.000	101.000	
	ST [C]	-3.732	4.853	-21.045	5.160	
	Latitude	53.643	0.957	50.979	55.237	
	Longitude	-115.507	1.388	-118.848	-113.233	
	Altitude	825.927	112.033	490.150	1161.700	
	Grip	0.675	0.219	0.119	0.820	
External Dataset	AT [C]	-5.656	6.287	-20.411	8.888	980
	RH [%]	74.067	16.517	28.156	100.000	
	ST [C]	-1.985	3.520	-11.923	9.374	
	Latitude	53.092	0.653	51.152	53.647	
	Longitude	-113.619	0.169	-114.157	-113.343	
	Altitude	754.957	124.051	620.653	1172.085	
	Grip	0.743	0.149	0.270	0.820	

5.2 Comparative Analysis Among Machine Learning Algorithms

5.2.1 Model Structure

During training, 3756 samples from the model development dataset were used to train the models, while 940 samples were reserved for validating model performance. To optimize the performance of the developed models, a grid search method was used to determine the optimal hyperparameters for each model. After fine-tuning, the hyperparameters that resulted in the lowest error were selected. The following are the optimal parameters used for model development: the Regression Tree model had a maximum depth of 14, a minimum sample leaf of 1, and a minimum sample split of 7. The Random Forest model had a maximum depth of 18, a minimum sample leaf of 1, a minimum sample split of 2, and 122 estimators. The Extreme Gradient Boosting (XGBoost) model had a gamma of 0, a learning rate of 0.1, a maximum depth of 7, 1000 estimators, and a subsample ratio of 0.7. Lastly, the Support Vector Regression (SVR) model had a cost of 10 and a gamma of 100.

5.2.2 Model Evaluation

Evaluation of the developed models using the test set showed performance scores of 84.17%, 89.99%, 91.39%, and 85.83% for Regression Trees, Random Forest, Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR), respectively based on r squared. Overall, all machine learning (ML) models exhibited good performance in accurately estimating road friction. Among them, the Regression Tree model obtained the lowest score due to its relatively simple structure, while the XGBoost model with the most complex structure achieved the highest performance, which is slightly higher than that of the Random Forest model. It is worth noting that the SVR model performed rather poorly, scoring only slightly higher than the Regression Tree model. These findings are in line with the model accuracy and interpretability rankings proposed by Herm [20].

To further compare and evaluate these four models, mean absolute error (MAE) and root mean square error (RMSE) were calculated for further comparison; the results obtained are shown in **Figure 5-2**. From **Figure 5-2**, we observe that Regression Tree and SVR models have similar estimation errors, while Random Forest and XGBoost models have significantly lower errors. This indicates that the friction estimation error decreases with increasing model complexity, which is in line with what we found previously. In terms of the SVR model, it was the second worth performer according to RMSE and the worst when using MAE.

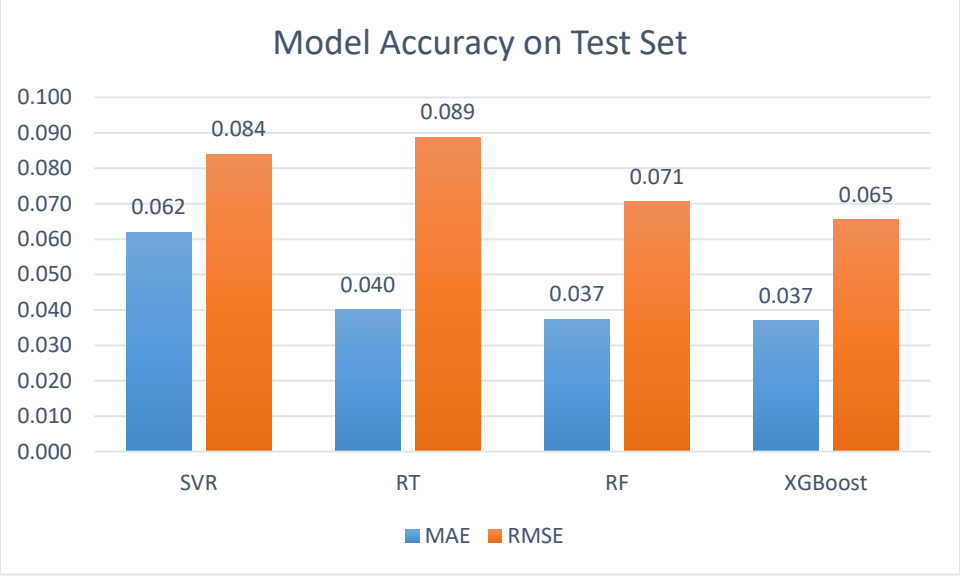


Figure 5-2 Model Accuracy on Test Set

To effectively prioritize winter road maintenance operations, we implemented a road friction coefficient classification system based on the relative risk level of potential traffic accidents. We employed a three-category system and a binary system [83], where the three-category system provides detailed insights by classifying the coefficient of friction into high, medium, and low risk levels using thresholds of 0.3 and 0.5. Friction values below 0.3 are marked as high risk, those between 0.3 and 0.5 are marked as medium risk, and values above 0.5 are marked as low risk. Additionally, we introduce a binary system with a threshold of 0.5 to quickly assess road risk by indicating whether it exists or not.

Next, we generated a confusion matrix to assess the accuracy of the classification of the friction estimates. The accuracy is shown in **Figure 5-3**. As illustrated, the performance of the four ML models is quite similar. Among them, XGBoost stands out as the best performer, demonstrating its excellent ability to estimate road friction. This is followed by Random Forest model, then SVR and Regression Tree with slightly worser performance.

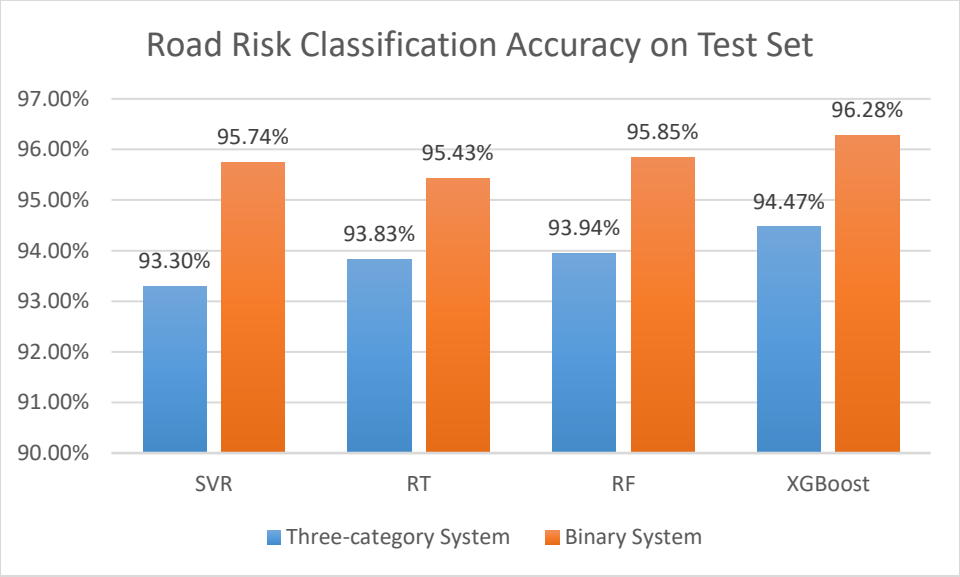


Figure 5-3 Road Risk Classification Accuracy

After evaluating the four models using metrics such as r squared, MAE, and RMSE, we have determined that XGBoost performed the best, followed by Random Forest, while SVR and Regression Tree exhibited poorer performance. However, except for the Regression Tree which is concise and easy to interpret, the XGBoost, Random Forest, and even SVR models have poor interpretability due to their complex mechanisms. This poses a challenge when assessing whether

their predictions align with logical reasoning. Therefore, in the next section, we will leverage SHAP to explore the internal workings of the models and gain further insights.

5.3 SHapley Additive ExPlanations Explanation

The use of the SHapley Additive exPlanations (SHAP) model serves two purposes in this thesis. First, it allows us to generate simple explanations for complex models, thereby increasing model interpretability. Second, by examining the contribution of each input feature via SHAP, the logical consistency and the correctness of the model structure can be assessed. For these reasons, we applied SHAP to the models developed previously in this study to evaluate their intuitiveness.

5.3.1 Global Explanation

The summary plot generated by SHapley Additive exPlanations (SHAP) visually represents the relationship between SHAP values and their corresponding features. For the analysis of this plot, the ranking of the features on the left side of the plot indicates their absolute importance with respect to the model. The color of the dots illustrates the magnitude of the feature values, with red representing high feature values and blue representing low feature values, while their distribution underlines the correlation between feature values and SHAP values. Red dots on the right and blue dots on the left indicate positive correlations. Conversely, red dots on the left and blue dots on the right indicate a negative correlation.

Figure 5-4 presents summary plots of the four models separately. By examining the diagram, it's clear that all three tree-structured models function utilize the input features in a similar manner. In contrast, the Support Vector Regression (SVR) model is drastically different, where the relationship between the input features and friction is close to the opposite of the three tree-structured models. In terms of feature importance ranking, the tree-structured models have identical order, i.e., longitude being the most important, followed by air temperature, relative humidity, latitude, surface temperature, and altitude; while the order of feature importance in the SVR model is longitude, latitude, air temperature, altitude, surface temperature, and relative humidity in the order of decreasing importance.

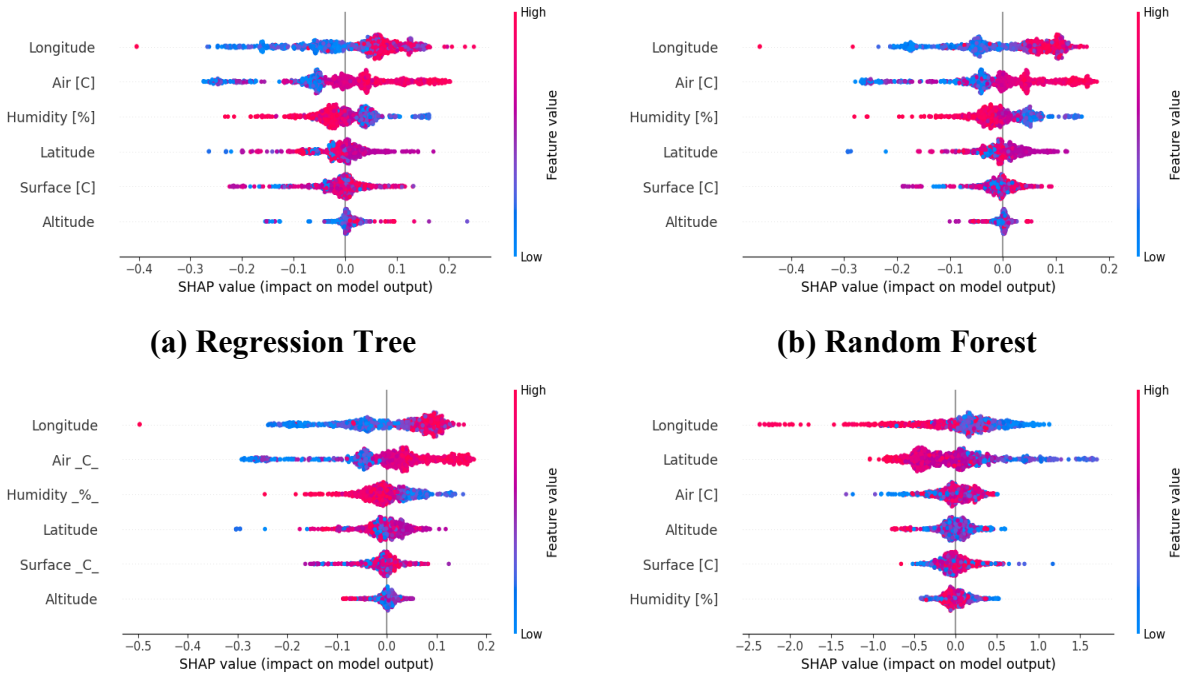


Figure 5-4 Summary Plot of SHAP Model on (a) Regression Tree; (b) Random Forest; (c) XGBoost; and (d) SVR

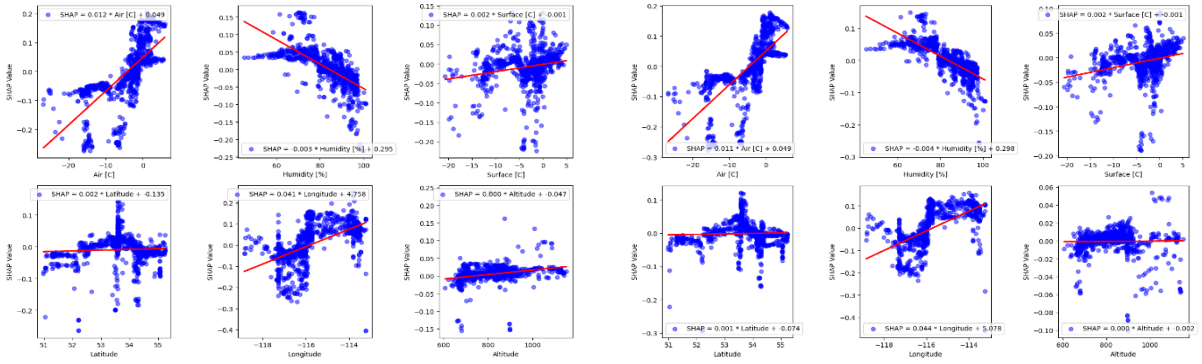
In addition to determining the order of feature importance for each model, the summary plot presented above also allows us to examine the model's assumed relationship between the predictor features and friction. An example of how this analysis can be done is presented below using the Extreme Gradient Boosting (XGBoost) model. The other two tree-structured models should operate in the same manner, given how similar their summary plots are.

- The longitude is positively correlated with SHAP values, namely the higher the longitude, the higher the estimated friction coefficient, implying that within Alberta, the low longitude area, i.e., the west, is more prone to friction loss. This can be explained by the fact that cities, primarily located in eastern Alberta, are more able to conduct timely winter road maintenance operations to prevent friction losses or restore road conditions to safe levels, whereas the sparsely populated western regions lack this capability.
- There is a positive correlation between air temperature and SHAP values; an increase in air temperature leads to an increase in friction coefficient. This is reasonable as rising

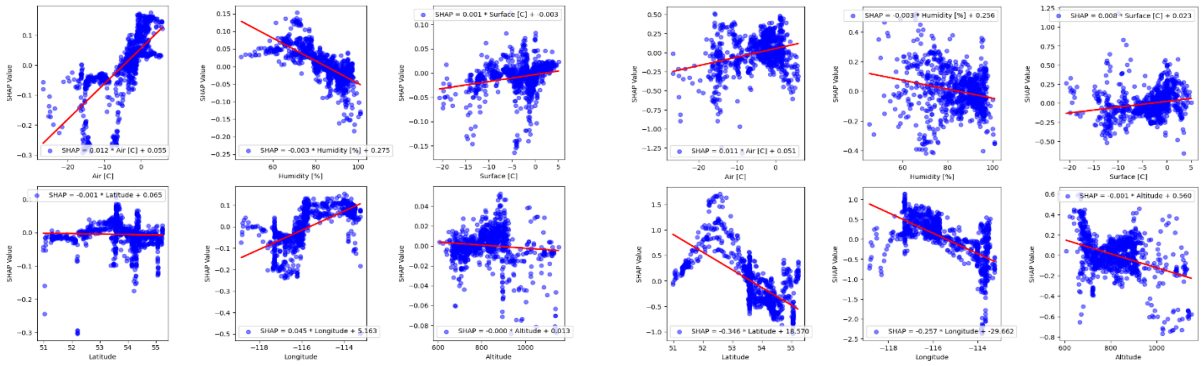
temperatures will bring more heat, accelerating the melting of ice and snow on the road or drying out the water on the road, thereby increasing road friction.

- Relative humidity is negatively correlated with friction estimates, meaning that high humidity causes low friction. Generally, high humidity causes the road surface to be moist, which reduces the friction coefficient. Especially in cold winter conditions, cold weather can cause moist road surfaces to freeze, resulting in a sudden drop in friction coefficient. Hence, this negative correlation makes sense.

Unlike the three features mentioned above, the distribution of latitude, surface temperature, and altitude points are clumped around 0 (SHAP value), indicating that they contribute little to the friction estimates. Further, as there is no distinct separation between blue and red dots, it's challenging to clearly understand the correlation between these features and friction. To solve this issue, we plotted the distribution of each feature value and its corresponding SHAP value and used linear regression to fit a trend line to further investigate this relationship (**Figure 5-5**).



(a) Regression Tree **(b) Random Forest**



(c) XGBoost **(d) SVR**

Figure 5-5 Best Fit of Features and SHAP Values in four models

The trend line in the plot shows a slight upward trend for surface temperature, suggesting that higher surface temperatures contribute to higher friction estimates, which is reasonable as surface temperature should have a similar effect as air temperature. In comparison, the trend line for altitude and latitude has a negative slope, indicating that friction estimates decrease with increasing altitude or latitude. This observation is intuitive as temperature decreases as these features increase, i.e., the decrease in temperature caused by these parameters results in lower friction. That being said, the flatness of the line indicates that their change has little effect on road friction. To further understand the reason why latitude and altitude contribute very little to friction predictions, we generated interaction plots to investigate the joint effect of these two features on the friction estimates.

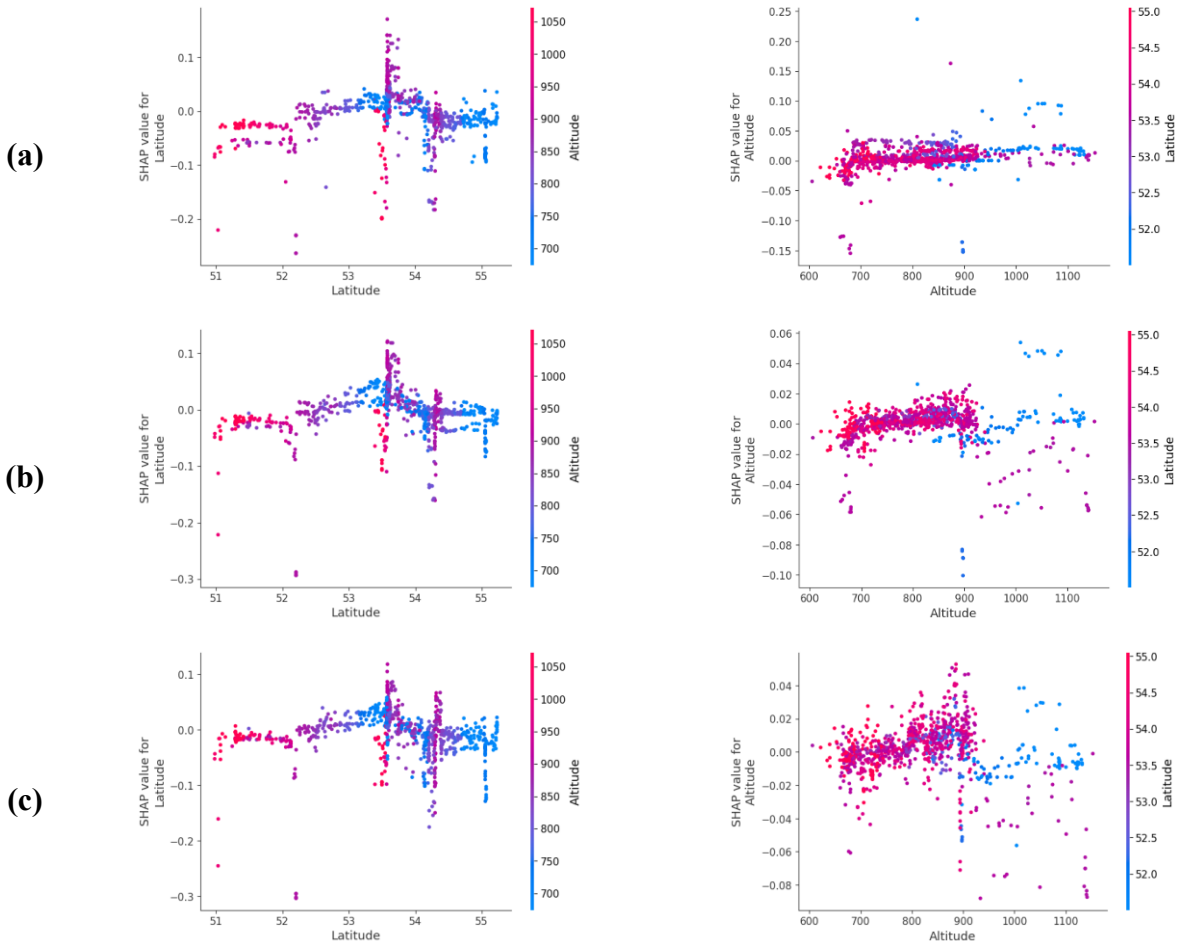


Figure 5-6 SHAP interactive plot of (a) Regression Tree, (b) Random Forest and (c) XGBoost with SHAP value for latitude (left) and SHAP value for altitude (right) on the y-axis

Figure 5-6 showcases the interaction plot of the three tree-structed models. Through this process, we found that most of the data were collected in low latitude and high-altitude regions, or high latitude and low altitude regions; hence the collected data doesn't have enough variation in these two predictors for the models to understand how they are related to friction.

Moving on to the SVR model, the feature with the largest contribution here is also longitude. However, unlike the tree-structured models, it's revealed that longitude is negatively correlated with friction. The second most important feature, latitude, shows a negative correlation, which is logical. However, the contribution of the remaining features is ambiguous as illustrated in the summary plot. This implies that SVR models may not be able to effectively use these features to produce accurate friction estimates. Therefore, such ambiguity in the feature contributions may be one of the reasons why SVR models perform poorly on the testing dataset. Similar to the tree-structured models, we further examined the trendlines for each feature and found that the contributions of these features are logically reasonable, indicating that these features do have some minor influence on the friction estimates. However, the distribution of the scatterplots also shows a more random distribution compared to the tree models, especially for air temperature and relative humidity. Therefore, simply using linear regression for fitting is invalid, which may be related to the principle of SVR mapping data to high-dimensional space.

In the global analysis, we analyzed the contribution of each feature to friction estimates for each of the four developed models, among which XGBoost was found to be the most robust model due to its intuitiveness and high performance.

5.3.2 Local Explanation

To ensure the reliability of the model, we expect the contribution direction of features to friction estimation in individual instances to match the actual situation. In particular, we expect features of air temperature, surface temperature, and longitude to positively correlate with road friction, while relative humidity, latitude, and altitude negatively correlate with road friction. Therefore, we generate SHapley Additive exPlanations (SHAP) local explanations to showcase the influence of each feature on the friction estimation in specific instances by using force plots. To avoid the randomness of instance selection and gain a more comprehensive understanding of the inner

working of the model, we randomly choose two instances from each road risk category for detailed analysis.

A force plot shows the effect feature value changes have on the friction value output. The bold value on the plot indicates the friction value estimated by the SHAP model, with the important features represented by the color red and blue, red indicating an increase in the estimate, and blue indicating the opposite. Regarding the size of the bars, they indicate the degree of contribution of a specific feature, with larger bars representing greater contribution.

By comparing feature values of specific instance with the mean feature values described in **Table 5-1**, we can determine whether a feature value is high or low and assess whether the contribution direction of each feature aligns logically with the friction estimate. The specific instances are discussed in detail below.

Figure 5-7 Force Plot of Instance 1 (High Risk)

A total of six instances were examined, where two instances per risk level were examined. Instance 1 (**Figure 5-7**), classified as high risk, has an observed friction coefficient of 0.1265, with the following feature values: air temperature of -15.8°C, relative humidity of 84.35%, surface temperature of -13.77°C, latitude of 55.95, longitude of -117.3, and altitude of 732.9 meters. Upon comparing the feature values with the overall dataset, several observations can be made. The air temperature and surface temperature in the selected instances are notably lower than the average

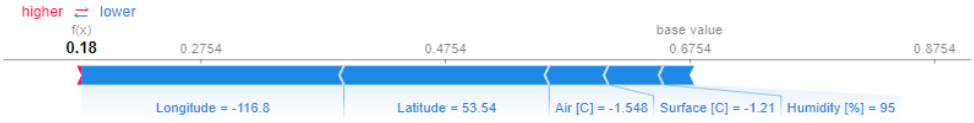
values of -4.53 and -3.73. Both the altitude and longitude are below the average of 825.93 and -115.51, and the latitude is slightly higher than the mean latitude of 53.64. As for relative humidity, it remains relatively consistent with the mean value of 84.03%, showing only a slight increase of approximately 0.3%. The following are the SHAP explanations for estimates generated by Regression Tree, Random Forest, Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) models.

In the Regression Tree, all features except relative humidity and altitude have a significant effect on the reduction in friction estimate. Specifically, they led to a decrease in road friction due to lower values of air temperature, longitude, and ground temperature, and the higher values of latitude. The contribution directions of these features are consistent with their correlation with the road friction. Likewise, the contribution direction of each significant influencing feature in Random Forest and XGBoost is the same as the Regression Tree. The difference in the Random Forest is that the additional feature of relative humidity also has a significant contribution, with high values leading to a decrease in estimated friction, consistent with its negative correlation with friction.

In the explanation of the SVR model, each feature makes a significant contribution to the estimation of friction. However, under the influence of these features, the generated friction estimate of 0.34 deviates greatly from the true value of 0.12, indicating a large error. In addition, not every feature contributes as logically expected. Specifically, longitude, air temperature, and altitude had opposite effects on the friction estimates. In the model, the low values of longitude and air temperature lead to an increased friction estimate, while a low altitude value leads to a decreased friction estimate. These unreasonable feature contributions are a reflection of poor model performance.

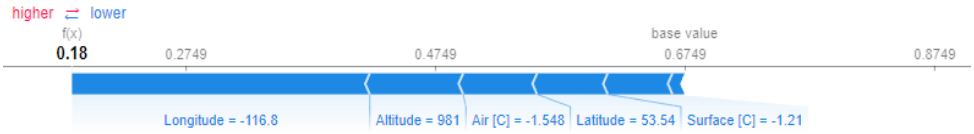
Instance 2 (**Figure 5-8**) also classified as high risk, but the significance of its characteristic contribution differs from that of Instance 1. The observed friction coefficient is 0.1333, with the following feature values: air temperature of -11.548°C, surface temperature of -1.21°C, relative humidity of 94%, latitude of 53.54, longitude of -116.8, and altitude of 981 meters. In this instance, air temperature, latitude and longitude are below their respective means, and relative humidity and elevation are above the mean. Surface temperature is higher than the mean, yet below 0 degrees.

(a) Regression Tree



(b) Random Forest

(c) XGBoost



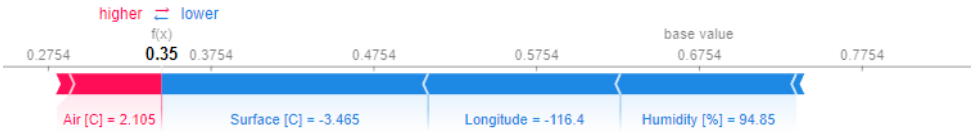
(d) SVR

Figure 5-8 Force Plot of Instance 2 (High Risk)

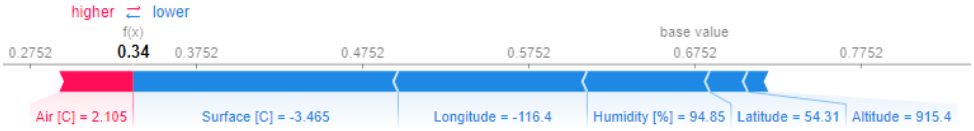
In the Regression Tree, all features play an important role in friction estimation, except for altitude. This implies that the contributions from the other features are sufficient to bring the friction estimate close to the true value, while the contribution of the altitude is minimal, consistent with the overall feature importance displayed in the feature graph. The remaining features that have significant contributions largely align with our expectations, as higher altitudes are negatively correlated with friction, while lower air temperature and longitude are positively correlated with friction, leading to lower friction values. Despite the surface temperature being above the average (-3.73), it is still reasonable to expect lower friction since it remains below 0 degrees. However, the low latitude value results in a decrease in friction, which contradicts the expected relationship. Nevertheless, considering that the latitude value is close to the dataset's mean of 53.64 and the observed instability of latitude's contribution in the global interpretation, we can attribute this to an acceptable error. Similarly, in the Random Forest model, each feature has a significant effect on the friction estimate with logical direction. Additionally, in this instance, high altitude has a logical effect on the reduced friction estimate. When using XGBoost to estimate friction, all features have a significant effect except for the minimal contribution of relative humidity, and the contribution directions of all features align with the analysis conducted for the Regression Tree and Random Forest. Notably, the contribution of latitude decreases in more complex models, corroborating that increasing model complexity improves accuracy. Moreover, the features of the SVR model that have a significant impact on friction estimation are consistent with XGBoost, but

the direction of contribution is different. That is, the low value of surface temperature and longitude lead to an increase in the friction estimate, which contradicts the actual logic.

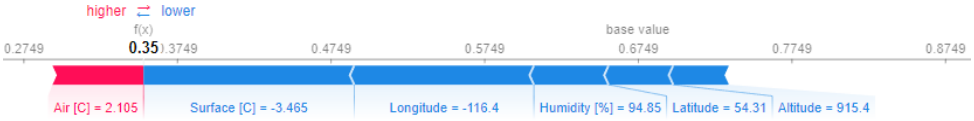
(a) Regression Tree



(b) Random Forest



(c) XGBoost



(d) SVR

Figure 5-9 Force Plot of Instance 3 (Med Risk)

Instance 3 (**Figure 5-9**) is classified as medium risk, with a true friction coefficient of 0.4115. The corresponding feature values are air temperature of 2.105°C, surface temperature of -3.465°C, relative humidity of 94.85%, latitude of 54.31, longitude of -116.4, and altitude of 915.4 meters. Of these, longitude is below the mean and the rest are above their respective means.

In the Regression Tree model, surface temperature, longitude, relative humidity, and air temperature are the features that contribute most to the friction estimates. High air temperature increases friction estimate, while lower longitude and high relative humidity lead to lower friction estimate, which agrees with our expectations. Although the surface temperature is slightly above the mean value of -3.73, as it is still below 0 degrees, its role in reducing friction is also reasonable. In Random Forest and XGBoost models, all features affect the estimated friction and contribute to a reasonable direction. Apart from that, latitude and altitude also play an important role in these models compared to Regression Trees. Since the correlation between friction and latitude and altitude is negative, it is reasonable that high values of these features would lower the estimated friction. In this instance, the SVR model shows conflicting contributions and rankings compared to tree models. Specifically, the low longitude contributes to a higher friction, while air temperature above 0 leads to a lower friction estimate, which is illogical.

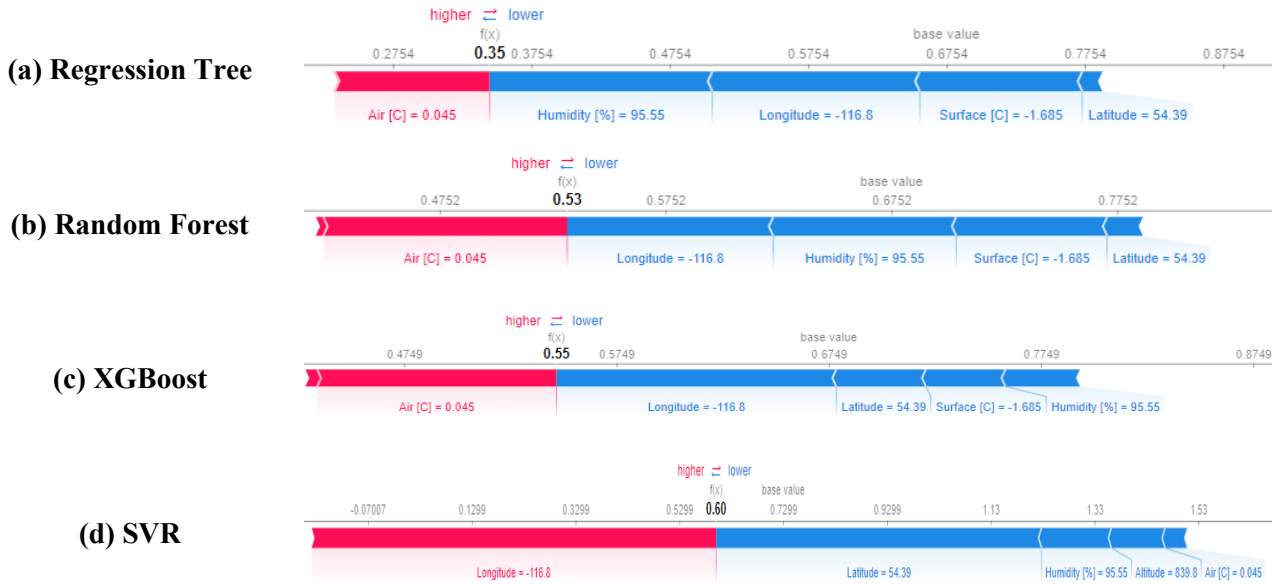


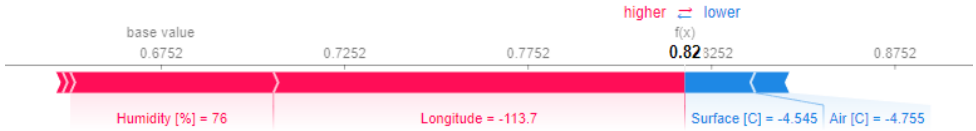
Figure 5-10 Force Plot of Instance 4 (Med Risk)

Instance 4 (**Figure 5-10**) represents a medium-risk scenario, with a true friction coefficient of 0.4875 and the following feature values: air temperature of 0.045°C, surface temperature of -1.685 °C, relative humidity of 95.55%, latitude of 54.39, longitude of -116.88, and altitude of 839.8 meters. Compared with the mean value of each feature, the longitude is lower and the rest of them are higher.

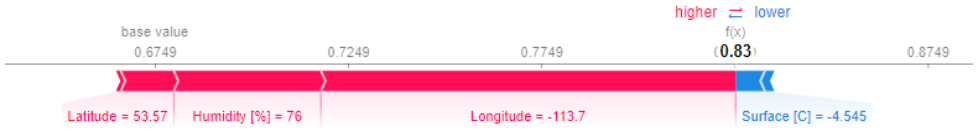
In the three tree models, all features except altitude have a significant effect on the friction estimate, and each feature contributes logically. High humidity and latitude, along with low longitude and surface temperature, lead to a low friction estimate, while air temperature above 0°C increases friction. Comparing the friction estimate generated by SHAP reveals that the estimates from the Random Forest and the XGBoost are closer to the observed true value than that from the Regression Tree. However, the SVR model differs from the tree model in several ways. First, surface temperature has a marginal effect on friction estimation in the SVR model, whereas it contributes significantly to the tree models. Second, altitude has a significant effect on the friction estimate in the SVR model, which is consistent with the logical expectation that a higher altitude would reduce friction. In addition, the contribution directions of longitude and temperature in SVR contradict those observed in the tree models, suggesting an unreasonable influence of them on the friction estimation.

(a) Regression Tree

(b) Random Forest



(c) XGBoost



(d) SVR

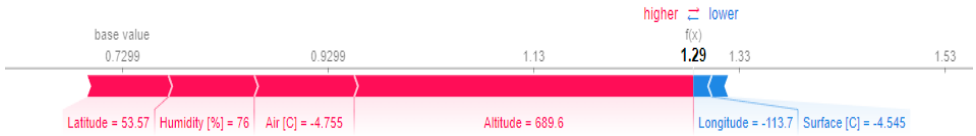


Figure 5-11 Force Plot of Instance 5 (Low Risk)

The low-risk instance 5 (**Figure 5-11**) has a true coefficient of friction of 0.82 and the features are a temperature of -4.755°C, a surface temperature of -4.545°C, a relative humidity of 76%, a latitude of 53.57, a longitude of -113.7, and an altitude of 689.6 meters. All are below the mean except longitude, which is above the mean.

In the three tree models, the contribution direction of each feature value is logical, where the features in the Regression Tree that contribute to the friction estimate are longitude, relative humidity, altitude, surface temperature, and air temperature. Among them, low altitude, low relative humidity, and high longitude lead to a high friction estimate, while low surface and air temperature lead to a low friction estimate. The Random Forest model places less emphasis on the effect of altitude on the friction estimate compared to the Regression Tree. Additionally, in the XGBoost model, the contribution of air temperature is negligible compared to the Random Forest, while the contribution of latitude is more prominent, with low latitude increasing the friction estimates, which is expected. In the SHAP force plot of the SVR model, we can see that the estimated friction coefficient exceeds 1, indicating that the model has the potential to produce over-range estimates.

Instance 6 (**Figure 5-12**), marked as low risk, has a true friction coefficient of 0.82, and the features values are: air temperature of 2.22°C, surface temperature of -2.46°C, relative humidity of 64%, latitude of 51.41, longitude of -114, and altitude of 1132 meters. In this instance, air temperature,

surface temperature, longitude, and altitude are higher than the mean value, while relative humidity and latitude are lower than the mean.

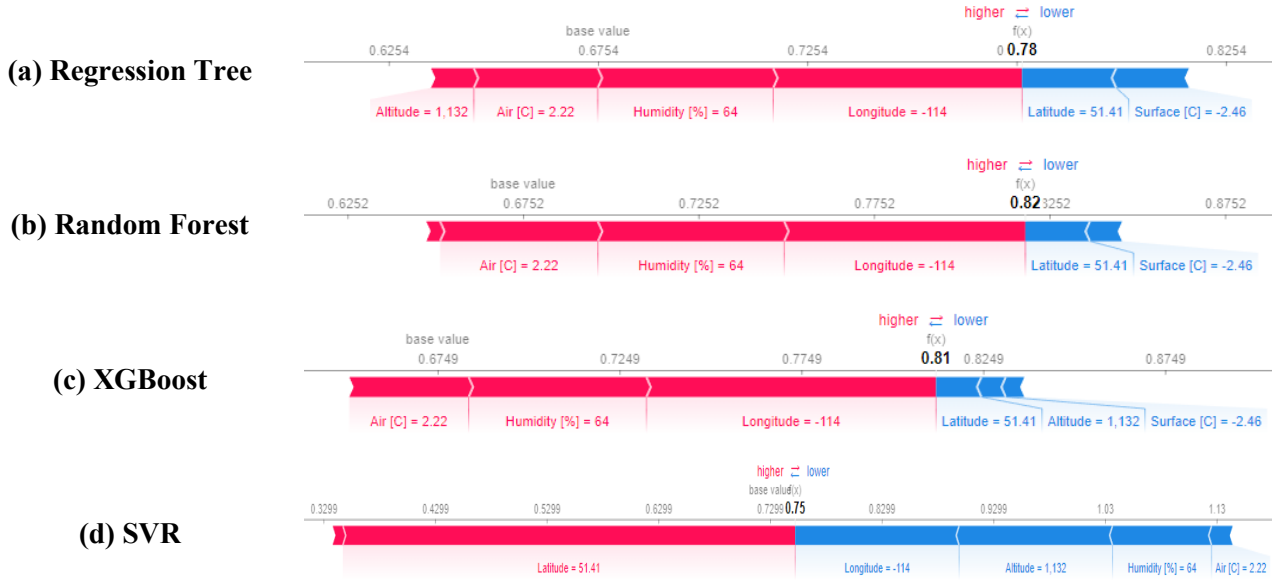


Figure 5-12 Force Plot of Instance 6 (Low Risk)

The force plot of the Regression Tree suggests that it is reasonable that high temperature, longitude, and low humidity increase the estimated friction, while high latitude and low surface temperature decrease the estimated friction. However, in this instance, a high altitude increases friction, contradicting our expectations and implying that the Regression Tree model may be flawed. Additionally, all features of the Random Forest contribute to the friction estimate in the same way as the Regression Tree; the most notable difference being that the contribution of altitude in the Random Forest is subtle and therefore can be ignored. In contrast, the importance of the features in XGBoost illustrates the superiority of XGBoost because high altitude helps to reduce the friction estimate rather than increase it like what was observed in Regression Tree and Random Forest. For the SVR model, the contribution direction for longitude, altitude, and air temperature are reasonable. However, it is irrational that low latitude leads to increased friction, and low humidity reduces friction.

In the analysis presented above, we examined two instances for each risk category, where the feature values exhibited distinct contributions. By doing so, we avoided the randomness of interpretations and demonstrated that the generation mechanism of each friction estimate differs in local instances. The analysis of force plots for local instances reveals that the structure of the

SVR model is relatively weak and prone to logical misestimation. Therefore, caution is needed when using this model. On the other hand, in most cases, the three tree models (Regression Tree, Random Forest, and XGBoost) work in a similar and logical way. In some of the cases, by examining the internal logic of the model, we can find that the overall reliability of the tree-based model is ranked as XGBoost, Random Forest, and Regression Tree, which confirms that an increase in model complexity can improve the performance of the model. The above results show that force maps with local examples can help us gain insight into the model's inner workings and provide guidance for choosing the right model.

As presented above, the application of SHAP models for both global and local interpretation equips us with tools to evaluate and fine-tune machine learning (ML) model performance. Our findings corroborate that the SHAP method is effective in augmenting the interpretability of ML models, enabling us to better understand their underlying mechanisms.

5.4 Transfer Learning

5.4.1 Model Generalizability on the External Validation Dataset

A model with strong generalization capabilities is essential as it ensures it performs well on new and unseen datasets. To evaluate the generalization capabilities of the models, we applied the four models to the external dataset. This evaluation will help us determine which model can be widely applied in real-world scenarios, maximizing its utility.

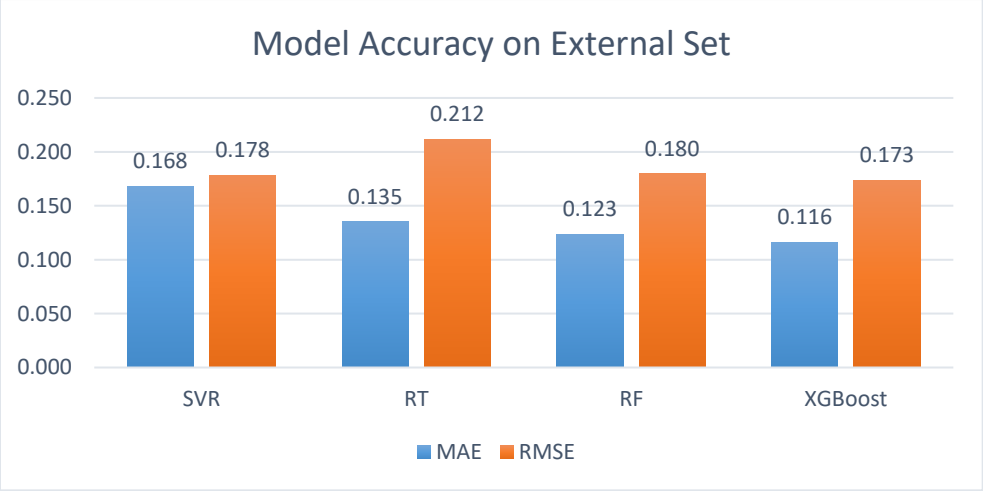


Figure 5-13 Model Accuracy on External Validation Set

After applying models to the external dataset, the mean absolute error (MAE) and root mean square error (RMSE) were calculated and shown in **Figure 5-13**. As the figure illustrates, the overall rank of model performance is consistent with the results observed in the previous section on the test set, though the overall accuracy dropped due to changes in dataset characteristics. Nevertheless, the estimation errors of the models remained low (below 0.2), signifying the robustness of the models developed.

To examine if this increase in error can lead to misclassifications, we converted the predicted friction values into risk levels based on the previously described system. The results from this analysis are shown in **Figure 5-14**. As depicted, Extreme Gradient Boosting (XGBoost) and Random Forest maintained a high accuracy of around 85%, while the classification accuracy of the Regression Tree model dropped significantly to 80% or lower, further confirming the negative relationship between model accuracy and complexity. Surprisingly, the Support Vector Regression (SVR) model achieves the highest accuracy on the testing dataset. To explore this anomaly, we carefully examined the confusion matrix and found that the SVR model's estimation ability on the external dataset is quite poor. As **Figure 5-15** suggests, the friction estimates generated by SVR are only in the low-risk category. It is evident that all the estimated values on this dataset are distributed in the low-risk region. Therefore, even though it exhibited an accuracy of 87.04%, the model is not trustworthy. Hence, the SVR model was rejected due to the lack of generalization capability. On the other hand, the other tree models were able to accurately estimate friction values for each risk category.

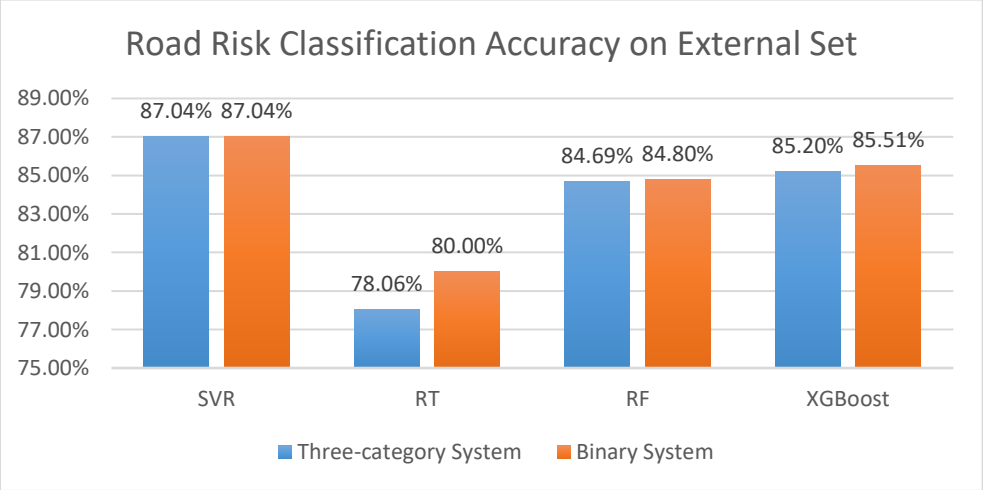
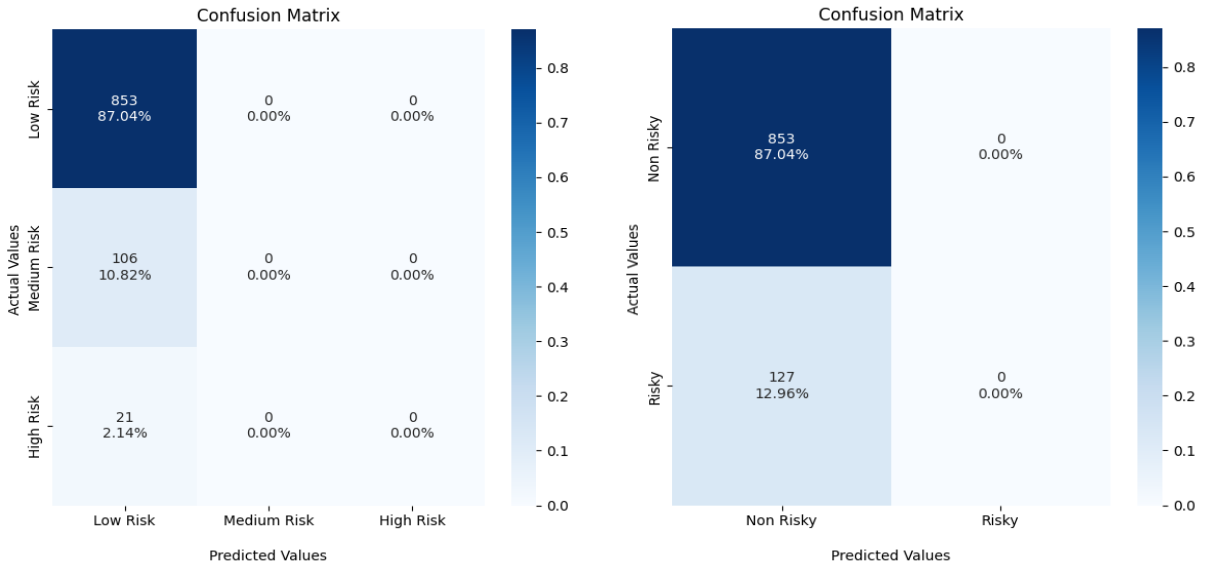


Figure 5-14 Road Risk Classification Accuracy on External Validation Set



(a) Three-Category System **(b) Binary System**

Figure 5-15 Confusion Matrix of SVR Model on Validation Set of (a) Three-Category System; (b) Binary System

Based on this evaluation, we can conclude that the SVR model built in this study lacks generalization capabilities, while the tree-based models are much better at adapting to new datasets. However, the performance of the tree models can still be somewhat affected due to the different dataset characteristics.

5.4.2 Transfer Learning

After evaluating the model performance and analyzing the SHapley Additive exPlanations (SHAP) explanations, Extreme Gradient Boosting (XGBoost) emerged as the top-performing model with the most credible inner logic, making it the optimal choice for road friction estimation in this study. With the optimal model identified, our focus shifts to enhancing its transferability to address the issue of decreased model performance due to different data characteristics for broader applications.

Given that the new dataset contains only 980 samples, which are significantly fewer than the 93,101 samples used to train the original model, it is reasonable to expect that a new model built solely on this smaller dataset would underperform compared to the original. To overcome this, we employed transfer learning. This technique leverages a small amount of new data to update the existing model, thus enhancing its adaptability to new datasets and producing more reliable estimates. This strategy effectively solves two problems: the inapplicability of older models to new

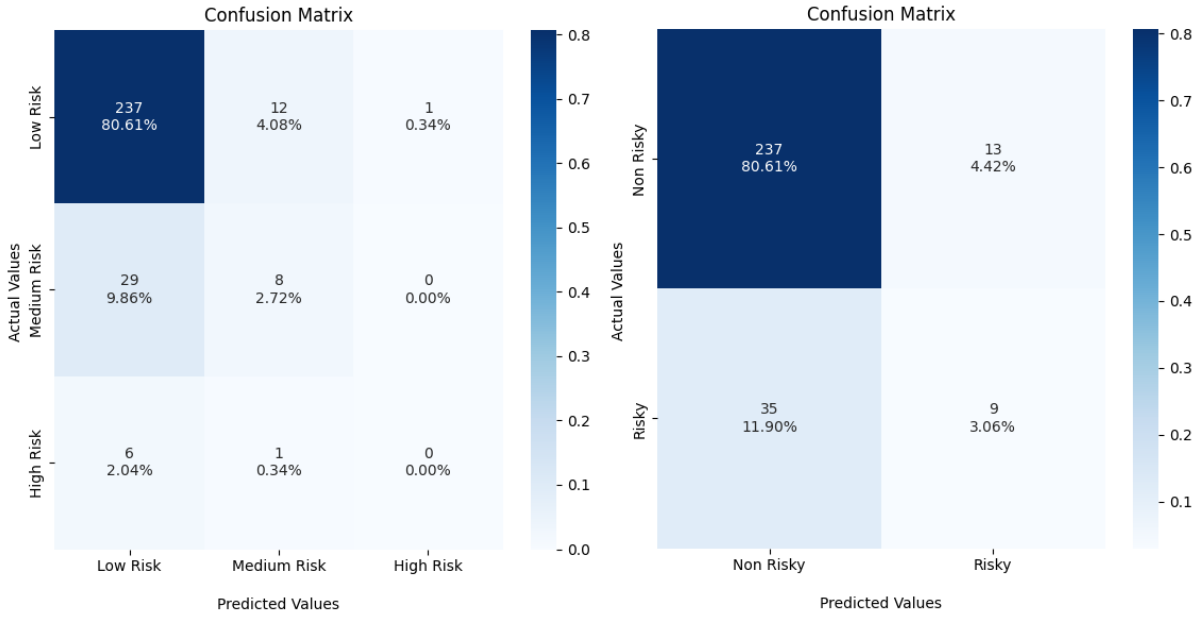
datasets and the insufficiency of new data for training a completely new model with satisfactory performance.

We began by splitting the external data set, which consisted of 980 samples, into a training set and a testing set. The training set, comprising 70% of the data, was utilized to develop the transfer learning model using the two-stage TrAdaBoost.R2 algorithm. Regarding the model hyperparameter values, 100 estimators, 15 steps, and 5 folds were selected; these values were obtained from parameter tuning via random search. After calibrating the model on the training set, it was applied to the testing set containing 295 samples for evaluation. We compared the confusion matrices of the primary XGBoost model and the improved XGBoost model (**Figure 5-16**).

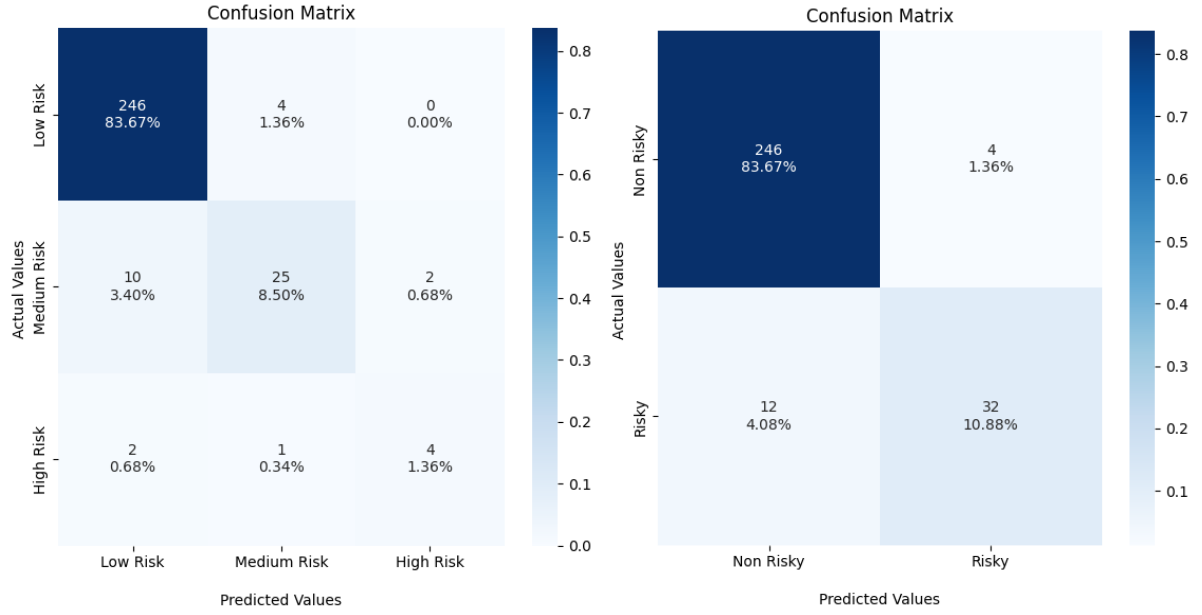
From **Figure 5-16**, it can be observed that the primary XGBoost model failed to identify high-risk road conditions in the three-category system, which is detrimental to providing accurate information to winter road maintenance (WRM) personnel and the general public. Misclassifying high-risk road conditions as lower risk can lead to complacency, leading to an increased risk of traffic crashes. Similarly, in the binary system, 35 out of 44 risky road conditions were incorrectly labeled as non-risky. As a result of the increase in mislabels, transfer learning is required to calibrate the model to the new dataset.

Following transferring learning using the TrAdaBoost.R2, model performance improved significantly, particularly in the three-category system, where the improved model demonstrated a much higher rate of correct classifications for each risk level. Furthermore, the binary system successfully identified 32 risky road conditions, which was a drastic performance boost compared to the previous 35 mislabels.

To quantify the improvements achieved through transfer learning, we compared model performance in terms of several error metrics (**Table 5-2)**. The evaluation results demonstrate the positive impact of transfer learning. Specifically, we observed a notable decrease in the mean absolute error (MAE) by 0.0965, the mean squared error (MSE) by 0.0289, and the root mean squared error (RMSE) by 0.1159. Moreover, the accuracy for the three-category classification increased by 10.21%, and there was a 10.89% improvement in binary classification accuracy. These significant enhancements in model performance highlight the necessity and effectiveness of applying transfer learning techniques.



(a) Primary XGBoost



(b) Improved XGBoost

Figure 5-16 Confusion Matrix of (a) Primary XGBoost Model; and (b) Improved XGBoost Model with three-Category System (left) and Binary System (right) on Test Set of External Data

Table 5-2 Comparison of Model Performance

	MAE	MSE	RMSE	Three-category Classification	Binary Classification
Primary XGBoost	0.1261	0.0333	0.1824	83.33%	83.67%
Improved XGBoost	0.0296	0.0044	0.0665	93.54%	94.56%
Model Improvement	-0.0965	-0.0289	-0.1159	10.21%	10.89%

6 CONCLUSIONS

In this chapter, we provide a detailed summary of the work that has been done in this thesis. Firstly, we give an overview of the thesis. After that, we present the main findings of the study. We will then summarize the contributions made by this work, highlighting the progress achieved in the field and the impact of the research. Lastly, we discuss the limitations faced during the study and provide suggestions for future research.

6.1 Overview of the Thesis

Efficient winter road maintenance (WRM) operations are essential for enhancing traffic safety and efficiency during winter conditions. However, achieving this objective necessitates the availability of accurate road surface condition (RSC) information, which can be challenging to obtain. This thesis proposes a framework for estimating road friction via weather and geographic data to address this issue. The model incorporates Ordinary Kriging (OK) to fill in missing data, enabling spatially continuous estimation of road friction. Additionally, a road risk classification system is developed based on friction estimates, providing concise and intuitive information for WRM personnel and road users.

Furthermore, efforts were made to explore more complex tree-based algorithms for predicting friction. Between Regression Trees, Random Forests, Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) models, the XGBoost model displayed the highest accuracy while also maintaining an intuitive decision-making process. Part of this process involved using explainable AI techniques, i.e., SHapley Additive exPlanations (SHAP), to investigate the underlying reasoning of the models, thereby making it more transparent and accessible to the user. To enhance the application scope of the friction estimation model further, transfer learning techniques were explored. The results from this analysis revealed that leveraging transfer learning allows the model to adapt to new datasets and scenarios, enhancing the model's transferability and utility in practical WRM operations.

6.2 Key Findings of the Thesis

The key findings of this thesis encompass two main aspects of road friction estimation and model enhancement. In the first part, we explore the development of a friction estimation model, its correlation with weather and geographic factors, and the production of road risk maps. In the second part, we discuss the enhancement of the model through the exploration of more complex algorithms, interpretability, and transferability. The details are provided below.

Road Friction Estimation and Risk Mapping

A friction estimation model was calibrated using a Regression Tree algorithm with weather and geographic factors as input features. The correlation matrix suggested road friction is positively correlated with air temperature, surface temperature, and longitude; and negatively correlated with relative humidity, latitude, and altitude. The features at the decision nodes shown in the model visualization were in the same order as the correlation rank, indicating a correct internal logic of the model. With a high accuracy rate of 93.3%, the model demonstrated that road friction can be explained by weather and geography. To predict features at unsampled locations between stationary road weather information system (sRWIS) stations, Ordinary Kriging (OK), a geostatistical interpolation method, was applied. The OK interpolations were similar to mobile RWIS (mRWIS) measurements despite certain discrepancies due to the omission of snow depth and predictor estimation error.

A road risk map was produced based on estimated road friction and defined risk level. OK was adopted to generate continuous weather information. These interpolated weather data were then inputted into the friction estimation model to obtain a road friction map. Risk thresholds of 0.3, 0.4, and 0.5 were chosen to categorize friction as risky and non-risky. The validation accuracy reached 84.25%, 89.76%, and 88.98% for three different thresholds in increasing order. By increasing the number of risk categories to three, high (<0.3), medium (0.3–0.5), and low (>0.5), we were able to generate a more comprehensive risk map with a 77.95% accuracy. The prediction errors consisted mostly of False Risky (18.90%) and low False Non-risky (3.15%), indicating our developed model is conservative and advantageous from a safety perspective.

The results obtained show that the proposed framework is robust and feasible. By generating a road friction estimation map, winter road maintenance (WRM) authorities can deliver more

comprehensive assessments of road conditions to road users and make their own operations more efficient.

Advanced Model Analysis via Explainable Artificial Intelligence and Transfer Learning

In order to improve the accuracy of the friction estimation model, we explored more suitable Machine learning (ML) algorithms and compared their performance. Four different models, namely Regression Trees, Random Forests, Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR), were selected and trained on the aggregated mRWIS dataset with a one-minute interval. The models achieved R^2 accuracies of 84.17%, 89.99%, 91.39%, and 85.83%, respectively. It is evident that as the complexity of the models increased, their accuracy improved, with XGBoost outperforming the others and Regression Trees exhibiting the lowest accuracy.

However, as the models became more complex, their interpretability worsened. To address this issue, we introduced SHapley Additive exPlanations (SHAP) to explain the complex models. Through SHAP global explanations, we observed that the feature contributions of the tree-structured models were similar, with minor variations, while the SVR model was completely different. When we examined the direction of the contribution, it became apparent that the tree-structured models aligned better with the real-world logic, while SVR exhibited notable discrepancies. To further validate this observation, we performed local explanations and examined selected model instances in detail. The results generated further confirmed our finding that only tree-based methods created intuitive models. Despite the SVR being slightly more accurate compared to the Regression Tree model, we rejected the SVR due to its limited interpretability. It is evident that although slight deviations from actual logic were observed, the tree-structured models demonstrated reasonable feature contributions in each specific instance. Among them, Regression Trees showed the greatest degree of deviation, followed by Random Forests. Only XGBoost consistently exhibited a reliable inner logic. By utilizing SHAP, we improved the interpretability of complex models and further confirmed the superiority of XGBoost for friction estimation.

Finally, we applied the optimal XGBoost model to an external mRWIS dataset and utilized the two-stage TrAdaBoost.R2 for transfer learning. The RMSE of the model after transfer learning was 0.066, which represented a significant reduction compared to the RMSE of 0.182 obtained by applying the primary XGBoost. These findings demonstrate the success of transfer learning and

prove its effectiveness in helping to migrate models to new datasets, thus expanding the range of applications for which the models can be used.

6.3 Contributions of the Thesis

Through the research conducted in this thesis, significant contributions have been made toward the accurate estimation of road friction coefficients, enabling the provision of timely and precise road surface condition (RSC) information. These contributions can be broadly classified into methodological and practical contributions as follows:

Methodological Contributions

- **Established an accurate estimation model of winter road friction coefficients by integrating stationary and mobile road weather information system (sRWIS and mRWIS) data.** Utilizing data from both RWIS systems, the models developed in this thesis overcome the drawbacks of each data source, offering a practical method for real-time estimation of road friction across a continuous space.
- **Conducted a comprehensive comparative analysis to identify the best-performing machine-learning algorithm.** By comparing models of varying complexity, this study identified that Extreme Gradient Boosting (XGBoost) model is the most accurate algorithm for estimating road friction coefficients.
- **Utilized SHapley Additive exPlanations (SHAP) explainable artificial intelligence (AI) to improve the interpretability and performance of complex machine learning (ML) algorithms in winter road friction estimation.** The integration of the SHAP model significantly improves the models' interpretability, highlighting the importance of combining explainable AI with ML in winter road friction estimation.
- **Adopted transfer learning techniques to extend the model's adaptability to new datasets, expanding its applicability.** Faced with limited new data, this thesis leveraged the two-stage TrAdaBoost.R2 technique with existing models and a small amount of new data, resulting in improved model performance. These findings serve as practical guidance for the application of transfer learning and substantiate the potential to broaden model applicability and performance across varied environments.

Practical Contributions

- **Enhanced Operational Efficiency for winter road maintenance (WRM) Personnel.** The methodologies developed in this thesis enable WRM personnel to prioritize maintenance activities based on the level of road risk, thereby maximizing operational efficiency. By offering the potential for real-time estimation of road friction, maintenance departments can deliver more targeted and timely responses to changing road conditions.
- **Improved safety and convenience for drivers during winter travel.** The work conducted in this thesis lays the groundwork for providing drivers with real-time information on road conditions. This has the potential to enhance safety and convenience during winter travel, contributing to more informed decision-making on the road and potentially reducing the risk of accidents related to adverse weather conditions.

In light of above, these methodological and practical contributions collectively advance the understanding and capabilities of WRM and safety, whereby offering valuable insights and tools that can be applied in both research and real-world settings.

6.4 Limitations and Future Work

While the findings of this thesis have broad implications, it is important to acknowledge that the study is not without limitations. The most evident limitation is the availability of datasets. The dataset used in this thesis does not contain parameters like snow depth, which could potentially contribute to improving the model performance. As mentioned in Chapter 4, the absence or presence of snow on the road surface has a strong correlation with road friction, meaning friction notably decreases with increasing snow depth and vice versa.

In addition, there is a lack of consideration of traffic attributes, as heat from heavy traffic is known to increase road surface temperature and thus affect road surface conditions (RSCs) (e.g., making snow-covered roads slushy). Furthermore, limited sample size results in insufficient spatial coverage of the data, thereby restricting model performance to local areas with reduced transferability potential. Although geographical characteristics have been taken into consideration to improve the spatial representation of variables under investigation, the use of more datasets

covering large areas could further improve the conclusiveness as well as the transferability of the findings documented herein.

Furthermore, for the comparative analysis section, only four machine learning algorithms were compared in this paper, while the inclusion of additional different machine learning algorithms could be considered in the future for a more comprehensive comparison.

On the other hand, there are also limitations in the application of SHapley Additive exPlanations (SHAP). This thesis did not address the potential model flaws revealed in the SHAP model explanations. Additionally, another limitation is related to the accuracy of SHAP. As a simplified interpretation model, SHAP cannot fully elucidate the intricate inner workings of the model.

Future research is therefore required to address the limitations mentioned above. First, the study area can be expanded to include more data points and more predictors, such as snow depth. It is also critical to implement more spatiotemporally comprehensive datasets that can capture the diversity of geographical and topographical characteristics. In addition, traffic volume and winter road maintenance (WRM) activities should be considered to account for the influence of human activities. Likewise, more advanced variants of kriging interpolation techniques, such as network kriging or regression kriging should be adopted to improve interpolation accuracy further. In terms of the SHAP application, addressing the model flaws identified by SHAP, such as removing features with illogical contributions, could improve model performance.

Bibliography

[1] International Transport Forum, "Detailed country profile for Canada," OECD, Paris, 2020.

[2] T. S. Alberta Transportation, "Alberta Traffic Collision Statistics 2018," Feburary 2018. [Online].

[3] Transport Canada, "National Collision Database Online 1.0," 13 April 2012. [Online]. Available: https://wwwapps2.tc.gc.ca/saf-sec-sur/7/ncdb-bndc/p.aspx?c=100-0-0&l=en.

[4] C.-G. Wallman and H. Åström, "Friction measurement methods and the correlation between road friction and traffic safety: A literature review," Swedish National Road and Transport Research Institute, Linköping, 2001.

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

[6] L. Goodwin and P. Pisano, "Weather-responsive traffic signal control," *ITE Journal,* vol. 74, no. 6, pp. 28-33, 2004.

[7] R. R. Blackburn, "Snow and ice control: Guidelines for materials and methods," *Transportation Research Board,* vol. 526, 2004.

[8] S. Du, M. Akin, D. Bergner, G. Xu and X. Shi, "Material application methodologies for winter road maintenance: a renewed perspective," *Canadian Journal of Civil Engineering,* vol. 49, no. 1, pp. 1-10, 2022.

[9] T. Usman, L. Fu and L. F. Miranda-Moreno, "Quantifying safety benefit of winter road maintenance: Accident frequency modeling," *Accident Analysis & Prevention,* pp. 1878-1887, 2010.

[10] C.-G. Wallman, P. Wretling and G. Öberg, "Effects of winter road maintenance: state-of-the-art," Swedish National Road and Transport Research Institute, Linköping, 1997.

[11] R. M. Hanbali, "Economic impact of winter road maintenance on road users.," *Transportation Research Record,* pp. 151-161, 1994.

[12] G. o. Alberta, "Alberta Transportation Annual Report 2021-2022," Government of Alberta, 2022.

[13] J. Ilkka, N. Pertti and H. Marjo, "Statistical modelling of wintertime road surface friction," *Meteorological applications,* vol. 20, pp. 318-329, 2013.

[14] T. J. Kwon, L. Fu and S. J. Melles, "Location optimization of road weather information system (RWIS) network considering the needs of winter road maintenance and the traveling public," *Computer-Aided Civil and Infrastructure Engineering,* vol. 32, no. 1, pp. 57-71, 2017.

[15] M. Rasol, F. Schmidt and S. Ientile, "FriC-PM: Machine Learning-based road surface friction coefficient predictive model using intelligent sensor data," *Construction and Building Materials,* p. 130567, 2023.

[16] F. Minges, "Neural Network-based road friction estimation using road weather information," *Chalmers tekniska högskola / Institutionen för mekanik och maritima vetenskaper,* 2020.

[17] G. Panahandeh, E. Ek and N. Mohammadiha, "Road friction estimation for connected vehicles using supervised machine learning," in *2017 IEEE intelligent vehicles symposium (IV)*, Los Angeles, 2017.

[18] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti and D. Pedreschi, "A Survey of Methods for Explaining Black Box Models," *ACM Computing Surveys,* pp. 1-42, 2018.

[19] V. Belle and I. Papantonis, "Principles and Practice of Explainable Machine Learning," *Frontier in Big Data,* 2021.

[20] L.-V. Herm, K. Heinrich, J. Wanner and C. Janiesch, "Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability," *International Journal of Information Management,* p. 102538, 2023.

[21] X. Shi and L. Fu, "Sustainable winter road operations (1st ed.)," Wiley Blackwell, 2018.

[22] H. Marjo, J. Ilkka and N. Pertti, "A statistical forecast model for road surface friction.," in *15th International Road Weather Conference, SIRWEC*, Quebec, 2010.

[23] M. Kangas, M. Heikinheimo and M. Hippi, "RoadSurf: a modelling system for predicting road weather and road surface conditions," *Meteorological Applications,* pp. 544-553, 2015.

[24] Y. Takasaki, M. Saldana, J. Ito and K. Sano, "Development of a method for estimating road surface condition in winter using random forest," *Asian Transport Studies,* 2022.

[25] M. Teke and F. Duran, "The design and implementation of road condition warning system for drivers," *Measurement and Control,* pp. 985-994, 2019.

[26] M. A. Linton and L. Fu., "Connected vehicle solution for winter road surface condition monitoring," *Transportation Research Record,* vol. 2551, no. 1, pp. 62-72, 2016.

[27] S. Kim, J. Lee and T. Yoon, "Road surface conditions forecasting in rainy weather using artificial neural networks.," *Safety science,* p. 105302, 2021.

[28] A. Novikov, I. Novikov and A. Shevtsova, "Study of the impact of type and condition of the road surface on parameters of signalized intersection," *Transportation Research Procedia,* vol. 36, pp. 548-555, 2018.

[29] M.-T. Do, V. Cerezo, Y. Beautru and M. Kane, "Influence of thin water film on skid resistance," *Journal of Traffic and Transportation Engineering,* vol. 2, no. 1, pp. 36-44, 2014.

[30] K. Ichihara and M. Mizoguchi, "Skid resistance of snow-or ice-covered roads," *Highway Research Board Special Report,* vol. 115, 1970.

[31] Y. Luo, "The Effect of Pavement Temperature on Frictional Properties of Pavement Surfaces at the Virginia Smart Road," Virginia Tech, Virginia, 2003.

[32] D. J. Mildrexler, M. Zhao and S. W. Running, "Satellite finds highest land skin temperatures on Earth," *Bulletin of the American Meteorological Society,* vol. 92, no. 7, pp. 855-860, 2011.

[33] "Temperature Over Time," NASA, [Online]. Available: https://www.ces.fau.edu/nasa/module-3/why-does-temperature-vary/elevation.php#:~:text=Temperature%20normally%20decreases%20as%20elevation,at%20higher%20elevations%20is%20cooler.. [Accessed 9 July 2023].

[34] M. Rasol, F. Schmidt, S. Ientile, L. Adelaide, B. Nedjar, M. Kane and C. Chevalier, "Progress and monitoring opportunities of skid resistance in road transport: a critical review and road sensors," *Remote Sensing,* vol. 13, no. 18, p. 3729, 2021.

[35] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608,* 2017.

[36] 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.

[37] R. Wexler, "When a computer program keeps you in jail: How computers are harming criminal justice," *New York Times,* 2017.

[38] M. McGough, "How Bad Is Sacramento's Air, Exactly? Google Results Appear at Odds with Reality, Some Say," 2018. [Online]. Available: https://www.sacbee.com/news/state/california/fires/article216227775.html. [Accessed 9 July 2023].

[39] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission.," in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, New York, NY, USA, 2015.

[40] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *arXiv preprint arXiv: 1705,* 2017.

[41] L. S. Shapley, "A value for n-person games," pp. 307-317, 1953.

[42] Y. Liu, Z. Liu, X. Luo and H. Zhao, "Diagnosis of Parkinson's disease based on SHAP value feature selection," *Biocybernetics and Biomedical Engineering,* vol. 42, no. 3, pp. 856-869, 2022.

[43] C.-T. Kor, Y.-R. Li, P.-R. Lin, S.-H. Lin, B.-Y. Wang and C.-H. Lin, "Explainable machine learning model for predicting first-time acute exacerbation in patients with chronic obstructive pulmonary disease," *Journal of personalized medicine,* vol. 12, no. 2, p. 228, 2022.

[44] Y. Zhang, D. Yang, Z. Liu, C. Chen, M. Ge, X. Li, T. Luo, Z. Wu, C. Shi, B. Wang, X. Huang, X. Zhang, S. Zhou and Z. Hei, "An explainable supervised machine learning predictor of acute kidney injury after adult deceased donor liver transplantation," *Journal of translational medicine,* vol. 19, no. 1, pp. 1-15, 2021.

[45] C. Duckworth, F. P. Chmiel, D. K. Burns, Z. D. Zlatev, N. M. White, T. W. Daniels, M. Kiuber and M. J. Boniface, "Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19," *Scientific reports,* vol. 11, no. 1, p. 23017, 2021.

[46] N. Farzaneh, C. A. Williamson, J. Gryak and K. Najarian, "A hierarchical expert-guided machine learning framework for clinical decision support systems: an application to traumatic brain injury prognostication," *NPJ digital medicine,* vol. 4, no. 1, p. 78, 2021.

[47] L. Roa, A. Correa-Bahnsen, G. Suarez, F. Cortés-Tejada, M. A. Luque and C. Bravo, "Super-app behavioral patterns in credit risk models: Financial, statistical and regulatory implications," *Expert Systems with Applications,* vol. 169, p. 114486, 2021.

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

[49] X. Xiaomao, Z. Xudong and W. Yuanfang, "A comparison of feature selection methodology for solving classification problems in finance," *Journal of Physics: Conference Series,* vol. 1284, no. 1, p. 012026, 2019.

[50] M. T. Ribeiro, S. Singh and C. Guestrin, "" Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, New York, NY, United States, 2016.

[51] A. Gramegna and P. Giudici, "SHAP and LIME: an evaluation of discriminative power in credit risk," *Frontiers in Artificial Intelligence,* vol. 4, p. 752558, 2021.

[52] X. Man and E. P. Chan, "The best way to select features? comparing mda, lime, and shap," *The Journal of Financial Data Science,* vol. 3, no. 1, pp. 127-139, 2021.

[53] Y. Hailemariam, A. Yazdinejad, R. M. Parizi, G. Srivastava and A. Dehghantanha, "An empirical evaluation of AI deep explainable tools," in *2020 IEEE Globecom Workshops*, Taipei, 2020.

[54] H. T. T. Nguyen, H. Q. Cao, K. V. T. Nguyen and N. D. K. Pham, "Evaluation of explainable artificial intelligence: Shap, lime, and cam," in *Proceedings of the FPT AI Conference*, 2021.

[55] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering,* vol. 22, no. 10, pp. 1345-1359, 2009.

[56] W. Dai, Q. Yang, G.-R. Xue and Y. Yu, "Boosting for transfer learning," in *Proceedings of the 24th international conference on Machine learning*, Corvallis, OR, 2007.

[57] H. He, K. Khoshelham and C. Fraser, "A multiclass TrAdaBoost transfer learning algorithm for the classification of mobile lidar data," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 166, pp. 118-127, 2020.

[58] D. Tang, X. Yang and X. Wang, "Improving the transferability of the crash prediction model using the TrAdaBoost. R2 algorithm," *Accident Analysis & Prevention,* vol. 141, p. 105551, 2020.

[59] G. De'ath and K. E. Fabricius, "Classification and regression trees: a powerful yet simple technique for ecological data analysis," *Ecology,* vol. 81, no. 11, pp. 3178-3192, 2000 November.

[60] H. Jorda , M. Bechtold, N. Jarvis and J. Koestel, "Using boosted regression trees to explore key factors controlling saturated and near-saturated hydraulic conductivity," *European Journal of Soil Science,* vol. 66, no. 4, pp. 744-756, 2015 July.

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

[62] G. Biau and E. Scornet, "A random forest guided tour," *Test ,* vol. 25, pp. 197-227, 2016.

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

[64] H. Drucker, C. J. Burges, L. Kaufman, A. Smola and V. Vapnik., "Support vector regression machines," *Advances in neural information processing systems,* vol. 9, 1996.

[65] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing,* vol. 14, pp. 199-222, 2004.

[66] R. A. Olea, "A six-step practical approach to semivar-iogram modeling," *Stochastic Environmental Research and Risk Assessment,* vol. 20, no. 5, pp. 307-318, 2006.

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

[68] P. Goovaerts, Geostatistics for Natural Resources Eval-uation, New York: Oxford University Press, 1997.

[69] D. Pardoe and P. Stone, "Boosting for regression transfer," in *Proceedings of the 27th International Conference on International Conference on Machine Learnin*, 863-870, 2010.

[70] Y. a. S. R. Freund, "A decision-theoretic generalization of online learning and an application to boosting," *Journal of Computer and System Sciences,,* vol. 55, pp. 119-139, 1997.

[71] K. Krivoruchko, "GIS and Geostatistics: Spatial Analysis of Chernobyl's Consequences in Belarus," Workshop on Status and Trend in Spatial Analysis, Santa Barbara, CA, 1998.

[72] K. Krivoruchko, "Using linear and non-linear kriging interpolators to produce probability maps," 2001.

[73] K. Krivoruchko, A. Gribov and J. Ver Hoef, "Predicting Exact, Filtered, and New Values using Kriging," *Stochastic Modeling and Geostatistics,* vol. 2, 2000.

[74] T. Haavasoja, V. Haavisto, M. Turunen, P. Nylander, V. Oyj and Y. Pilli-Sihvola, "A Field Trial of a Vehicle's Grip Compared with RWS Data," vol. 159, 2002.

[75] P. Saarikivi, M. Hippi, P. Nurmi and J. Sipilä, "Observing the variability of road and weather conditions with hybrid mobile and fixed sensors," 2008.

[76] A. Abohassan, K. El-Basyouny and T. J. Kwon, "Effects of Inclement Weather Events on Road Surface Conditions and Traffic Safety: An Event-Based Empirical Analysis Framework," *Transportation Research Record,* 2022.

[77] I. Al-Qadi, A. Loulizi, G. Flintsch, D. Roosevelt, R. Decker, J. Wambold and W. Nixon, "Feasibility of Using Friction Indicators to Improve Winter Maintenance Operations and Mobility (Tech. Rep.).," 2002.

[78] L. Fu, L. Thakali, T. J. Kwon and T. Usman, "A risk-based approach to winter road surface condition classification," *Canadian Journal of Civil Engineering,* vol. 44, no. 3, pp. 182-191, 2017.

[79] T. A. o. Canada, "Winter Road Condition Terminology User Guide," 2011.

[80] "511 Alberta," [Online]. Available: https://511.alberta.ca/. [Accessed 15 August 2023].

[81] K. Ito, K. Hashimoto and Y. Shibata, "V2X communication system for sharing road alert information using cognitive network," in *Proceedings of the 8th International Conference on Awareness Science and Technology (iCAST)*, Taichung, Taiwan, 2017.

[82] M. Pomoni, "Exploring Smart Tires as a Tool to Assist Safe Driving and Monitor Tire–Road Friction.," *Vehicles,* vol. 4, pp. 744-765, 2022.

[83] X. Ding and T. J. Kwon, "Winter Road Friction Estimations via Multi-Source Road Weather Data—A Case Study of Alberta, Canada."," *Future transportation,* vol. 2, no. 4, pp. 970-987, 2022.

[84] Waterdown Collision, "Interesting Accident Statistics," October 2014. [Online]. Available: https://waterdowncollision.com/auto-repair/interesting-accident-statistics/.

[85] Highway Statistics Publications, "Highway Statistics Publications," U.S. Department of Transportation Federal Highway Administration, 2020. [Online]. Available: https://www.fhwa.dot.gov/policy/ohpi/hss/hsspubs.cfm.

[86] Z. Ye, C. K. Strong, L. Fay and X. Shi, "Cost Benefits of Weather Information for Winter Road Maintenance," Western Transportation Institute , Bozeman, 2009.

[87] F. Feng, L. Fu and M. S. Perchanok, "Comparison of alternative models for road surface condition classification," in *Transportation Research Board 89th Annual Meeting*, Washington DC, United States, 2010.

[88] K. Katko, "Goals and Methods of Winter Maintenance in Finland.," *Transportation Research Record,* vol. 1387, pp. 8-11, 1993.