GEORGIA DOT RESEARCH PROJECT 21-03

Final Report

ENHANCING THE GDOT'S MAINTENANCE DECISION TREES CONSIDERING THE EFFECTIVENESS OF VARIOUS TREATMENT OPTIONS IN DIFFERENT GEOGRAPHICAL LOCATIONS AND OVER TIME

Office of Performance-based Management and Research
600 West Peachtree Street NW | Atlanta, GA 30308

July 2024

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-GA-24-2103	2. Government Accession No. N/A	3. Recipient's Catalog No. N/A
4. Title and Subtitle Enhancing the GDOT's Maintenance Decision Trees Considering the Effectiveness of Various Treatment Options in Different Geographical Locations and Over Time		5. Report Date July 2024 6. Performing Organization Code N/A
7. Author(s) Brian Moore, Ph.D., KSU; Minsoo Baek, Ph.D., KSU; Baabak Ashuri, Ph.D., GT; Frederick Chung, GT; Yejee Paik, GT		8. Performing Organ. Report No. 21-03
9. Performing Organization Name and Address Kennesaw State University Office of Research 1000 Chastain Road, MD 0111 Kennesaw, GA 30144-5591		10. Work Unit No. N/A 11. Contract or Grant No. PI# 0018123
12. Sponsoring Agency Name and Address Georgia Department of Transportation (SPR) Office of Performance-based Management and Research 600 W. Peachtree Street NW Atlanta, GA 30308		13. Type of Report and Period Covered Final Report (April 2022 – July 2024) 14. Sponsoring Agency Code N/A
15. Supplementary Notes Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration.		
16. Abstract The Georgia Department of Transportation (GDOT) is responsible for pavement management of 7,100 centerline miles, encompassing diverse terrain types. The conventional "one size fits all" approach is no longer viable as pavement condition depends on several internal and external factors that are different from district to district, e.g., material, construction, environment, etc. Hence, the maintenance and rehabilitation criteria for pavements need to be studied and updated to account for regional effects, cost variation, and decision timing. This study seeks to enhance the GDOT's maintenance strategies through: (1) empirically analyzing the pavement deterioration in different geographical locations and climate conditions across the state and (2) improving life cycle cost estimates for the treatment options. The decision analysis methods will be capable of helping support maintenance and rehabilitation prioritization, optimizing budget allocation, and enhancing the reliability of statewide pavement performance.		
17. Key Words Transportation Asset Management, Pavement Management System, Machine Learning, Pavement Condition Prediction, Cost Forecasting Model	18. Distribution Statement No restrictions.	

19. Security Classification (of this report):	20. Security Classification (of this page):	21. No. of Pages:	22. Price:
Unclassified	Unclassified	96	Free

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

GDOT Research Project 21-03

Final Report

ENHANCING THE GDOT'S MAINTENANCE DECISION TREES CONSIDERING THE EFFECTIVENESS OF VARIOUS TREATMENT OPTIONS IN DIFFERENT GEOGRAPHICAL LOCATIONS AND OVER TIME

By

Brian Moore, Ph.D., P.E.
Professor. Kennesaw State University

Minsoo Baek, Ph.D.
Assistant Professor. Kennesaw State University

Baabak Ashuri, Ph.D., DBIA
Professor, Georgia Institute of Technology

Frederick Chung
Graduate Student, Georgia Institute of Technology

Yejee Paik
Graduate Student, Georgia Institute of Technology

Kennesaw State University Research and Service Foundation

Contract with
Georgia Department of Transportation

In cooperation with
U.S. Department of Transportation, Federal Highway Administration

July 2024

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
	NOTE: volumes greater than 1000 L shall be shown in m^3			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The Georgia Department of Transportation (GDOT) faces the critical challenge of managing and maintaining its extensive network of roadways amidst evolving environmental conditions, varying regional factors, and constrained budgets. In response, this project endeavors to enhance the GDOT's pavement management strategies through a comprehensive research initiative spanning two key areas: pavement deterioration modeling and maintenance cost estimation. It is essential to capture and differentiate pavement deterioration rates and analyze the effectiveness of various treatments and improvement options on pavement condition indicators at the district (or subdistrict) level to ultimately evaluate the trade-offs among various treatment options.

The research began with a comprehensive review of academic and professional literature relevant to the pavement asset management plans and experiences of other departments of transportation. The task mainly focused on collecting information on the state of knowledge about the decision-making in the maintaining and updating of pavement assets, along with pavement condition forecasting models and cost forecasting models. This effort integrated scholarship from such sources as the Transportation Research Board (TRB), Federal Highway Administration (FHWA), American Association of State Highway and Transportation Officials (AASHTO), and the American Society of Civil Engineers (ASCE) library.

Based on the literature review, the researchers conducted an in-depth analysis of pavement deterioration dynamics, leveraging empirical data from the GDOT's pavement inspection records spanning 2017–2021.The complex interplay of factors influencing pavement condition transitions were examined through empirical analysis using Markov chain pavement deterioration models. By identifying leading indicators and exploring the impacts of multiple

factors on pavement degradation, machine learning (ML)–based classification models were developed for enhanced pavement performance predictions. The findings offer valuable insights into the regional variations in pavement deterioration behavior and serve as a foundation for the development of probabilistic deterioration models and streamlined asset management strategies, enabling the GDOT to prioritize maintenance interventions and allocate resources more effectively.

The next initiative focuses on the challenging task of estimating maintenance project costs in the face of uncertain market conditions and evolving project requirements. Leveraging ML algorithms, particularly tree-based methods such as the extra trees (ET) algorithm, researchers formulated early-stage cost prediction models tailored to highway maintenance projects. By considering pavement characteristics and other relevant factors, the models aim to achieve high accuracy in cost estimation, facilitating long-term financial planning and resource allocation for the GDOT. Moreover, the exploration of tree-based algorithms highlights their practicality and applicability, offering insights into improving the robustness and efficiency of cost forecasting models.

The overarching goal of this project is to provide the tools, methodologies, and insights for the GDOT's effective pavement management. These approaches help enhance the GDOT's decision-making processes, optimize maintenance strategies, and ultimately ensure the resilience and sustainability of Georgia's transportation infrastructure. The findings and recommendations serve as a roadmap for the GDOT to develop cost-effective treatment plans, prioritize maintenance interventions, and uphold its commitment to safe and efficient transportation for all Georgians.

CHAPTER 1. INTRODUCTION

BACKGROUND

The Georgia Department of Transportation (GDOT) has adopted a comprehensive pavement management measure called the Overall Condition Index (OCI). It replaced the Computerized Pavements Condition Evaluation System (COPACES) in 2019. The OCI serves as a basis to trigger maintenance and rehabilitation. It is derived by averaging six distress indices—load cracking, edge cracking, block cracking, reflective cracking, rutting, and raveling—and an additional adjustment index score for asphalt pavements. This new measure takes advantage of the automated data collection method; however, the corresponding maintenance and rehabilitation criteria for pavements need to be studied and updated considering regional effects, cost variation, the effectiveness of different treatment options, and decision timing.

The GDOT is responsible for pavement management of 7100 centerline-mile, along with different terrain types. A “one size fits all” strategy no longer works because increasingly pavement condition depends on internal and external factors that differ from district to district (i.e., material, construction, environment). More specifically, variations in temperature and terrain, evolving traffic conditions and the distribution of vehicle types, anticipated growth in population, and economic activities all lead to different roadway deterioration paths. Therefore, it is crucial to capture and differentiate pavement deterioration rates at the district (or subdistrict) level. Historical pavement condition data can be utilized to identify critical regional factors and analyze their impacts on pavement performance in order to update the existing maintenance and rehabilitation criteria. In this research, the regional effects will be clearly defined and properly incorporated to enhance maintenance and rehabilitation decision-making.

In developing cost-effective maintenance plans, estimating the cost of maintenance and rehabilitation projects can be a challenging task, especially when the cost estimate is needed for a project anticipated in 4–7 years or more in the improvement program. Maintenance and rehabilitation project costs are subject to high variations over time. A major source of uncertainty is related to the construction market; for instance, the level of project cost depends on how busy the construction market is in the area of the project. Although consideration of the broader economic drivers of project cost, such as macroeconomic indicators (e.g., inflation rate and gross domestic product), major supply sources for critical materials, labor shortage issues, and energy market conditions (especially the trend in oil price), are also important, the actual contemplation of such indicators in the preliminary stage is difficult for cost prediction of a project farther down the road. The unalignment between short-term maintenance budgeting periods and long-term asset management plan periods inhibits effective maintenance programming. In this research, both temporal and spatial variations will be considered to characterize future scenarios for the cost of maintenance and rehabilitation projects.

This research aims to help the GDOT's Office of Maintenance better allocate funding and improve maintenance strategies at the network level through updating the decision tree considering temporal and spatial variation in cost, deterioration rate, and maintenance effectiveness.

OBJECTIVES

The major objective of this research is to enhance the GDOT's maintenance strategies through the following:

1. Empirically analyzing the pavement deterioration in different geographical locations and climate conditions across the state.

2. Improving life cycle cost estimates for the treatment options.

The enhanced decision analysis methods will provide the GDOT with comprehensive tools to better consider regional effects and cost variations in developing cost-effective treatment plans for statewide maintenance and rehabilitation projects.

CHAPTER 2. PAVEMENT DETERIORATION MODEL

BACKGROUND

State highway agencies face significant challenges in generating cost-efficient pavement maintenance plans for long-term periods due to inherent difficulties in forecasting pavement conditions. As infrastructure systems mature and the need to optimize the asset-performance-to-operating-cost ratio becomes more pronounced; the traditional "one size fits all" approach for estimating pavement conditions has proven inadequate. Although the conventional approach carried out maintenance and refurbishment work on fixed time intervals, pavement deterioration is not a uniform process but a dynamic process that depends heavily on the specific characteristics of individual roadways. Previous studies have investigated the potential of applying probabilistic models to be used in the asset management process (Black et al. 2005, Carer 2006, Fickler et al. 2009, Hu et al. 2022, Merigó et al. 2018). Yet, despite this progress, modeling the condition of pavements remains a complex task, compounded by factors such as uncertain environmental conditions, largely unknown operating history, and periodic availability of data. Previous work organized and illustrated pavement condition forecasting models and emphasized the need for identifying leading indicators, the factors that influence the pavement conditions, to improve the performance of probabilistic models (Justo-Silva et al. 2021, Kaloop et al. 2022, Pérez-Acebo et al. 2019, Shtayat et al. 2022).

Moreover, data-driven machine learning (ML) models have gained attention for predicting pavement conditions, leveraging historical data to enhance forecasting accuracy (Marcelino et al. 2021). ML is a branch within artificial intelligence that enables computers to learn and behave through input of data derived from observations. These sophisticated algorithms are capable of processing substantial amounts of data and resolving nonlinear

challenges to extract patterns, trends, and insights from the data (Bashar and Torres-Machi 2021). These ML models have been explored to draw relationships between the future state of pavement conditions and independent factors, such as pavement attributes, traffic volume, and weather conditions.

Most prior research has focused on developing forecasting models in regression matter, estimating the exact condition index values (Attoh-Okine 1994, Bashar and Torres-Machi 2021, Georgiou et al. 2018, Kargah-Ostadi and Stoffels 2015, Mazari and Rodriguez 2016, Yang et al. 2003). Yet, in practice, transportation asset management plans tend to direct maintenance projects based on ranges of pavement condition indices rather than their precise values. For example, the GDOT adopted a pavement management measure, OCI, to quantify the serviceability and condition for the predefined maintenance sections (Georgia Department of Transportation 2023). OCIs for asphalt pavements are calculated by integrating six distress measurements: load cracking, edge cracking, block cracking, reflective cracking, rutting, and raveling. Based on the pavement condition classes (in ranges of OCI) of the predefined maintenance sections, the GDOT's transportation asset management plan suggests a specific maintenance type: no treatment, lite treatment, minor preservation, major preservation, minor rehabilitation, or major rehabilitation. Then, the required maintenance expenditure for assets can be estimated based on the anticipated cost for each maintenance type. Therefore, making accurate predictions of condition classes is crucial to determine the correct maintenance type and allocate the appropriate time and budget.

This chapter comprises two portions. The researchers (1) present empirical analysis of differences in pavement condition transition likelihood, with the aim of enhancing the understanding of how degradation of pavement depends on multiple factors, and (2) develop a

pavement condition classification model using ML classifiers. The GDOT's pavement inspection records spanning 2017–2021 are used for the analyses. Empirical Markov chain pavement deterioration models are constructed to explore how transition probabilities of pavement condition are influenced by multiple factors. The findings of this research will provide insights into how different factors impact pavement deterioration behavior and can be utilized as backbone material when developing probabilistic deterioration models of the pavement condition. The deterioration classification models will facilitate streamlined asset management strategies by predicting pavement condition levels and determining the right maintenance treatment for transportation pavement assets. These more precise projections of conditions are able to promote timely and necessary maintenance interventions, possibly leading to a more economical asset management plan.

LITERATURE REVIEW

In recent years, there have been studies that covered pavement performance prediction models. Researchers have attempted to use traditional pavement deterioration models (e.g., linear regression [LR]) to forecast pavement condition indices, but they have found limitations in producing promising results due to the complex nature of the data (Flintsch and Chen 2004). Thus, more advanced ML algorithms were employed as an enhanced alternative to the traditional models.

In the transition from traditional statistical methods to ML, more attention is drawn to the improved accuracy and practicality of prediction models. Attoh-Okine (1994) demonstrated that employing an artificial neural network (ANN) for roughness prediction was feasible and could serve as the basis for developing a pavement deterioration process. Yang et al. (2003) predicted pavement condition for 1 to 5 years using an ANN and compared those predictions with

autoregressive models. The ANN model exhibited a higher R-squared (R^2) value compared to the autoregressive models, and the outperformance was more emphasized with longer forecasting intervals. Bashar and Torres-Machi (2021) also illustrated the capabilities of ML algorithms in pavement performance models that forecast international roughness index (IRI). The authors compared the correlation coefficients of three ML algorithms—ANN, random forest (RF), and support vector machine (SVM)—and traditional approaches including LR, multiple LR, quadratic LR, and others. The ML techniques showed better performance in that the models captured 15.6 percent more variability than those of traditional techniques, on average.

Kargah-Ostadi and Stoffels (2015) used Federal Highway Administration (FHWA) Long-Term Pavement Performance (LTPP) data for building and comparing pavement prediction models using an ANN, SVM, radial basis function (RBF), and the exponential form of the nonlinear regression (NLR) model. All three ML models showed higher performance than the regression model in quantitative analysis, which resulted in lower mean squared error (MSE) and standard deviation of the error. Qualitative evaluation depicted that the model using Bayesian regularization of an ANN best captured the measured values.

Mazari and Rodriguez (2016) used a hybrid technique that combined gene expression programming based on an ANN to predict the IRI. Georgiou et al. (2018) also predicted pavement roughness with an ANN and RBF kernel-based SVM. Data were collected annually for a high-volume motorway for 7 years. Both models were capable of predicting with high accuracy despite the variation of data. Ziari et al. (2016) employed SVM to predict IRI using nine variables from LTPP. Through 10-fold cross-validation, the results exhibited that Pearson VII universal kernel of SVM is capable of prediction in both the short and long terms.

The majority of previous studies have been conducted to develop prediction models in regression matter and evaluate the performance of models using R^2 and/or MSE. There is a gap of knowledge in employing ML classification models for predicting pavement condition levels. As transportation asset management plans are often structured based on the intervals of pavement condition indices instead of their exact values, classification models are suitable for determining the specific maintenance type in need for each asset, leading to a more streamlined development of asset management strategy. Additionally, previous studies demonstrated the potential of improving forecasting abilities within the traffic monitoring and project planning domain by integrating multiple ML models through formulation of ensemble models (Cao et al. 20182018, Zheng et al. 2019). However, there is a gap in research on utilizing ensemble approaches to enhance performance of pavement condition predictions. This study seeks to formulate pavement condition classification models utilizing ML classifiers and explore the benefits of employing ensemble techniques to enhance pavement performance predictions.

DATA DESCRIPTIONS

The empirical pavement condition scores of the GDOT's 3197 road segments recorded annually from 2017 to 2021 were used to analyze Markov chain transition probabilities and develop an ensemble ML classification model. The OCI ranges 0–100 to indicate the scores of road segments' pavement conditions (Georgia Department of Transportation 2023). It is calculated by accommodating asphalt distresses including load cracking, block cracking, edge distress, reflection cracking, raveling, and rutting. The GDOT's transportation asset management plan assigns a particular maintenance type for pavement assets, including no treatment, lite treatment, minor preservation, major preservation, minor rehabilitation, or major rehabilitation, according to the assets' condition expressed in ranges of OCI. The ranges of OCI for select maintenance

types were used as a reference to categorize the OCI into states of condition. The OCI values are categorized into six classes (states) based on the ranges of OCI that the GDOT's asset management plan typically uses to select type of maintenance. The OCI classes are provided in table 1. Moreover, figure 1 shows the number of data points of condition classes contained in each year.

Table 1. Pavement condition classes in ranges of OCI.

Overall Condition Index	Condition Class
100–90	A
90–80	B
80–70	C
70–60	D
60–50	E
50 and below	F

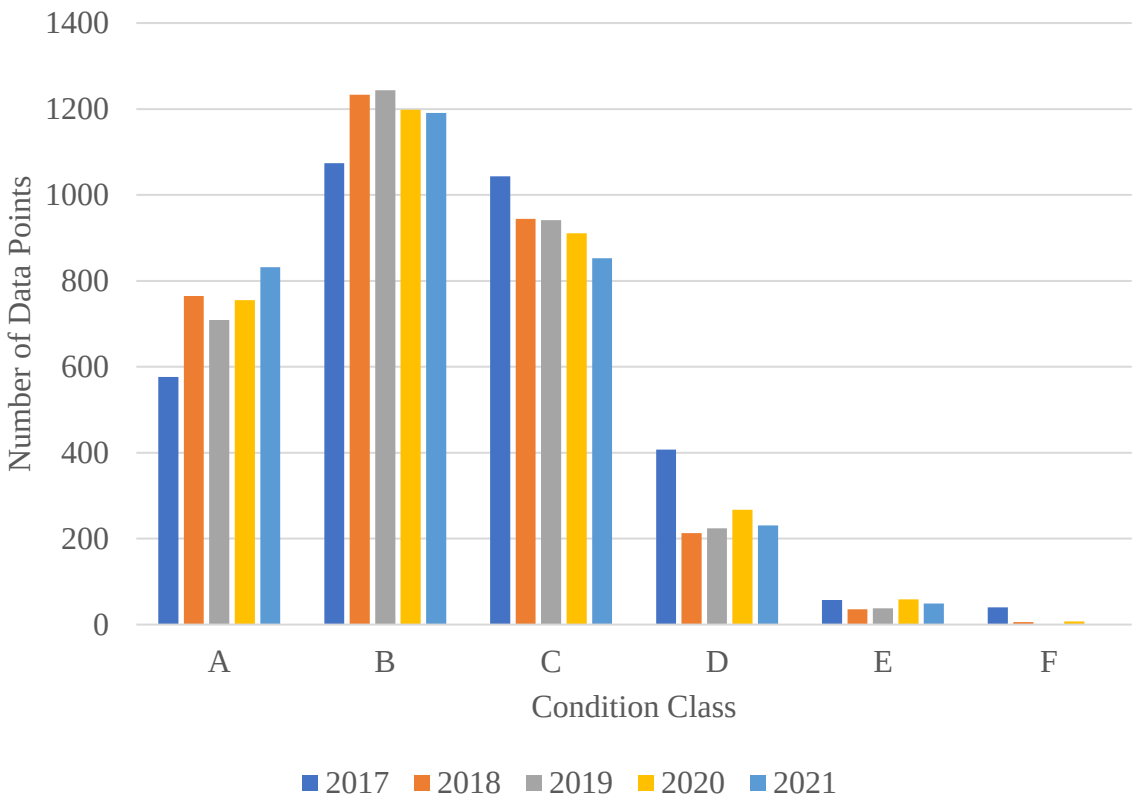


Figure 1. Chart. Number of data points of annual condition class.

Maintenance sections that have not undergone maintenance or rehabilitation were analyzed to demonstrate the ongoing deterioration relative to time. Yearly changes in pavement conditions, along with a set of variables from the GDOT's maintenance section's management system, were gathered to formulate the pavement deterioration models. Additionally, the precipitation and temperature information were added, as climate conditions affect the pavement deterioration (Titus-Glover et al. 2019). The climate conditions were gathered by matching Georgia county climate information collected from the National Weather Service. Descriptions of variables used in this study are provided in table 2.

Table 2. Variable descriptions.

Potential Variables	Units/Categories	Description
Overall Condition Index	Score	Numerical, ranging 20.28–100.00; mean = 81.92; standard deviation = 9.94
Average Annual Daily Traffic	Count	Numerical, ranging 3–288,954; mean = 11,670.84; standard deviation = 20,871.55
Number of Lanes	Categorical: {One, Two, Three, Four, Five, Six}	One: 1678 sections
		Two: 1307 sections
		Three: 158 sections
		Four: 11 sections
		Five: 5 sections
		Six: 1 section
District	Categorical: {District 1, District 2, District 3, District 4, District 5, District 6, District 7}	District 1: 410 sections
		District 2: 522 sections
		District 3: 609 sections
		District 4: 529 sections
		District 5: 557 sections
		District 6: 378 sections
		District 7: 192 sections
Interstate	Categorical: {Interstate, Non-Interstate}	Interstate: 187 sections
		Non-Interstate: 3010 sections
National Highway System (NHS)	Categorical: {NHS, Non-NHS}	NHS: 1541 sections
		Non-NHS: 1656 sections
Annual Precipitation	Inches	Numerical, ranging 45.25–78.49; mean = 55.68; standard deviation = 5.90
Average Temperature	Fahrenheit	Numerical, ranging 57.08–69.57; mean = 65.16; standard deviation = 2.72
Minimum Temperature	Fahrenheit	Numerical, ranging 45.97–59.35; mean = 54.07; standard deviation = 2.66
Maximum Temperature	Fahrenheit	Numerical, ranging 67.98–80.78; mean = 76.23; standard deviation = 2.85

NHS = National Highway System.

MARKOV CHAIN MODEL

Methodology

Markov chain analysis is a statistical technique used to model the behavior of a system in which the future state of the system depends only on the current state and not on any of the previous

states (Valles-Valles and Torres-Machi 2023). Markov chain models open the opportunity to predict the condition of pavements by modeling the transition probabilities of pavement states from one time period to the next. This study investigated the condition transition probabilities for pavements with varying attributes. By comparing the probabilities of degradation, the analysis aimed to identify which pavement categories are more susceptible to deterioration.

The GDOT's transportation asset management plan assigns a particular maintenance type for pavement assets, including no treatment, lite treatment, minor preservation, major preservation, minor rehabilitation, and major rehabilitation, according to their condition categories (expressed in ranges of OCI score). Subsequently, by referencing the expected cost associated with each maintenance type, practitioners project the necessary maintenance expenditure for assets. Hence, understanding how the pavement deterioration probability differs depending on pavement characteristics is important to ensure selection of the correct maintenance type and optimally allocation of resources to maintain assets. Given that OCI scores are documented annually, the Markov chain model was employed to analyze condition transition probabilities, facilitating comprehension of deterioration patterns over discrete time intervals.

The Markov chain transition probabilities are typically expressed as the matrix in equation 1. Each element represents the probability of transitioning from the initial state to another. The probabilities are calculated by the ratio between transition observations from the initial state to a specific state and all of the observations from the initial state. Therefore, the probabilities in each column will sum up to 1.

$$\begin{bmatrix} p_{AA} & 0 & 0 & 0 & 0 & 0 \\ p_{AB} & p_{BB} & 0 & 0 & 0 & 0 \\ p_{AC} & p_{BC} & p_{CC} & 0 & 0 & 0 \\ p_{AD} & p_{BD} & p_{CD} & p_{DD} & 0 & 0 \\ p_{AE} & p_{BE} & p_{CE} & p_{DE} & p_{EE} & 0 \\ p_{AF} & p_{BF} & p_{CF} & p_{DF} & p_{EF} & p_{FF} \end{bmatrix} \quad (1)$$

This study investigated how three different properties of Markov chain model differ based on the following variables: Interstate, National Highway System (NHS), Average Annual Daily Traffic (AADT), Number of Lanes, District, and Annual Average Temperature.

First, the transition probabilities were calculated under different groups of maintenance sections, such as interstate or non-interstate, NHS or non-NHS, traffic volume, number of lanes, districts, and temperature levels. For a better visualization, the transition probabilities were presented in Markov chain model diagrams. Probabilities that could not be calculated due to lack of number of observations are expressed as "N/A".

Next, degradation probabilities for each pavement group were compared to determine which group is more prone to deterioration. The probability of degradation was calculated by determining the ratio of the number of observations of transition to a lower grade to the total number of observations. This total includes both the observations of transition to a lower grade and the cases where the grade remained the same. The formula used to calculate the degradation probability for each group is as follows:

$$p_{degrade} = \frac{number\ of\ transitions\ to\ lower\ grade}{total\ number\ of\ observations} \quad (2)$$

In addition, the chi-squared (χ^2) test was employed to determine whether the distribution of pavement condition transition frequencies differs depending on pavement characteristics

(Washington et al. 2011). The chi-squared test statistic is calculated by summing the squared differences between the observed and expected frequencies, divided by the expected frequencies for k number of each cell in the table, as follows:

$$\chi^2 = \frac{\sum_{i=1}^{k}(O_i - E_i)^2}{E_i} \tag{3}$$

where χ^2 is the chi-squared test statistic, O_i is the observed count in cell i, and E_i is the expected count in cell i. This sum is then compared to the chi-squared distribution with $(\text{number of rows} - 1) \cdot (\text{number of columns} - 1)$ degrees of freedom to obtain a *P*-value. In this study, the null hypothesis was that transitions in pavement conditions are independent of the specific pavement characteristics under consideration. Conversely, the alternative hypothesis suggests that transitions in pavement conditions are dependent on the characteristics of the pavement. The 0.05 level of significance was used to reject the null hypothesis.

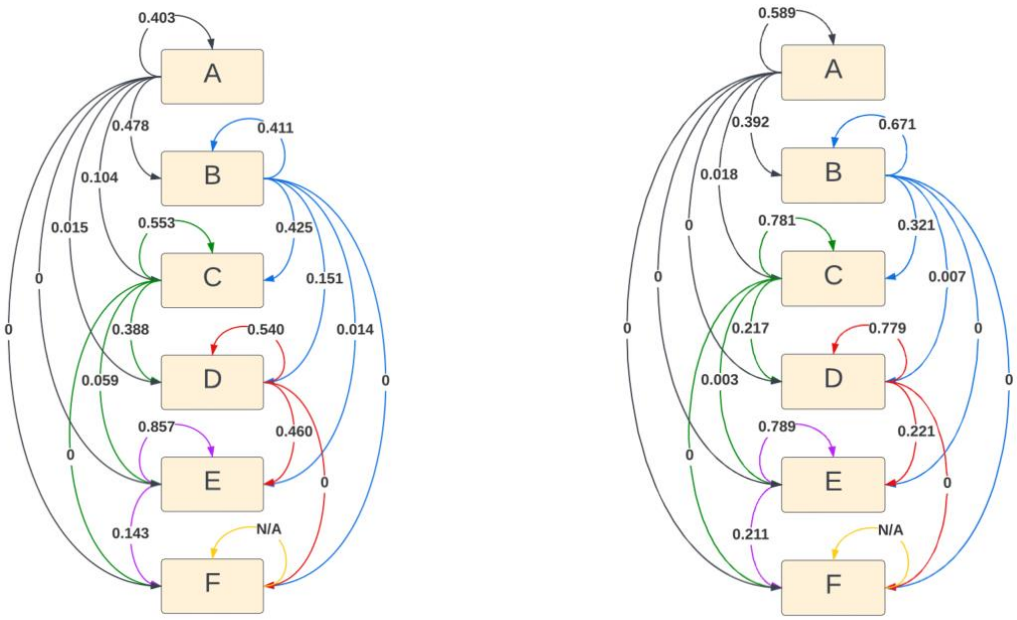
One of the assumptions of the chi-squared test is that the expected frequency for each cell should be 5 or higher in at least 80 percent of the cells (Bewick et al. 2003). This condition is more likely to be satisfied if the sample size is at least five times the number of cells. However, as shown in the Markov chain diagrams (see figure 2, figure 3, figure 4, figure 5, figure 6, and figure 7), the majority of the probabilities are concentrated either on remaining in the same state or deteriorating by one grade. This results in low frequencies of transitioning to two or more grades lower, potentially violating the previously mentioned chi-squared test assumption. Consequently, this study concentrates on where most observations occur, i.e., cases of staying at the same state and deteriorating by one grade, and aims to determine if the frequency of staying in the same state or deteriorating by one grade varies based on the characteristics of pavements.

Markov Chain Model Results

The Markov chain model diagrams, showing degrading transition probabilities based on each variable value, and the pairwise chi-squared test results are presented in the following subsections.

Interstate Variable

Figure 2 presents the Markov chain model for the Interstate and Non-Interstate groups of maintenance sections. The direction of arrows indicates the direction of transition from the initial state to the next states, and the numbers represent the probabilities of the corresponding transition. Table 3 shows the annual degrading probabilities of Interstate and Non-Interstate sections. Table 4 shows the chi-squared test results of Interstate and Non-Interstate groups of road segments. It is found that the degrading probability for Interstate exceeds that of Non-Interstate by more than 18 percent. Furthermore, the chi-squared test results indicate dependency between the frequency of pavement condition transitions and the Interstate variable, suggesting Interstate and Non-Interstate have significantly different distributions of transition probabilities. Interstate roads tend to carry more traffic than Non-Interstate roads. This means that the pavement is subjected to more wear and tear, which can cause it to deteriorate faster. Also, Interstate roads are often used by heavy vehicles, such as trucks and buses. These vehicles cause more damage to the pavement than lighter vehicles, which can contribute to faster deterioration. Lastly, the higher permissible speeds and the increased occurrence of abrupt braking by users to rapidly reduce their speed in emergency situations could be additional causes for Interstate roads to degrade faster than Non-Interstate roads.



(a) Interstate (b) Non-Interstate

Figure 2. Diagrams. Markov chain transition probabilities – Interstate.

Table 3. Degrading probabilities for Interstate variable.

Interstate Categories	Degrading Probability
Interstate	0.502
Non-Interstate	0.317

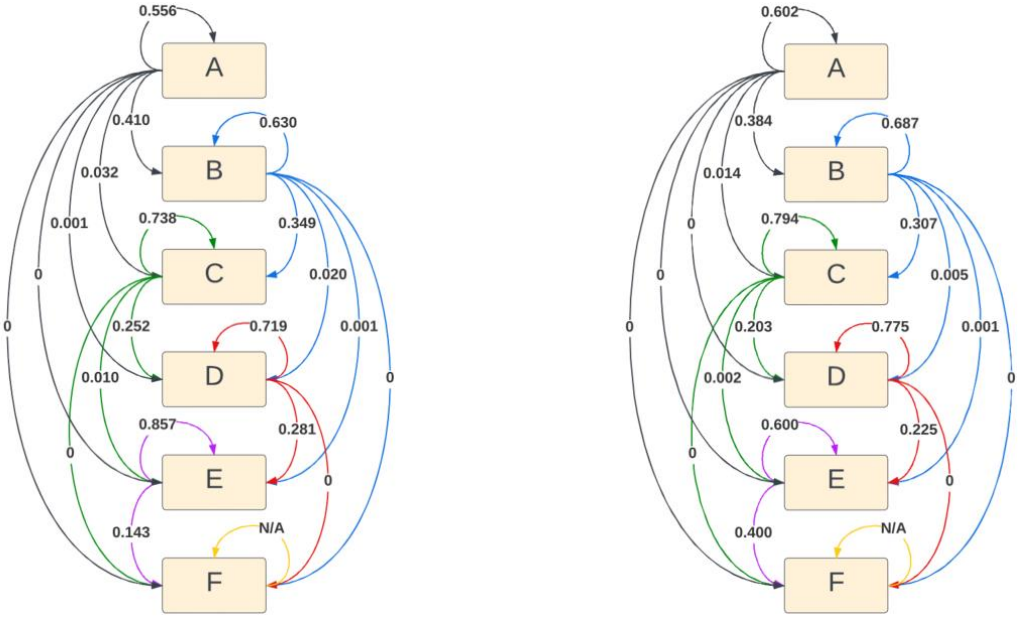
Table 4. Chi-squared test for Interstate variable.

Group 1	Group 2	*P*-value
Interstate	Non-Interstate	<0.05

National Highway System Variable

The NHS refers to roadways important to the nation's economy, defense, and mobility established by the U.S. Department of Transportation (USDOT). Figure 3 presents the Markov chain model for the NHS and Non-NHS groups of maintenance sections. Table 5 shows the

annual degrading transition probabilities of NHS and Non-NHS road segments. Table 6 shows the chi-squared test results of NHS and Non-NHS road segments. The degrading probability is slightly higher for NHS roads than for Non-NHS roads. Furthermore, the chi-squared test rejects the null hypothesis, indicating that the frequency of pavement condition transitions is dependent on the NHS variable. This result suggests that the distributions of transition probabilities are different between NHS and Non-NHS groups.



(a) NHS (b) Non-NHS

Figure 3. Diagrams. Markov chain transition probabilities – NHS.

Table 5. Degrading probabilities for NHS variable.

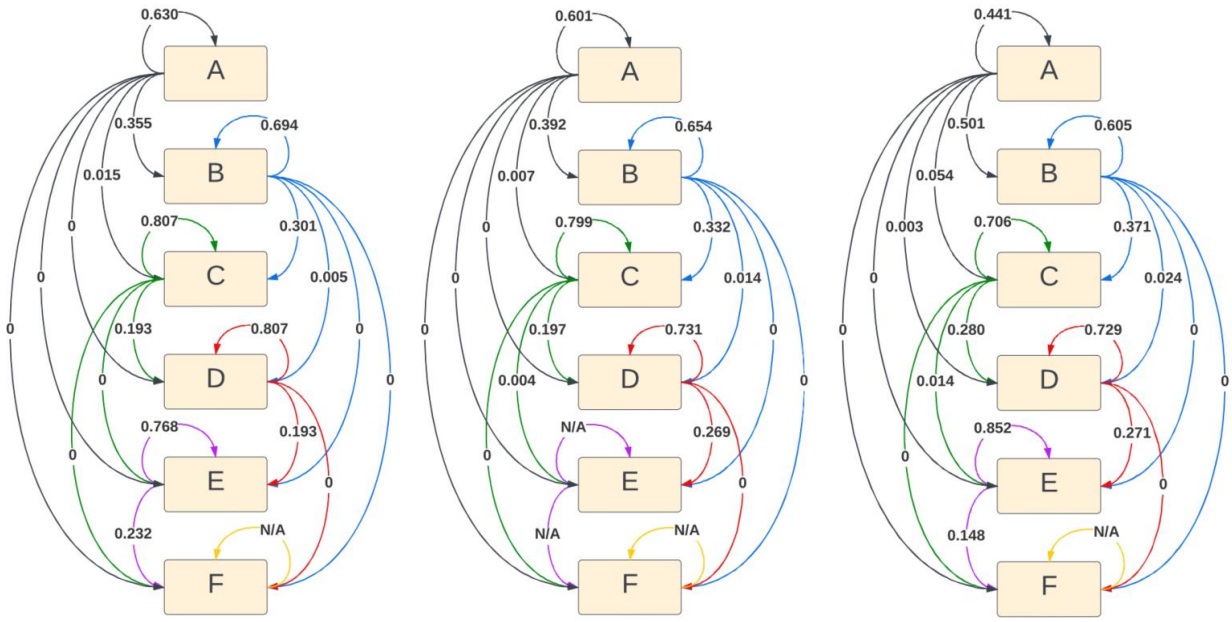
NHS Categories	Degrading Probability
NHS	0.351
Non-NHS	0.305

Table 6. Chi-squared test for NHS variable.

Group 1	Group 2	*P*-value
NHS	Non-NHS	<0.05

Annual Average Daily Traffic Variable

The AADT is categorized based on thresholds as follows: Low = ≤5000, Medium = 5000–10,000, and High = ≥10,000 (Tsai et al. 2010). Figure 4 presents the Markov chain model for the Low, Medium, and High AADT groups of maintenance sections. Table 7 shows the degrading probabilities of road segments with different AADT categories. Table 8 shows pairwise chi-squared test results for different AADT category groups. It is apparent that the degrading probability increases as the AADT increases. The *P*-values from chi-squared tests show that the frequencies of pavement condition transitions are significantly influenced by the groups with Low and Medium AADT and Low and High AADT. Conversely, the transitions between the Medium and High AADT categories appear to be independent. These results suggest that the transition probabilities of pavement condition for sections with Low AADT are statistically different from those with Medium and High AADT, whereas the transition probabilities between Medium and High AADT sections do not show significant statistical differences. The pairwise degrading probability differences are less than 10 percent, which is not as significant as the Interstate variable.



(a) Low (b) Medium (c) High

Figure 4. Diagrams. Markov chain transition probabilities – AADT.

Table 7. Degrading probabilities for AADT variable.

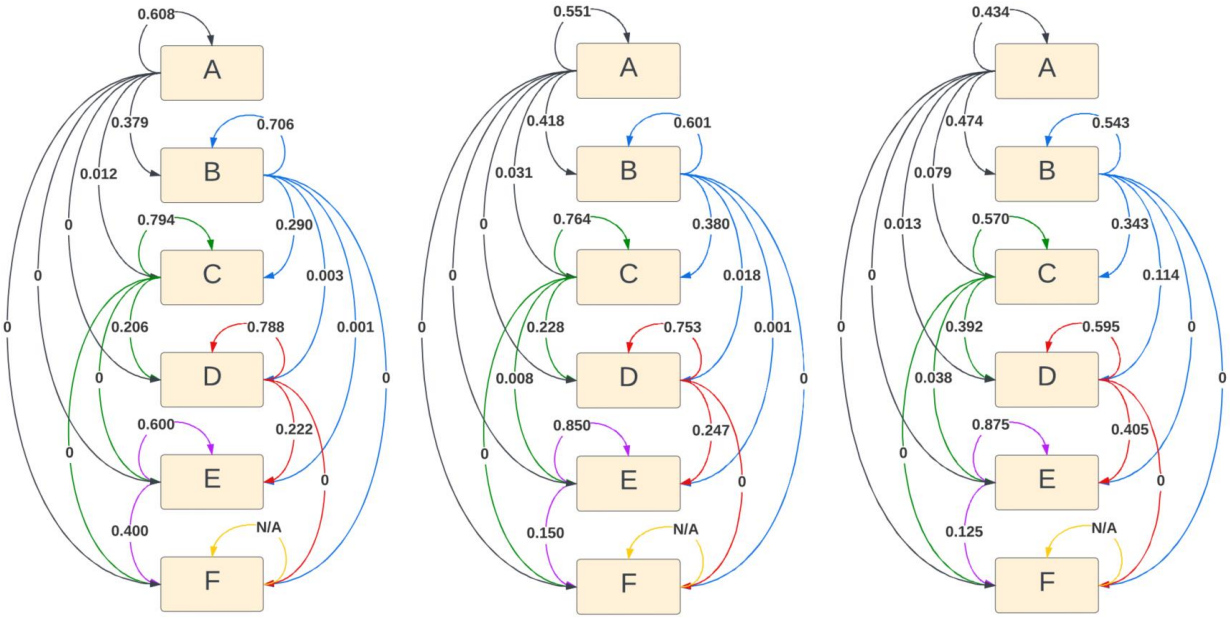
AADT Categories	Degrading Probability
Low	0.298
Medium	0.319
High	0.380

Table 8. Chi-squared test for AADT variable.

Group 1	Group 2	*P*-value
Low	Medium	<0.05
Low	High	<0.05
Medium	High	0.32

Number of Lanes Variable

The effect of the number of lanes on pavement deterioration also has been investigated. As shown in table 2, the Number of Lanes variable has six categories. Due to the lack of sample size for the Four, Five, and Six categories, this study concentrated on evaluating the One, Two, and Three categories. Figure 5 presents the Markov chain model for the One, Two, and Three Number of Lanes groups of maintenance sections. Table 9 shows the degrading probabilities of road segments with different Number of Lanes categories. Table 10 shows pairwise chi-squared test results for different Number of Lanes category groups. It is found that the degrading probability increases as the number of lanes increases. The maintenance sections with three lanes exhibit a 16.3 and 11.2 percent higher annual probability of degrading compared to sections with two lanes and one lane, respectively. The observation that roads with more lanes deteriorate faster is due to their design to accommodate higher traffic volumes, which subjects the pavement to increased wear and tear. Furthermore, the chi-squared test results indicate dependency between the frequency of pavement condition transitions and the Number of Lanes variable, suggesting sections with one, two, and three lanes have significantly different distributions of transition probabilities.



(a) One (b) Two (c) Three

Figure 5. Diagrams. Markov chain transition probabilities – Number of Lanes.

Table 9. Degrading probabilities for Number of Lanes variable.

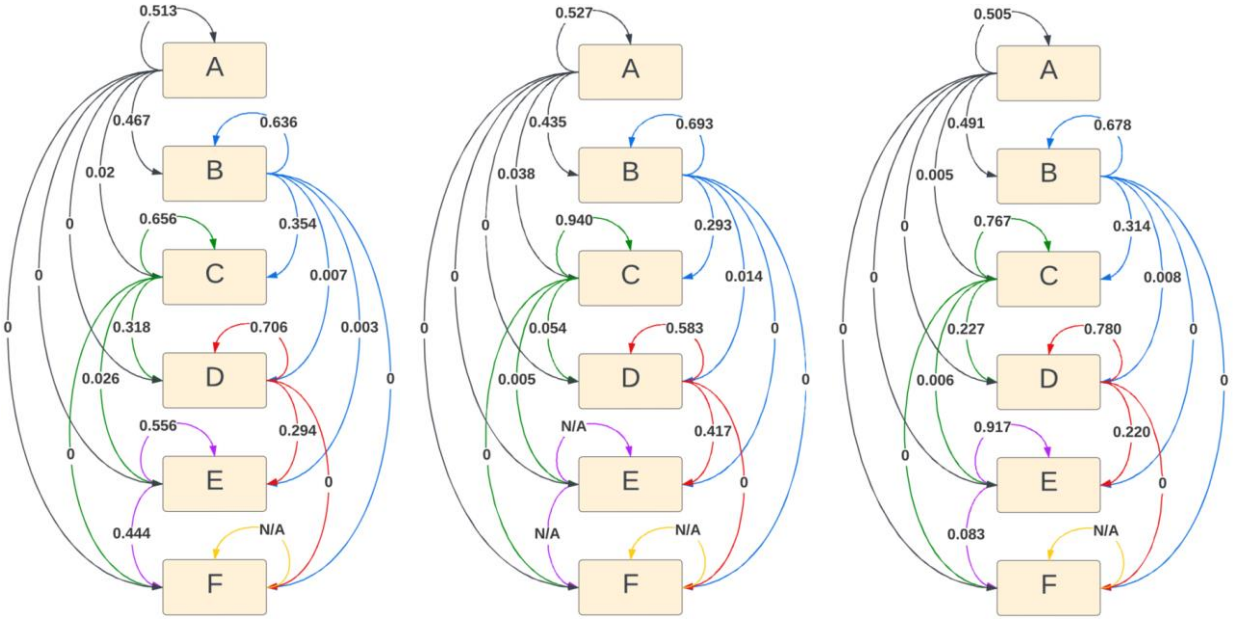
Number of Lanes Categories	Degrading Probability
One	0.299
Two	0.350
Three	0.462

Table 10. Chi-squared test for Number of Lanes variable.

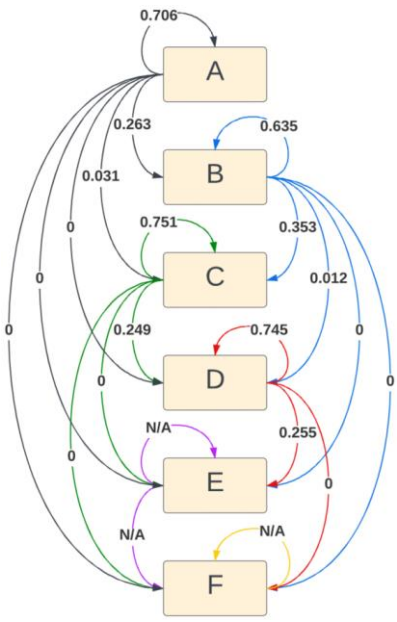
Group 1	Group 2	*P*-value
One	Two	<0.05
One	Three	<0.05
Two	Three	<0.05

District Variable

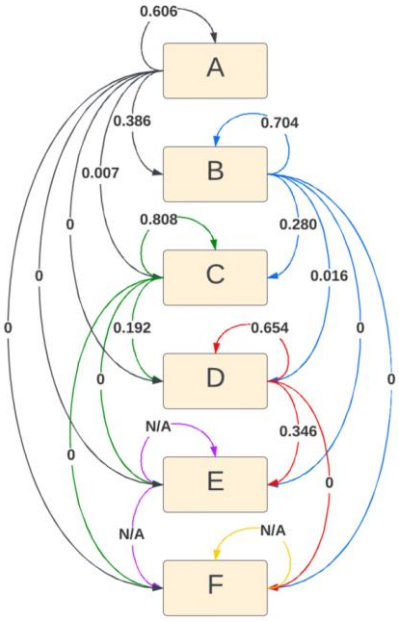
Figure 6 presents the Markov chain model for the maintenance sections located in different districts. Table 11 shows the annual degrading probabilities of each district's maintenance sections. With 7 districts, 21 pairwise chi-squared tests were performed. Among the 21 tests, 3 pairwise tests failed to reject the null hypothesis, and the remaining pairwise tests rejected the null hypothesis. Table 12 only shows pairwise chi-squared tests that failed to reject the null hypothesis. The majority of pairwise chi-square tests reveal statistically significant variances in pavement condition transitions among districts.



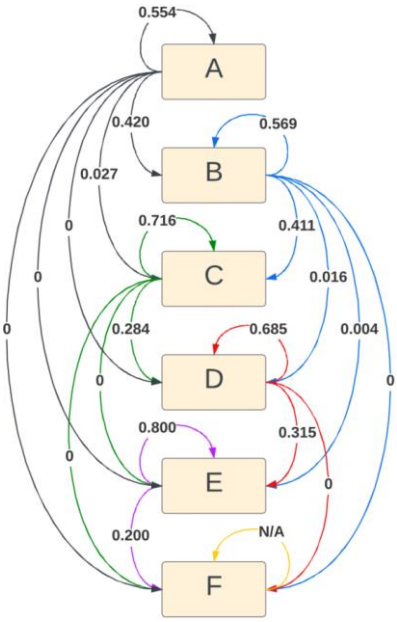
(a) District 1 (b) District 2 (c) District 3



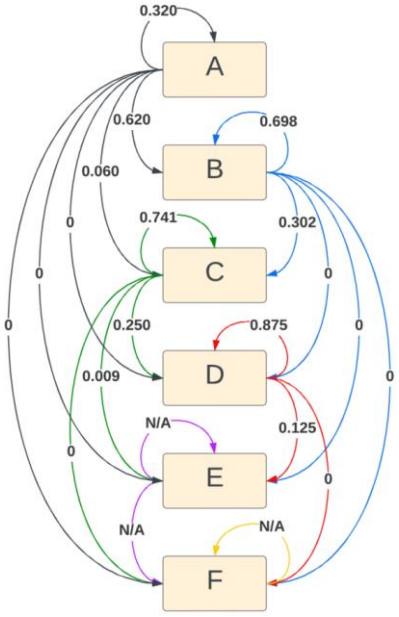
(d) District 4



(e) District 5



(f) District 6



(g) District 7

Figure 6. Diagrams. Markov chain transition probabilities – District.

Table 11. Degrading probabilities for District variable.

Number of Lanes Categories	Degrading Probability
District 1	0.381
District 2	0.306
District 3	0.318
District 4	0.310
District 5	0.311
District 6	0.371
District 7	0.321

Table 12. Chi-squared test for Number of Lanes variable.

Group 1	Group 2	*P*-value
District 1	District 6	0.18
District 2	District 3	0.09
District 4	District 5	0.25

Combination of Interstate and Annual Average Temperature

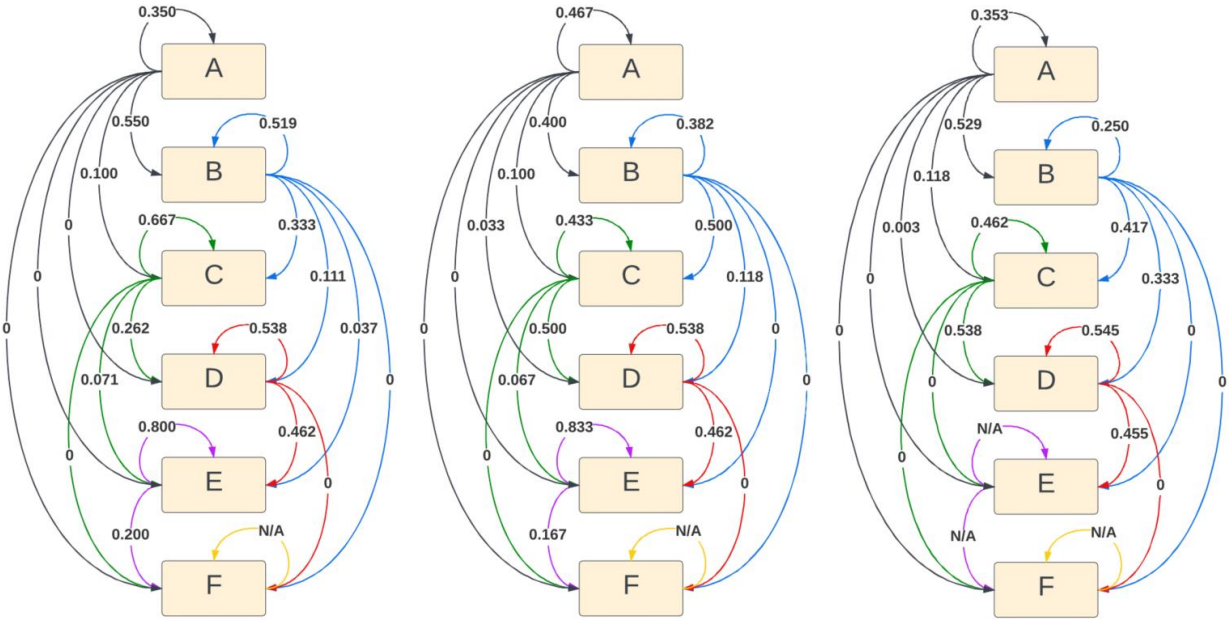
This study also investigated the influence of annual average temperature on the probabilities of pavement degradation. The geographical variation in temperature across Georgia was considered, with higher average temperatures observed in South Georgia compared to North Georgia. Depending solely on average temperature can lead to segmenting road sections by region. This segmentation can introduce bias into the analysis results, as transportation usage varies across different population centers throughout the state. To mitigate this bias, it was deemed more appropriate to assess the effect of average temperature in conjunction with another variable that would lessen partitioning by regions. Consequently, the impact of temperature was examined separately for Interstate and Non-Interstate maintenance sections, which was found to have the largest difference in degrading probabilities. However, due to the lack of number of observations for the Interstate group to meet the chi-squared test's assumption, only the Non-

Interstate group has been analyzed. Because of the absence of established thresholds for defining temperature effects on pavement deterioration, this study established thresholds based on the 25th and 75th percentiles and categorized the average temperature as follows:

- Low: ≤63.03°F (<25th percentile)
- Medium: 63.03°–67.43°F (25th–75th percentile)
- High: ≥67.43°F (>75 percentile)

Error! Reference source not found. presents the Markov chain model for the Low, Medium, and High Annual Average Temperature groups of Non-Interstate maintenance sections. **Error! Reference source not found.** shows the annual degrading transition probabilities of Non-Interstate sections with different Annual Average Temperature categories. **Error! Reference source not found.** shows pairwise chi-squared test results for Non-Interstate pavements with different Annual Average Temperature groups. The degrading probabilities for Non-Interstate sections with low average temperature decrease are found to be higher than those with medium and high average temperatures.

All pairs of chi-squared test results indicate dependency between the frequency of pavement condition transitions and the Annual Average Temperature variable, suggesting that the distributions of degrading probabilities are different among the temperature categories.



(a) Low (b) Medium (c) High

Figure 7. Diagrams. Markov chain transition probabilities – Annual Average Temperature (Non-Interstate).

Table 13. Degrading probabilities for Annual Average Temperature variable (Non-Interstate).

Average Temperature Categories (Non-Interstate)	Degrading Probability
Low	0.361
Medium	0.304
High	0.308

Table 14. Chi-squared test for Annual Average Temperature variable (Non-Interstate).

Group 1	Group 2	*P*-value
Low	Medium	<0.05
Low	High	<0.05
Medium	High	<0.05

ENSEMBLE MACHINE LEARNING MODEL

Methodology

This research analyzed performances of supervised ML classification models and ensemble models in predicting pavement condition classes. Five classification models, including RF, gradient boosting (GB), SVM, *k*-nearest neighbors (KNN), and ANN, were investigated in this study based on their verified performance in previous studies. Furthermore, two ensemble methods, voting and stacking, were employed to improve the prediction performance. The flowchart of research methods is provided in figure 8.

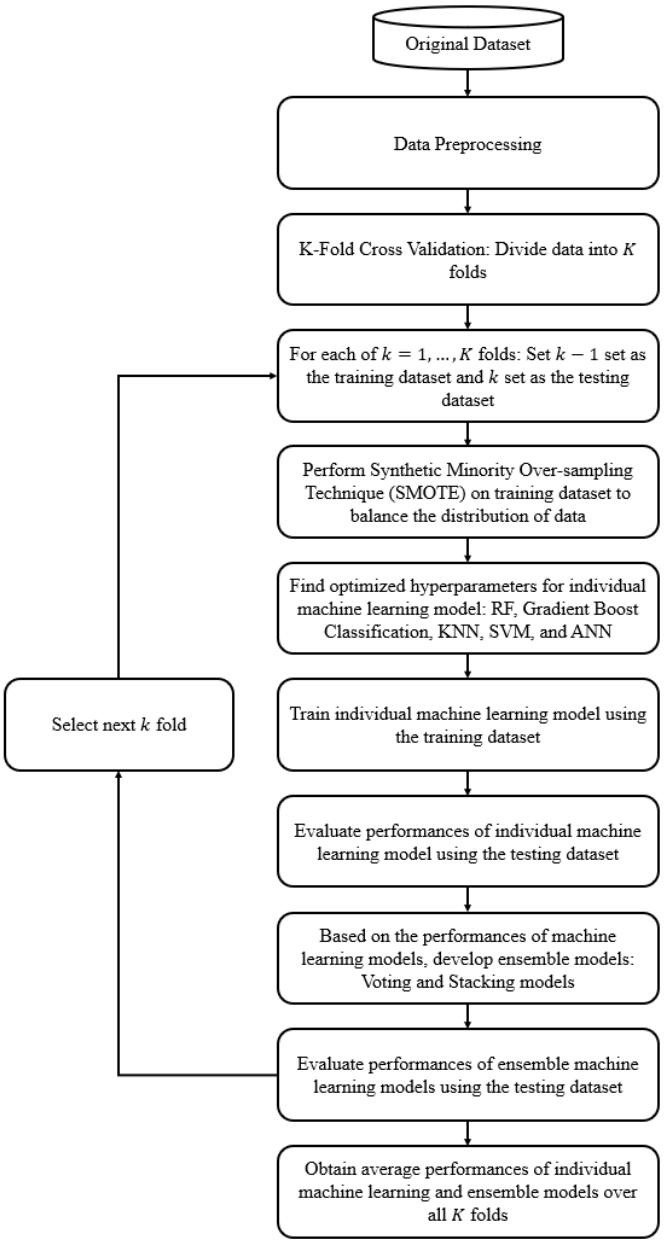


Figure 8. Flowchart. Research methodology (RF = random forest, KNN = *k*-nearest neighbors, SVM = support vector machine, ANN = artificial neural network).

As will be discussed in Machine Learning Algorithm and Ensemble Model Results, the analysis results suggest that it is challenging to definitively determine which classification model described above outperforms the others in predicting pavement condition classes. To ensure enhanced predictive accuracy, this research employed ensemble methods, specifically the voting

and stacking approaches, to integrate multiple classification models for generating predictions. Selecting a response based on a combination of models can mitigate the risk of choosing from an unfavorable hypothesis (Dietterich 2000).

Machine Learning Algorithms

Random Forest

RF is a ML algorithm that capitalizes on the strengths of multiple decision trees for improved accuracy and robustness. RF seeks to make a collective decision based on the outputs of multiple decision trees to ensure a higher accuracy and lower risk of overfitting compared to just one tree (Cano-Ortiz et al. 2022). The technique uses bootstrap aggregating, commonly known as "bagging." In this process, for each of the trees in the forest, a random sample of the training data is drawn with replacement, ensuring that every tree gets trained on a varied dataset (Bashar and Torres-Machi 2021).

The decision tree uses least squares deviation as its splitting criterion, which serves to measure the impurity within a dataset (Breiman et al. 1984). This model categorizes pavement condition classes by analyzing similar patterns through data segmentation using various branches based on input factors. It employs specific "if-then" rules to assign the pavement conditions to predefined classes and to predict the condition ratings. A tree structure is developed from the training dataset through recursive partitioning based on the least squares deviation values until the node satisfies certain conditions and becomes a leaf node. A lower least squares deviation in a node indicates greater sample purity. Therefore, the attribute splitting aims to minimize the average least squares deviation across child nodes, with each leaf node's prediction being the average of its sample values. Stopping conditions such as maximum tree depth, minimum sample

count per leaf node, and minimum decrease in impurity are determined via tuning hyperparameters to avoid overfitting.

Each decision tree in the RF operates on a unique sample drawn with replacement from the initial dataset, ensuring that each tree encounters different aspects of the data during training. This variation among trees helps to average out biases and variances, promoting a more balanced model. When it comes to making predictions, RF takes a democratic approach. For classification tasks, it aggregates the predictions from all the trees and selects the outcome based on majority voting. Thus, if a majority of trees predict a certain class, RF will output that class as the model's prediction.

Gradient Boosting Classification

GB refines predictive accuracy through an iterative method that systematically reduces errors by adding a series of weak learners, typically using the gradient descent approach. This technique builds upon the foundation laid by decision trees, which serve as the weak learners. Each tree in the sequence is tasked with selecting optimal split points based on metrics such as Gini impurity or minimal loss, generating real-valued outputs at each division (Friedman 2001).

The process begins with an initial weak model that evaluates the dataset and provides preliminary predictions. The residuals from these predictions are then leveraged to develop subsequent models. Each new model focuses on amending the inaccuracies identified in its predecessor's performance. To enhance the precision of future predictions, the algorithm assigns increased weights to instances that were previously failed to predict, boosting their likelihood of correct classification in subsequent iterations (Zhang et al. 2020). This strategic addition of learners continues until the overall error of the model can no longer be significantly reduced.

K-Nearest Neighbor

The KNN algorithm is a nonparametric, instance-based learning approach that is widely used in classification tasks, although it can also be effectively applied to regression problems (Inkoom et al. 2019). This algorithm operates on a straightforward premise: it classifies new instances based on a plurality vote of its neighbors, with the instance being assigned to the class most common among its nearest neighbors, as determined by a selected number of K neighbors.

KNN makes predictions by calculating the distances between a query instance and all the instances in the training dataset, then selecting the closest K instances to determine the most likely category (Cano-Ortiz et al. 2022). The distances can be measured using various metrics, the most common being Euclidean, Manhattan, and Minkowski distances. After computing these distances, KNN ranks them and predicts the class based on the prevalence of each class within the closest K instances.

One of the primary advantages of the KNN algorithm is its simplicity and effectiveness in handling multi-class cases. Its nonparametric nature means it does not assume anything about the underlying data distribution, which can be particularly advantageous in situations where the actual distribution is unknown and difficult to model.

Support Vector Machine

SVM operates by constructing a hyperplane or a set of hyperplanes in a high-dimensional space to segregate different classes with the widest possible margin between them. The primary objective of SVM is to maximize this margin, enhancing the model's ability to generalize well to new unseen data (Ziari et al. 2016).

SVM is effective in situations where the boundary between classes is not linearly separable by employing the kernel, a technique that transforms the original data into a higher-

dimensional space where a linear separator is viable. This transformation occurs using kernel functions such as polynomial, RBF, and sigmoid, which help adapt the SVM to various data complexities without explicit high-dimensional computations (Georgiou et al. 2018). SVM requires careful tuning of the kernel and its parameters, such as the regularization parameter and kernel-specific parameters. The performance of SVM is highly sensitive to these settings, often necessitating extensive grid search and cross-validation to optimize.

Artificial Neural Network

ANNs are adept at discerning patterns and relationships within datasets by mapping inputs to their corresponding outputs. An ANN draws inspiration from the structure and functioning of the nervous system of the human brain, consisting of a series of interconnected neurons organized into layers.

In an ANN, neurons are organized into three main layers: the input layer, which receives the initial data; one or more hidden layers, where the actual processing is done through weighted connections; and the output layer, which delivers the final decision or prediction (Kargah-Ostadi and Stoffels 2015). Data are processed in an ANN by passing signals from the input layer through the hidden layers to the output layer. Each neuron in these layers applies a mathematical activation function to the incoming signals, thus enabling the network to handle complex nonlinear relationships that simpler models might miss.

The connections between neurons carry weights that signify the strength or importance of the connection. These weights are fine-tuned during the training process to minimize the error between the predictions of the network and the actual data outputs. They are set randomly at first but are adjusted iteratively through a process known as backpropagation. Backpropagation is a systematic method of training an ANN by adjusting weights in reverse order from the output

toward the input layer, significantly reducing error by moving in the direction of the steepest gradient of the error surface (Kargah-Ostadi and Stoffels 2015).

Ensemble Models

Voting Model

Voting is an ensemble technique used in ML to enhance the predictive performance by combining the strengths of diverse learning algorithms. By aggregating the results of multiple models, the voting ensemble method aims to achieve higher accuracy and stability than any single model could on its own (Gandhi and Pandey 2016). This method is particularly effective in reducing the likelihood of overfitting by blending the predictions from multiple models.

There are two principal approaches to voting: hard voting and soft voting. Hard voting, also known as majority voting, simply counts the votes of each classifier in the ensemble and assigns the final prediction based on the class that gets the majority of the votes. Each model gets one vote per instance, and the class with the most votes becomes the prediction of the ensemble. This method is straightforward and often effective, particularly when combining models that have different error characteristics.

On the other hand, soft voting offers a more nuanced approach by considering the confidence level of the predictions made by each individual model. In soft voting, each model estimates probabilities for each class, and these probabilities are then averaged out. The final prediction is the class with the highest average probability. This approach leverages the probability estimates of the individual classifiers, making it generally more flexible and typically resulting in higher performance than hard voting, as it considers not just the final class outcomes but the underlying confidence of the classifiers (Kumari et al. 2021).

The advantage of soft voting is particularly pronounced in scenarios where models are well calibrated, meaning their probability estimates are reliable. Soft voting can effectively harness the strengths of both high-bias and low-variance models within the ensemble, smoothing out individual predictions and reducing the overall error rate.

Stacking Model

Stacking is an ensemble learning method that enhances prediction accuracy by combining the strengths of multiple base ML models. This technique involves a two-level structure: the first level consists of several base estimators, and the meta-level utilizes a meta-model that processes the outputs from the base level as its inputs to produce the final prediction (Wolpert 1992).

The process begins by training a variety of base models on the same dataset. The five ML models introduced in Machine Learning Algorithms are selected as the base models for their unique abilities to capture different characteristics of the data. The predictions from the base models are subsequently used as inputs for the meta-model. The meta-model, such as logistic regression, is tasked with learning the optimal way to amalgamate the predictions from the base models. This integration helps to correct any biases or inaccuracies presented by individual base models, thereby improving the overall accuracy. Essentially, the meta-model serves as an arbitrator, synthesizing various patterns and strengths from the base models into a cohesive output.

The advantage of stacking is its capacity to leverage the distinct capabilities of each model within the ensemble. For example, one model may excel at identifying specific data patterns, and another may be better suited to dealing with noise or anomalies. The meta-model is strategically employed to balance these varying strengths and weaknesses effectively, optimizing the overall predictive power.

Data Preprocessing

Data preprocessing was performed to convert categorical variables into numeric form and scale the continuous variables. It is essential to transform categorical attributes into numerical values to enable arithmetic operations during model training (Kim and Hong 2017). This transformation was achieved by randomly allocating distinct numerical identifiers to each category before initiating the model training. Additionally, the continuous variables have been scaled into the range of 0 and 1 based on each variable's minimum and maximum values. Scaling continuous variables enhances the speed of the ML training process (Kotsiantis et al. 2007).

Then, 5-fold cross-validation was performed to select hyperparameters for each classification model and evaluate performances of models. The cross-validation technique helps to prevent the overfitting problem and diminish bias that might arise from arbitrary data splits (Berrar 2019, Zhang et al. 2020). In *k*-fold cross-validation, the training set is partitioned into *k* equally sized subsets. During each iteration, one subset is used as the testing set, and the other $k-1$ subsets form the training set. This ensures that every data point is tested exactly once and is part of the training set $k-1$ times.

Finally, for each training set, Synthetic Minority Over-sampling Technique (SMOTE) was applied on imbalanced pavement condition data. SMOTE is used to address the issue of class imbalance in ML datasets, which is particularly effective when the data are skewed toward one class more heavily than another (Fernández et al. 2018). This imbalance often leads to models that are biased toward the majority class, potentially misclassifying the minority class. SMOTE works by creating synthetic samples rather than replicating existing samples. This approach helps to create a more balanced dataset, which enables ML algorithms to perform more equitably across all classes.

SMOTE operates by selecting examples that are close in the feature space, drawing a line between the examples in the feature space and generating new examples along that line (Fernández et al. 2018). This technique enhances the representation of the minority class in the training set without causing significant information loss, in contrast to under-sampling of the majority class. By enriching the dataset in this manner, SMOTE allows ML models to learn more about the minority class's characteristics, leading to better detection and classification of this class in new unseen data.

Machine Learning Algorithm and Ensemble Model Results

To evaluate the performance results, the accuracy metric was measured with the consideration that every class is equally important. The accuracy is measured as follows:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Numer of Predictions}} \quad (4)$$

The accuracy metric was measured for every classification model using the test set derived from every cross-validation set. Subsequently, four of each voting and stacking ensemble classification models were constructed and analyzed: the first using the two best-performing models, the second using the top three, the third using the top four, and the last ensemble integrating all five of the classification models. Table 15 shows the accuracy metrics of individual and ensemble classification models obtained by taking the average of the results retrieved from the cross-validation sets.

Table 15. Accuracy metrics of classification models.

Classification Model	Accuracy (mean ± σ)
Random Forest	0.768 ± 0.037
Gradient Boosting	0.776 ± 0.026
Support Vector Machine	0.778 ± 0.024
K-Nearest Neighbor	0.774 ± 0.023
Neural Network	0.724 ± 0.034
Voting with 2 Models	0.830 ± 0.015
Stacking with 2 Models	0.823 ± 0.016
Voting with 3 Models	0.814 ± 0.011
Stacking with 3 Models	0.811 ± 0.025
Voting with 4 Models	0.807 ± 0.012
Stacking with 4 Models	0.810 ± 0.018
Voting with 5 Models	0.791 ± 0.015
Stacking with 5 Models	0.815 ± 0.027

Among the individual classification models, the SVM model was found to have the highest accuracy, followed by the GB and KNN models. However, the differences were small, so it is challenging to assert one model as distinctly superior. Ensemble models generally demonstrated higher accuracies and lower variance compared to single ML models, aligning with previous research highlighting the advantages of ensemble models (Dietterich 2000).

Of all the models assessed, the voting classification model, which was constructed using the two best-performing classification models, was identified as the most accurate model, predicting the correct class with an 83 percent accuracy rate. This was an improvement of 5.2 percent over the best-performing standalone model, SVM, and 11.8 percent over the least-accurate model, ANN. The stacking ensemble model, created using the two top-performing classification models, was identified as the second-most-accurate model. The advantages of the ensemble method were evident not just in the accuracy measurements but also in the consistent top-tier performance they offered across all cross-validation sets. Although the accuracy ranking of individual models fluctuated based on the sampling variations from the cross-validation,

ensemble models consistently achieved top-tier performance (figure 9; for clarity in visualization, only ensemble models constructed using the two highest performing individual models are illustrated). This finding indicates the advantage of ensemble approaches in ensuring top-tier performance for pavement condition classification models, alleviating concerns that other individual models might outperform under varying datasets.

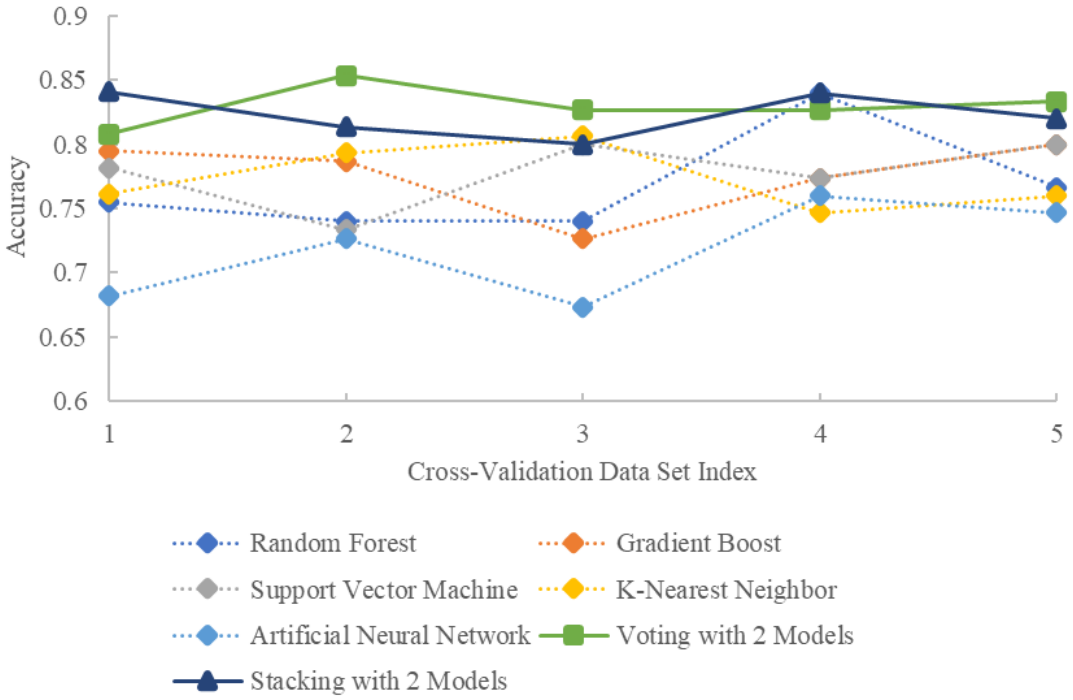


Figure 9. Graph. Accuracy of individual and ensemble models for each cross-validation set.

As described previously, ensemble models achieved the best performance when developed using the two top-performing individual models. Figure 10 illustrates the variations in accuracy of ensemble models depending on the count of individual models incorporated. As more individual models were integrated into the voting model, the accuracy tended to decrease. This trend could be due to each individual model being given equal importance in the voting process, allowing models with lower accuracy to sway the final predictions toward incorrect

classifications. In future works, exploring methods to assign appropriate weights to votes from each individual model could be beneficial in enhancing the accuracy of voting ensemble models built using multiple classifiers. On the other hand, the performance of stacking models decreased when the models were developed with three and four individual models but improved as five individual models were involved. Further research is needed to determine whether stacking models benefit from a greater number of individual models or if specific combinations of classification models enhance predictions for pavement condition classes.

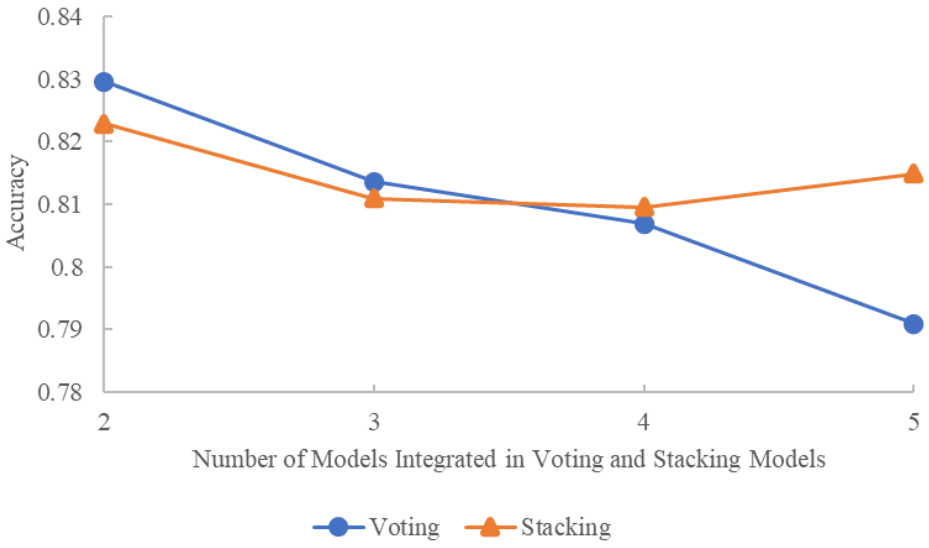


Figure 10. Graph. Accuracy of voting and stacking models based on integrated number of classifiers.

The voting and stacking ensemble models demonstrate enhanced predictive capabilities over singular ML models. Ensemble models leverage the combined strengths of multiple ML approaches, enhancing their ability to make predictions. The integration of diverse ML models allows for a variety of hypotheses to inform predictions, promoting the unique ability of each model to detect different patterns within the data (Dietterich 2000). This multifaceted approach

can be particularly beneficial when the amount of data is limited, as it helps to prevent overfitting (Sagi and Rokach 2018). Overfitting occurs when a model is excessively tailored to the training data but is yielding unsatisfactory predictive accuracy on unseen data. By balancing out the weaknesses of individual ML models and reducing the likelihood of relying on a biased hypothesis, ensemble models are better equipped to provide reliable predictions and mitigate overfitting (Polikar 2006).

Although ensemble models have the potential to deliver superior performance, developers must be aware of the challenges inherent in constructing such models. Involving multiple individual learners adds complexities in developing ensemble models. First, the heightened computational resources needed for training ensemble models, along with the time it takes for these models to make predictions on new data, are critical considerations, particularly with stacking models that involve training an additional meta-learner. Moreover, the complexity of ensemble models can hinder the interpretability of effects of explanatory variables. The multi-layered interaction among various predictors can make it difficult to discern the effects of independent variables on the predicted outcome. Furthermore, creating ensemble models demands a more profound knowledge and greater effort than is required for implementing a single ML model. Model developers are tasked with careful preprocessing of data to ensure compatibility with each base learner, and they must also dedicate time to fine-tune the hyperparameters for each model (Sagi and Rokach 2018).

Therefore, in practice, the employment of ensemble models to produce more reliable and consistent predictions should take into consideration factors such as the available data, computational resources, the requirement for model interpretability, and the proficiency of the modelers in training and tuning multiple ML models.

Feature Analysis

After constructing pavement condition prediction models, feature importance was employed to rank the critical variables influencing the predictions. Feature importance is gauged by the reduction of impurity, a measure indicating how each feature contributes to purifying the dataset at each decision node within the trees in the decision tree–based algorithms. Impurity is quantified using metrics such as Gini impurity, which measures the frequency at which any element of the dataset will be wrongly labeled if it was randomly labeled according to the distribution of labels in the node (Kazemitabar et al. 2017).

Among the ML algorithms discussed in the previous section, the RF classification model, which is a decision tree–based algorithm, was employed to perform the feature importance. In the RF classification model, each time a feature is used to split a node, it reduces the impurity of that node. The magnitude of this reduction across all trees is indicative of the feature's effectiveness. A greater decrease in impurity due to a feature suggests a stronger capability of that feature to segregate the data, thereby improving the model's accuracy.

To quantify feature importance, the sum of impurity reductions attributed to each feature is measured and then normalized by the total impurity reduction across all features and trees in the forest. This normalized value represents the feature importance ratio, which illustrates the relative contribution of each feature to the overall reduction in impurity.

Figure 11 illustrates the importance of the variables to predict the outcomes from the RF classification model.

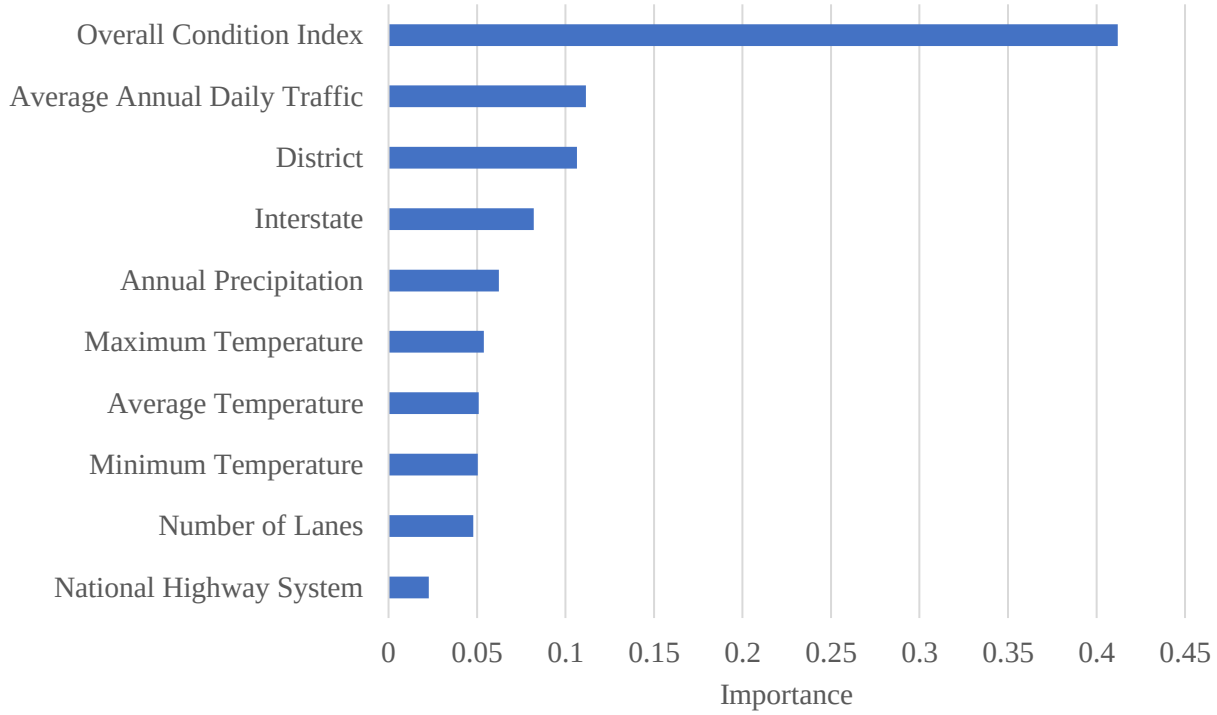


Figure 11. Graph. Feature importance from the RF classification model.

As anticipated, the current OCI is crucial for predicting future pavement condition classes, accounting for 40 percent of the total impurity reduction. Following OCI in significance are AADT, the specific district, and whether the road is part of the interstate system. Together, these top four features contribute to over 70 percent of the total impurity reductions, underscoring their substantial impact on the model's predictive capabilities. Factors such as temperature and precipitation, the number of lanes, and NHS criteria have a relatively minor influence on prediction outcomes. These features collectively contribute less to the model's ability to reduce impurity and, thus, are considered less pivotal in predicting pavement conditions.

Despite certain variables exhibiting low importance in our feature analysis, including all variables in the model could potentially enhance performance by capturing subtle, complex

interactions that might otherwise be overlooked. However, this approach involves a trade-off, as incorporating a comprehensive set of variables requires more extensive data collection, which can be both time-consuming and costly. Additionally, the inclusion of more variables increases the computational demands of the model. It would extend the time required for training and prediction, thus escalating the overall cost of the modeling process. Therefore, a balanced approach must be considered, weighing the potential benefits of improved model accuracy against the practical constraints of data collection and computational efficiency.

CHAPTER 3. MAINTENANCE COST ESTIMATION MODEL

BACKGROUND

Highways are a large part of transportation, and their well-function plays a crucial role in economic and social development. However, many state highways built around the 1950s have surpassed their initial design lifespan. Although studies have shown that damaged roads lead to substantial social costs such as reduced travel speeds and increased chances of vehicle crash rates, the investment in pavement maintenance and the necessity of endorsing strategic management has been emphasized (Bock et al. 2021, Santero and Horvath 2009, Wang et al. 2014, Small and Winston 1988)

Transportation-related legislation has continued to be established and updated to facilitate such functions in a safe, innovative, and sustainable manner. Those have been enacted to set the framework for transportation policy, funding, and planning, and the rising budgets show the increased importance and interest in the transportation and infrastructure sectors. Such recognition is notable from the recent legislations during the past decade: Moving Ahead for Progress in the 21st Century (MAP-21) from 2013 to 2014, Fixing America's Surface Transportation (FAST) Act from 2016 to 2020, Infrastructure Investment and Jobs Act (IIJA) from 2022 to 2026, which allocated approximately $105 billion, $305 billion, and $1.2 trillion, correspondingly. The most recent IIJA, known as the Bipartisan Infrastructure Law, has made the most significant investment in transportation infrastructure since the construction of the Interstate Highway System in the 1950s and 1960s.

Nonetheless, despite the increasing budget of the transportation and infrastructure sector, the expenditure on state and local roads has remained stable compared to the other sectors, such as public welfare, health and hospitals, and housing. The U.S. Census Bureau's Annual Survey

of State and Local Government Finances data show that the share of state and local spending on highways and roads also fell from 8 percent in 1977 to 6 percent in 2020 (Urban Institute 2022). In cases of federal spending for highways, the Congressional Budget Office anticipates that the Highway Trust Fund's balance, allocated for both highway and transit accounts, will be depleted in 2028 (Shirley 2023). The demand is outpacing the budget in the context of pavement maintenance. Moreover, the coronavirus disease 2019 pandemic in 2020 induced comprehensive shifts and complications in the construction sector. The confusion affected various perspectives, such as workforce, supply chain, material price, etc., causing higher costs in projects as a result (Biörck et al. Biörck, Alsharef et al. 2021, Assaad and El Adaway 2021). With possible remnants and other unprecedented uncertainties from the market, the ability to establish thorough financial planning became more critical for state departments of transportation (DOTs) to attain successful asset management.

State DOTs are required to build a more strategic transportation asset management plan (TAMP) to utilize the budget allocations effectively. However, the National Cooperative Highway Research Program (NCHRP) research report mentions that the state DOTs have had trouble in the path of establishing financial plans (Allen et al. 2023). The challenges ranged widely from lack of quality data to inconsistent definitions of activities, which led to short interpretation of maintenance on assets. The unalignment between short-term maintenance budgeting periods and long-term TAMP periods was identified as one of the barriers to establishing an effective TAMP. Therefore, the previous approach, utilizing multiple indices representing market conditions to address the indigenous uncertainty in the construction industry, should be revised. The project focuses on satisfying the following conditions for building a cost prediction model to be successfully incorporated into asset maintenance plans:

- Cost predictions should achieve high accuracy if not acceptable accuracy.
- Prediction models should consider the various asset conditions.
- Long-term predictions should be feasible and straightforward enough for a long-term financial plan.

For such objectives, the framework can be largely divided into two sections whereby (1) the authors investigate the application of ML algorithms for early-stage cost predictions of highway maintenance projects and (2) the authors delve into the tree-based algorithms to enhance the overall performance of cost forecasting and explore the contributions of features.

LITERATURE REVIEW

Previous research has focused on developing accurate cost estimates because incorrect cost assessments result in significant problems when allocating resources efficiently and completing projects on schedule. When the distributed budget falls short, the project may experience delays or fail to meet the anticipated outcomes (Asmar et al. 2011). The converse situation may lead to unused funds, depriving other projects needing financial resources (Alavi et al. 2009, Chou 2011). Previous studies sought to integrate the risks in the prediction model to achieve high accuracy by employing various models including time-series models, regression models, case-based reasoning, and advanced ML algorithms.

Although some may not be directly applicable to highway maintenance projects, numerous studies have been conducted regarding the prediction of construction costs. Ilbeigi et al. (2017) built a univariate time-series forecasting model focusing on the impracticality and erroneousness of the future explanatory variable values. The authors built a cost-effective asphalt cement cost prediction model that requires only one historical input with the best mean absolute

percentage error (MAPE) of 2.26 percent. Zhang et al. (2017) used the least absolute shrinkage and selection operator (LASSO) regularized regression to forecast completed highway construction costs with a MAPE of 7.10 percent. Employed variables were related to project and economic conditions, which included the consumer price index (CPI) as one of the primary features selected. Li et al. (2021) developed multivariate time-series models to forecast the ratio of a low bid to an owner's estimate for highway projects. The authors also used market variables representing the local highway construction market, the construction market, macroeconomic conditions, etc. Li and Ashuri (2021) also used Cox regression models to quantify the likelihood of underestimation using several groups of variables, which again included environmental and market-related variables.

Advanced ML algorithms have also been actively engaged in studies, showing high accuracy. Cao et al. (2018) predicted the unit price bids of resurfacing highway projects using ensemble ML. With features regarding project characteristics and various market indices, they reached an average estimation error of 7.56 percent. Wilmot and Mei (2005) then implemented neural network modeling to estimate the annual overall costs and calculate the Louisiana highway construction index. The authors built five submodels that each estimate the price of representative pay items: embankment, concrete pavement, asphalt pavement, reinforcing steel, and structural concrete. The results of the submodels were aggregated to calculate the Louisiana highway construction index, which captured about 95 percent of the variation.

Although the aforementioned studies have had significance in enhancing the performance of cost prediction models, the barriers to cost estimation in such an objective still prevail. With sophisticated data and model-wise requirements, it is difficult for the owners to apply the current suggestions because the project still needs to be developed and lacks details. For example, Cao

et al. (2018) predicted the unit price bids of resurfacing highway projects using ensemble ML with an average estimation error of 7.56 percent. However, the 57 employed features include those that cannot be utilized for the objective of this research, which are those relevant to the market information, such as the CPI. Likewise, many studies use the factors that will not be accessible for the upcoming years at the stage of planning, such as external or environmental factors, as their input (Ashuri et al. 2022, Chou et al. 2015, Li and Ashuri 2021, Teicholz 1993, Zhang et al. 2017). Throughout this chapter, ML algorithms are revisited to investigate the suitability for long-term prediction and budgeting. The models aim to achieve an acceptable accuracy rate despite the absence of such variables to consider their use in the early stages of project design.

Moreover, no research has yet been conducted on the preliminary cost prediction of highway pavement maintenance projects. Because the project is designed based on the pavement condition, it is natural for the prediction model to reflect the conditions and characteristics of the asset. This research focuses on the cost estimation of maintenance projects by accounting for the pavement characteristics. Although Ilbeigi et al. (2017) used univariate time series to overcome the problem related to feature availability, the objective was to build an asphalt cement cost prediction model. Sonmez and Ontepeli (2009) also built a preliminary cost estimation model, but the model was targeted for urban railway projects. Similarly, previous work has not focused on highway maintenance projects in the early stages but on other subjects such as construction cost index, highway construction cost index, and bridge construction bid award cost (Cao and Ashuri 2020, Cheng et al. 2010, Chou et al. 2015, Herbsman 1983).

The second part of this chapter shows the significance of tree-based algorithms, especially the ET method, in terms of the practicality and applicability of the model. Recent

research focused on reducing prediction errors, especially utilizing neural networks (NNs; Adeli and Wu 1998, Cheng et al. 2010, Hegazy and Ayed 1998). For example, Sonmez and Ontepeli (2009) used NN models for predesign estimation of railway projects, achieving the best MAPE of 33.3 percent. Chou et al. (2015) also demonstrated the superiority of artificial intelligence models, especially ANN with optimization through the genetic algorithm, compared to the case-based reasoning to forecast bidding prices in transportation projects.

However, despite its capability, the NN suffers from several disadvantages (Tu 1996). Not only is NN computationally expensive compared to other models, but its performance is not guaranteed. The NN has an empirical nature in the model development, which makes the model performance subject to various factors such as the model structure or data (Hegazy et al. 1994). Although the model is developed for high accuracy based on historical data, the significant fluctuation and uncertainty of the project costs may raise a model performance problem related to the overfitting or the sensitivity to outliers. The research suggests that the robustness, speed, and ease of implementation of the tree-based algorithms are conducive to the shortcomings of the NN. The tree-based algorithm presents reasonably consistent accuracy without many assumptions, preprocessing, or parameter tuning. The tree algorithms usually require the selection of two parameters disparate from the NN, which requires multiple tuning of parameters. The algorithms also do not assume the normality of data, nor are they sensitive to the scaling of input features.

DATA DESCRIPTIONS

The dataset consists of a total of 394 samples with 49 features, including the continuous and categorical variables. The data sample utilized in the cost forecasting models can be divided into

three categories: bidding-related data, asset-related data, and project-related data, as shown in table 16.

Table 16. Feature summary.

Group	Variables	Units/Categories	Description
Bidding Related	Bid Awarded Cost	Numerical ($)	Awarded cost of the project (calculation of cost per lane mile)
	Year	Numerical	Year of project letting date
	Month	Categorical: {1,2,3,4,5,6,7,8,9,10,11,12}	Month of project letting date
Asset Related	Latitude	Numerical (degrees)	Latitude of the center of the selected maintenance section
	Longitude	Numerical (degrees)	Longitude of the center of the selected maintenance section
	Interstate	Categorical: {Interstate, Non-Interstate}	Interstate: 6
			Non-Interstate: 388
	State Route Priority	Categorical: {Critical, High, Medium, Low}	Critical: 66
			High: 131
			Medium: 98
			Low: 99
	National Highway System (NHS)	Categorical: {NHS, Non-NHS}	NHS: 172
			Non-NHS: 222
	District	Categorical: {District 1, District 2, District 3, District 4, District 5, District 6, District 7}	District 1: 55
			District 2: 39
			District 3: 68
			District 4: 70
			District 5: 56
			District 6: 63
			District 7: 43
	Average Annual Daily Traffic (AADT)	Numerical (vehicles per day)	Average annual daily traffic for selected maintenance section
	Length of Sections	Numerical (miles)	Length of the maintenance section
	Number of Lanes	Numerical (count)	Number of lanes in the maintenance section
	Previous Year Pavement Condition Class (OCI)	Categorical: {A, B, C, D, E, F}	A: 11
			B: 44
			C: 164
			D: 167
			E: 8
			F: 0
Project Related	Total Miles	Numerical (miles)	Total length of the project
	Project Duration	Numerical (days)	Days from bid letting date to project completion date
	Number of Sections	Numerical (count)	Number of maintenance sections in the project
	Number of Counties	Numerical (count)	Number of counties involved in the project
	Treatment Type	Categorical: {Nothing, Lite Treatment, Minor Preservation, Major Preservation, Minor Rehabilitation, Major Rehabilitation}	Nothing: 5
			Lite Treatment: 42
			Minor Preservation: 152
			Major Preservation: 173
			Minor Rehabilitation: 18
			Major Rehabilitation: 1

Note: Numerical, continuous variables were examined before input to the model. When exhibiting skewness in distribution, logarithmic transformations were performed and added to the initial pool of features.

Bidding information includes the letting date and bid awarded cost. These were extracted from the BidTabs database from Oman Systems (OSI), where the FHWA organizes the historical bidding records for each DOT project. Letting date is separated into a Year and a Month variable to implement possible trend or seasonality. Bid Awarded Cost will be the target variable for the final cost prediction model, which uses tree-based algorithms. The Cost per Mile Lane will be calculated and used as a target variable for the first step of examining the applicability of various ML models. The Cost per Mile Lane is calculated as follows:

$$C_{ml} = Bid\ Awarded\ Amount/(Total\ Miles * Lanes\ in\ Section) \quad (5)$$

The asset-related data were collected from GDOT-owned geospatial information systems (GIS) shape files where highway assets are managed based on the maintenance sections. Binary and categorical variables, such as Interstate, NHS, State Route Priority, and District, can depend on the characteristics of the selected maintenance section. Numerical variables, such as Length of Sections, Number of Lanes, and Average Annual Daily Traffic (AADT), are also extracted based on the target maintenance sections. On the other hand, the asphalt pavement data consisted of records regarding pavement conditions expressed in a GDOT-specific index (OCI) for each maintenance section. The OCI is calculated by averaging six pavement serviceability scores, ranging 0–100. The asset conditions from 2017 to 2021 will be utilized in a categorized grade format (table 17).

Table 17. Overall Condition Index scores and classes.

Condition Class	**A**	**B**	**C**	**D**	**E**	**F**
OCI score	100−90	80−90	70−80	60−70	50−60	Below 50

Finally, Total Miles and Project Duration, which are collected through project proposals and GeoPI, where the GDOT's project information is stored in a geographical map, are included in the project-related variables. The number of sections and counties are also calculated from the data in OSI. The project descriptions and objectives are supposedly determined through the GDOT's decision tree in practice. The GDOT's decision tree of proceeded maintenance treatments can be simplified in this study, as shown in table 18. The treatments are decided (assigned) based on the OCI score of each segment and whether the segment is included in the NHS. The maintenance type is judged based on the simplified decision tree and used as an input for the forecasting models (table 18).

Table 18. The GDOT's simplified decision tree for maintenance treatment types.

Interstate (Binary)	Overall Condition Index	Maintenance Treatment
Interstate	100−75	No Treatment
	75−60	Major Preservation
	50−60	Minor Rehabilitation
	1−50	Major Rehabilitation
	<1	Reconstruction
Non-Interstate	100−90	No Treatment
	90−80	Lite Treatment
	80−70	Minor Preservation
	70−60	Major Preservation
	50−60	Minor Rehabilitation
	1−50	Major Rehabilitation
	<1	Reconstruction

MODEL ALGORITHMS

Supervised Machine Learning

Supervised learning is a ML technique by which the models are trained with the labeled data. The learner can then map between input data to their desired outcomes (Nasteski 2017). This

paper leverages six supervised learning algorithms: LR, RF regression, ET regression, GB, eXtreme GB (XGB), and KNN.

Apart from LR, the RF and ET models are ensemble methods built on decision trees. Both train multiple decision trees and aggregate the results, comprising numerous weak decision tree learners (Ahmad et al. 2018). However, whereas RF greedy searches to select split values and uses bootstrapped samples for training decision trees, ET randomly selects splits and uses the entire dataset for training (Breiman 2001, Geurts et al. 2006). The randomness allows ET to reduce the model bias and lessen the probability of overfitting. GB also combines learners by iterating and adding weak learners using gradient descent (Friedman 2001). The decision trees are fitted and modified to reduce the error from the previous model. XGB uses advanced regularization and parallel processing to implement GB efficiently (Chen and He 2024). Conversely, KNN makes predictions by relying on the proximity of data points in feature space (Peterson 2009). The regression value is assessed by local interpolation of the selected number of nearest neighbors.

Tree-based Algorithms

This paper applies and compares two types of tree-based models, RF and ET models, which are both ensemble methods based on decision trees. The RF and ET train multiple decision trees and then amass the results, incorporating the numerous weak decision tree learners (Ahmad et al. 2018). Decision trees recursively split the dataset based on the optimal split of each node, which maximizes the information gain. This optimal selection makes the decision tree prone to overfitting and sensitive to the input data.

Although the decision tree has high variance, RF combines multiple decision trees for generalization (Breiman 2001). RF builds decision trees based on the best split from a randomly

selected subset of predictors for each bootstrapped sample of data. This randomness not only helps with the high prediction accuracy, but the algorithm is also robust against overfitting and less sensitive to outliers (Ali et al. 2009). ET uses the entire dataset for training and randomly selects the splits, unlike RF, which uses bootstrapping and greedy searching (Geurts et al. 2006). This difference reduces variance and bias, resulting in superior prediction with noisy data. ET also shows much faster computation time (Geurts et al. 2006)

Feature Selection Algorithm

Feature selection has been widely used in ML as a part of the preprocessing step for enhanced performance and cost-effectiveness of the models (Guyon and Elisseeff 2003). Out of various feature selection methods, this research uses the greedy approach of selecting the most pertinent *k* number of features among the initial set of features based on mutual information (MI; Battiti 1994). The MI, or information gain, quantifies the dependency between a pair of variables regardless of the linearity of the variable relationships. It measures a variable's average degree of uncertainty given information about the other variable (Shannon 1949). The calculations for MI given features *x* and *y* are as follows (Guyon and Elisseeff 2003, Vergara and Estévez 2014):

$$I(x;y) = H(x) - H(x|y), \tag{6}$$

where,

$$H(x) = -\sum_{i=1}^{n} p(x(i)) \cdot \log\left(p(x(i))\right) \tag{7}$$

$$I(x;y) = \sum_{i=1}^{n}\sum_{j=1}^{n} p(x(i),y(j)) \cdot \log\frac{p\big(x(i),y(j)\big)}{p\big(x(i)\big)\cdot p\big(y(j)\big)} \quad (8)$$

Here, H(x) is the entropy, or the uncertainty, of variable x. The MI is 0 when the variables are independent as $p\big(x(i),y(j)\big) = p\big(x(i)\big)\cdot p\big(y(j)\big)$, and a higher MI value indicates that more information is provided between the features.

APPLICABILITY OF MACHINE LEARNING ALGORITHMS

Methodology

The overall framework of research is illustrated in figure 12. As noted in the data collection section, the three data categories (bidding, asset, and project-related data) are collected from corresponding sources. After data collection and categorization, initial data cleaning and preprocessing are performed. The data cleaning process includes calculating additional variables or handling incomplete data, leaving 231 projects, 384 maintenance sections, and 394 project-maintenance section samples. Different preprocessing is performed based on the data types, but in this stage, one-hot encoding is applied to the categorical or binary features. All categorical values are mapped into and assigned separate columns. They are transformed into binary vectors for computing compatibility.

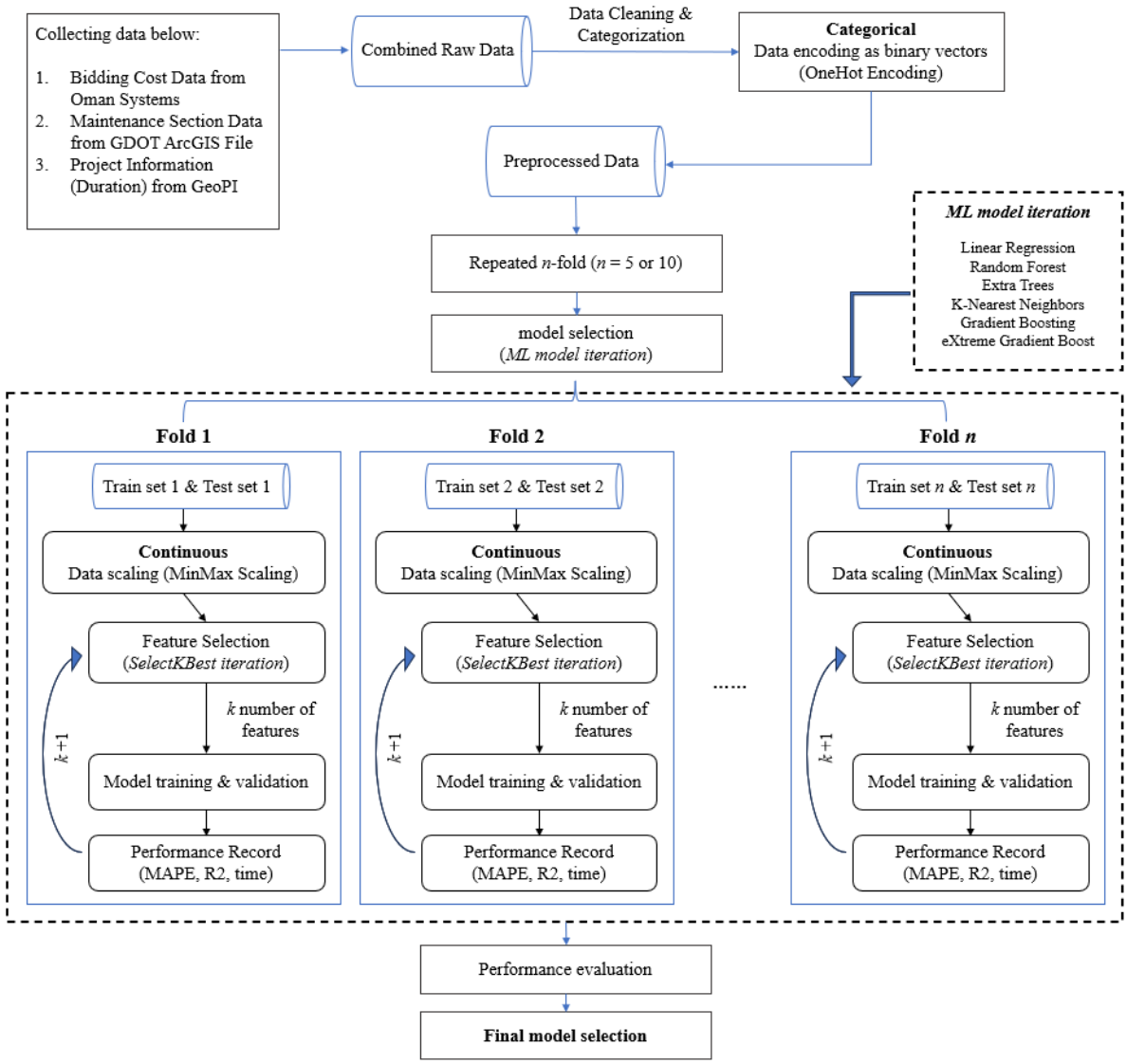


Figure 12. Flowchart. Cost estimation model framework.

Data are then divided into non-overlapping 5 and 10 folds for cross-validation of the model. Each fold is set as a test set, whereas other $n-1$ folds are used to train the model. A total n number of models can be fit and evaluated on n different testing sets and the average performance is calculated. This cross-validation diminishes possible bias from data splitting and prevents overfitting (Rodríguez et al. 2010). Although a single n-fold cross-validation may still have the possibility of reporting a noisy estimate of model performance, the authors repeat the

process three times to overcome the possible noisy estimate of model performance when running a single *n*-fold cross-validation.

The continuous variables were scaled based on the minimum and maximum values of the dataset so that all features could have the same range of numbers from 0 to 1. The base parameter values for each feature are retrieved from the training set of each case of *n* folds and then applied to the whole dataset to prevent any information leaks. The scaling equation is as follows:

$$x_{scaled} = \frac{x_{original} - x_{min}}{x_{max} - x_{min}} \tag{9}$$

selection is also performed for each *n*-fold. The *k* features with the highest estimated MI for the target variable are selected and input to the model (Kraskov et al. 2004). The refined data are input to train models for each 3·*n* fold to get the average model performance. This process will be repeated for six ML algorithms, and each *k* iteration (*k* ranges 0–49) will be compared based on three model performance metrics.

Results

Performance Metrics

The metrics include MAPE, R^2, and computation time. The computation time (seconds) is the sum of the time spent fitting and scoring for each fold in cross-validation. The equation for MAPE is as follows:

$$MAPE = \frac{C_{true} - C_{pred}}{C_{true}} * 100 \tag{10}$$

Here, C_{true} is the actual bid awarded cost, and C_{pred} is the predicted cost of the project.

The R^2, known as the coefficient of determination, is also recorded to be interpreted as the fraction of variance that can be predicted from the model or the independent variables (Wright 1921). The equation for R^2 is given as:

$$R^2 = 1 - \frac{residual\ sum\ of\ squares}{total\ sum\ of\ sqaures} = 1 - \frac{\sum_i \left(C_{true,i} - C_{pred,i}\right)^2}{\sum_i \left(C_{true,i} - \overline{C_{true}}\right)^2} \quad (11)$$

where $\overline{C_{true}}$ is the mean of the actual value. If the model outcomes show an exact match to the actual values, R^2 will be 1.

Model Performance

Table 19 shows the top five cases in terms of low MAPE. The model results include R^2, calculation time, and MAPE with the top *k* number of features. The Time column exhibits the computation time as one of the performance metrics.

Table 19. Top five results for each ML algorithm.

	5-fold				10-fold			
	k	MAPE (%)	R^2 (%)	Time	*k*	MAPE (%)	R^2 (%)	Time
Linear Regression	45	30.678	23.266	0.086	42	28.939	22.006	0.097
	46	30.708	23.178	0.084	44	29.078	21.438	0.093
	48	31.092	22.149	0.086	45	29.246	20.925	0.104
	44	31.376	23.609	0.085	43	29.300	20.598	0.095
	43	31.388	23.569	0.086	41	29.358	21.825	0.101
Random Forest	45	24.711	43.711	0.601	21	22.831	44.270	0.477
	28	24.805	48.610	0.536	20	22.847	45.138	0.469
	16	24.872	47.727	0.402	39	22.897	42.993	0.744
	19	24.876	48.403	0.431	23	22.903	44.421	0.531
	47	24.909	44.821	0.605	19	22.915	44.659	0.463
Extra Trees	9	18.513	51.948	0.207	19	16.891	49.305	0.295
	10	18.690	53.377	0.217	12	16.953	54.068	0.252
	12	18.741	53.600	0.224	17	16.991	50.379	0.280
	11	18.823	51.006	0.223	7	17.026	47.888	0.226
	17	19.081	52.525	0.259	16	17.078	50.028	0.276
K-Nearest Neighbors	33	33.843	15.190	0.091	35	32.235	13.839	0.093
	29	33.880	14.415	0.094	32	32.365	14.321	0.094
	31	33.897	12.327	0.092	36	32.451	13.242	0.095
	30	33.963	14.348	0.093	33	32.554	14.625	0.095
	32	34.211	14.303	0.091	34	32.758	13.216	0.095
Gradient Boosting	38	24.752	48.615	0.229	40	23.877	43.128	0.260
	21	24.818	48.082	0.189	21	23.895	45.183	0.214
	23	24.854	46.290	0.195	20	23.920	42.893	0.216
	37	24.857	47.035	0.229	43	24.023	43.630	0.260
	22	24.859	47.615	0.194	41	24.130	43.027	0.261
eXtreme Gradient Boosting	9	21.271	36.275	0.160	12	18.515	44.198	0.161
	10	21.272	41.965	0.168	8	19.129	31.567	0.155
	11	21.339	42.537	0.167	9	19.352	34.168	0.160
	17	21.356	45.266	0.169	11	19.456	39.408	0.163
	18	21.462	46.044	0.169	7	19.480	32.932	0.154

The results of the final model and their *k* values are organized in table 20 and table 21. The *k* values for each model were selected when the result showed the lowest MAPE. Overall, the MAPE results for 10-fold cross-validation were smaller than those for 5-fold. However, the 10-fold performance was slightly worse than 5-fold in terms of R^2. The ET model demonstrated the highest accuracy in both cases, with the XGB model following closely behind.

Looking deeply into the 5-fold cross-validation results (table 20), ET and XGB achieved the best results using only nine features, the fewest of all cases. The ET model shows the best

performance of MAPE at 18.513 percent, followed by 21.271 percent from XGB. In terms of R^2, the ET model explained the most considerable variability of data, which is 51.948 percent. Although GB and RF models explained over 40 percent data variability, the R^2 of the XGB model was 0.36, falling short. Inferring from the time metric, the XGB and ET algorithms took third and fourth place in computation speed, respectively.

Table 20. Final models (5-fold cross-validation).

	LR	RF	ET	KNN	GB	XGB
k	45	45	9	33	38	9
MAPE (%)	30.678	24.711	18.513	33.843	24.752	21.271
R^2 (%)	23.266	43.711	51.948	15.190	48.615	36.275
Time	0.086	0.601	0.207	0.091	0.229	0.160

Table 21. Final models (10-fold cross-validation).

	LR	RF	ET	KNN	GB	XGB
k	42	21	19	35	40	12
MAPE (%)	28.939	22.831	16.891	32.235	23.877	18.515
R^2 (%)	22.006	44.270	49.305	13.839	43.128	44.198
Time	0.097	0.477	0.295	0.093	0.260	0.161

Predictive Result Analysis

In early project development, during which there is limited project information, the expected cost is used for the initial feasibility and funding studies of long-range programs (Anderson et al. 2006). The estimates serve as an indicator of the expected degree of project cost. Due to the lack of project definition, they are expected to have a ±40 percent confidence range (Merrow et al. 1981). Also, Schexnayder and Mayo (2004) proposed that the prediction in the early stage of a project is expected to have an error of ±25 percent. In this context, the models ET, XGB, RF, and GB show the overall applicability of ML algorithms in the early-stage highway maintenance cost

forecasting models and long-term cost management. These models show MAPE below 25 percent under both cases of cross-validation.

In particular, the ET model showed the lowest MAPE: 16.891 percent in 10-fold and 18.513 percent in 5-fold cross-validation. As an ensemble algorithm, ET is less likely to be significantly influenced by the high variability and noisiness in the target variable, Cost per Lane Mile. The inherent randomness in the algorithm also contributes to its resilience to noisy data, reducing sensitivity to individual data points. For similar reasons, the ET model has the highest R^2, indicating the superior explanatory power among all the models. Nonetheless, there remain opportunities for improvement in further studies, given that the R^2 values do not exceed 0.5.

Regarding model fitting time, the ET model was not entirely superior, yet it still managed to reduce 65.56 percent (5-fold) and 38.08 percent (10-fold) of computational time compared to RF due to its inherent randomness. The LR and KNN models had the shortest computation time. Although LR models may have limitations in capturing nonlinear and complex variable relationships, they demonstrated the highest computational efficiency. On the other hand, KNN, which did not build separate models for training, exhibited fast computation, particularly due to moderate-sized datasets such as the one in this study.

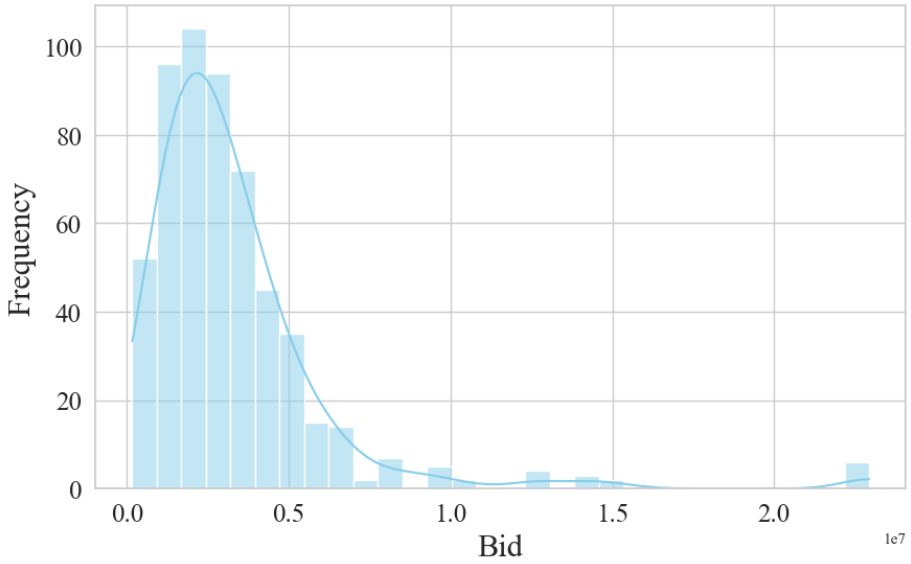
TREE-BASED ALGORITHMS

Methodology

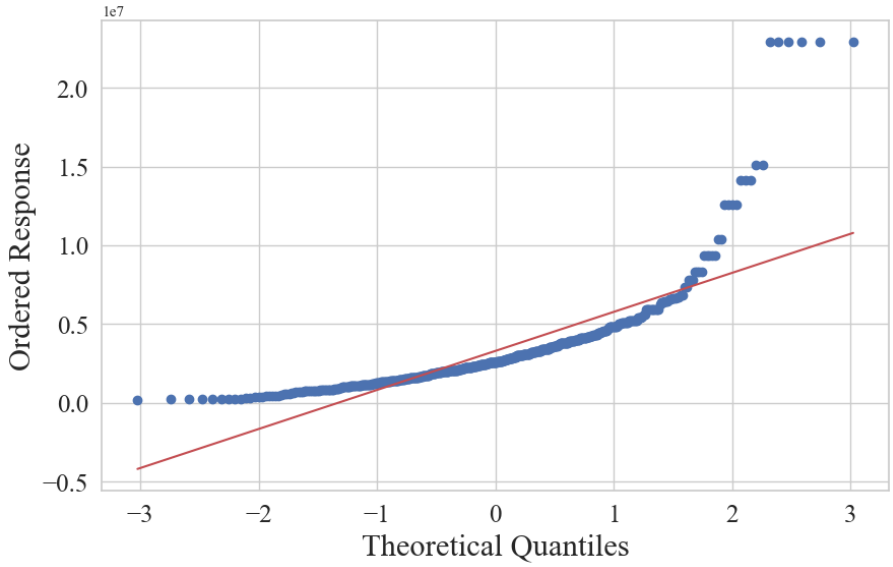
The main framework follows the methodology presented in the previous section (see figure 12). Data preprocessing consists of additional variables, handling incomplete data, and applying one-hot encoding to the categorical or binary features. However, the scaling processes are excluded because the decision trees are invariant to the scale of the features.

Along with preprocessing of the input variables, the target feature, the bid awarded amount, also goes through transformation. The histogram and probability plot of the target variable exhibit right skewness and heavy tail (see figure 13). The variable will be log-transformed during the model training and testing before calculating the accuracy. The histogram and probability plot of the transformed data are shown in figure 14. The R^2 value from the probability plot increased from 0.81 to 0.99, showing an improvement in the fit to normal distribution. The decrease in the absolute value of skewness from 3.69 to 0.31 proved that the transformation caused the distribution to be more symmetric.

Again, the data are then divided into non-overlapping 10 folds for cross-validation, and the validation is repeated three times. The MI for each feature is calculated, and the *k* number of features with the highest MI are selected as the inputs while *k* iterates through the range 2–46. Both the RF and ET models are fit to the selected features, resulting in four cases: (1) RF model without log transformation, (2) ET without log transformation, (3) RF with log transformation, and (4) ET with log transformation, denoted as RFW, ETW, RFT, and ETT, respectively. Each model's performance will be measured by comparing prediction results and the actual data, and the feature subset of the best models will also be investigated.

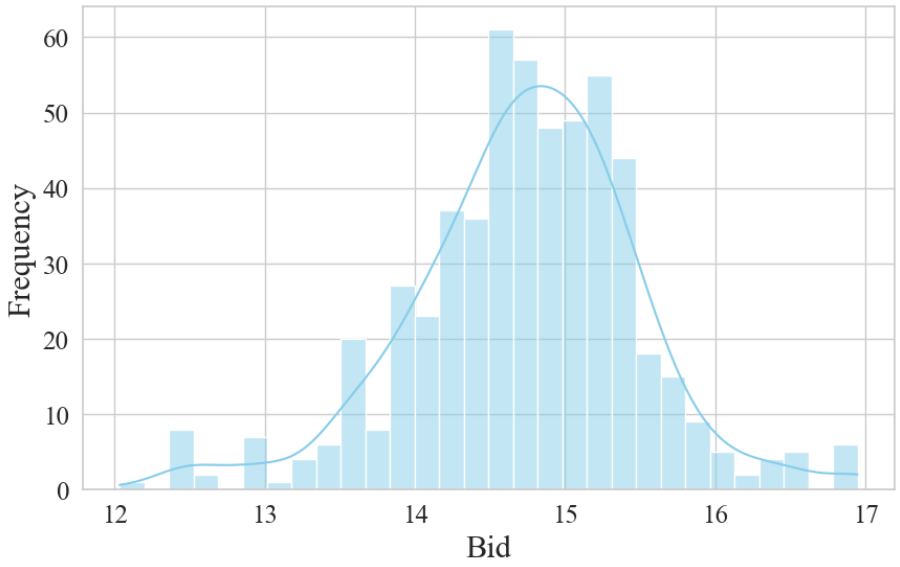


(a) Histogram

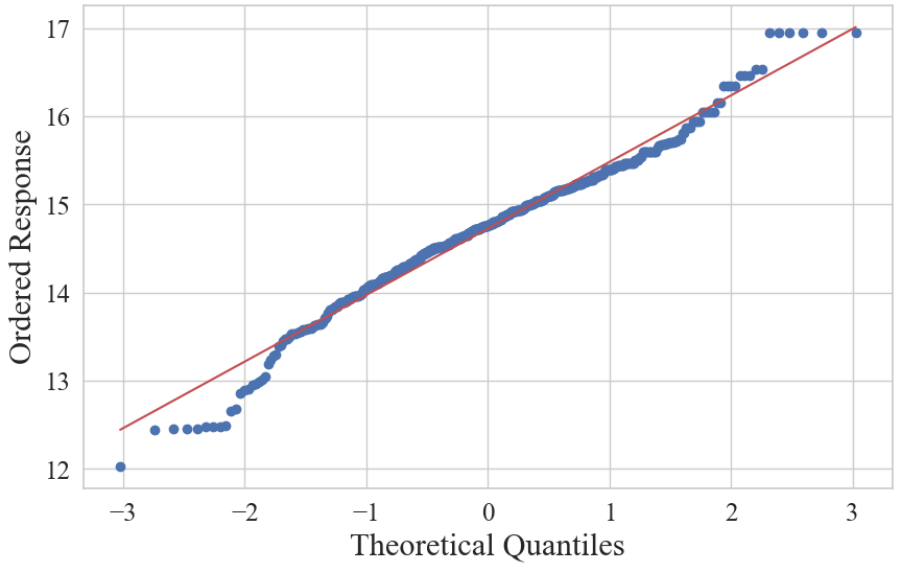


(b) Probability Plot

Figure 13. Graphs. Histogram and probability plot of the original target variable.



(c) Histogram



(d) Probability Plot

Figure 14. Graphs. Histogram and probability plot of target variable after log transformation.

Results

Performance Metrics

Three metrics for model comparison include MAPE, root mean square error (RMSE), and R^2. Smaller values of MAPE and RMSE indicate better performance according to the following:

$$MAPE = \frac{C_{true} - C_{pred}}{C_{true}} * 100 \tag{12}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(C_{true} - C_{pred})^2}{n}} \tag{13}$$

The R^2, known as the coefficient of determination, is also recorded and interpreted as the fraction of variance that can be predicted from the model or the independent variables (Wright 1921). The equation for R^2 is the following:

$$R^2 = 1 - \frac{residual\ sum\ of\ squares}{total\ sum\ of\ sqaures} = 1 - \frac{\sum_i (C_{true,i} - C_{pred,i})^2}{\sum_i (C_{true,i} - \overline{C_{true}})^2}, \tag{14}$$

where C_{true} is the mean of the actual value. If the model outcomes show an exact match to the actual values, R^2 will be 1.

Model Performance

Figure 15 and figure 16 show the MAPE and R^2 of the models based on the iteration of *k*, the number of selected features. ETT shows significant superiority in accuracy, followed by ETW, RFT, and RFW. The graphs are similar to the logarithmic functions, exhibiting the least error around the *k* value of around 20. The behavior is more decisive for ET models, giving a brief view of the importance of feature selection or reduction.

Similar behavior can be noticed in the R^2 scores of the models. The ET models show generally higher R^2 scores compared to the RF models, although the effect of target variable transformation is not always significant. The highest R^2 is reached around the *k* value of 18 for ET models, whereas RF models do not show a distinct peak of R^2. Significantly, neither the accuracy nor data explainability of the model implicitly increases linearly with more features.

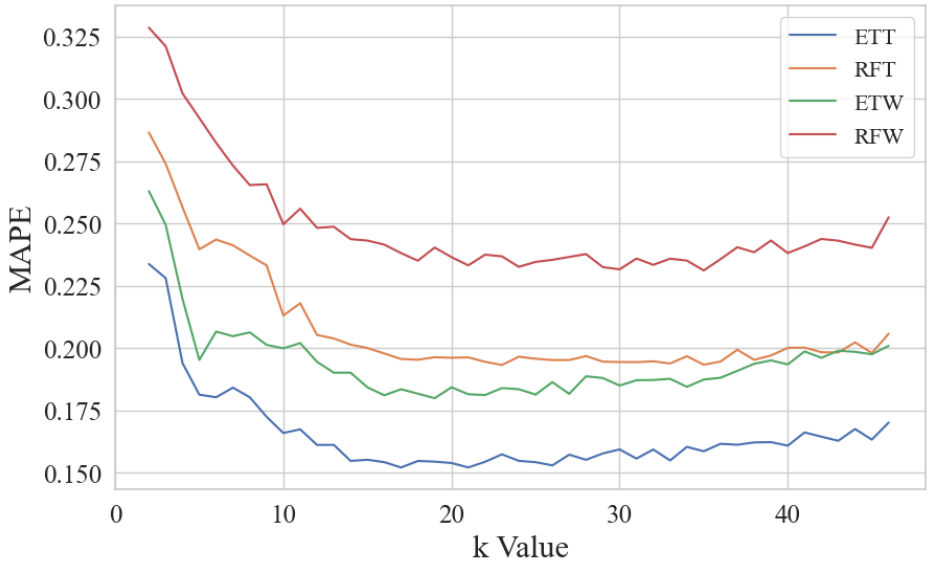


Figure 15. Graph. MAPE for different numbers of features.

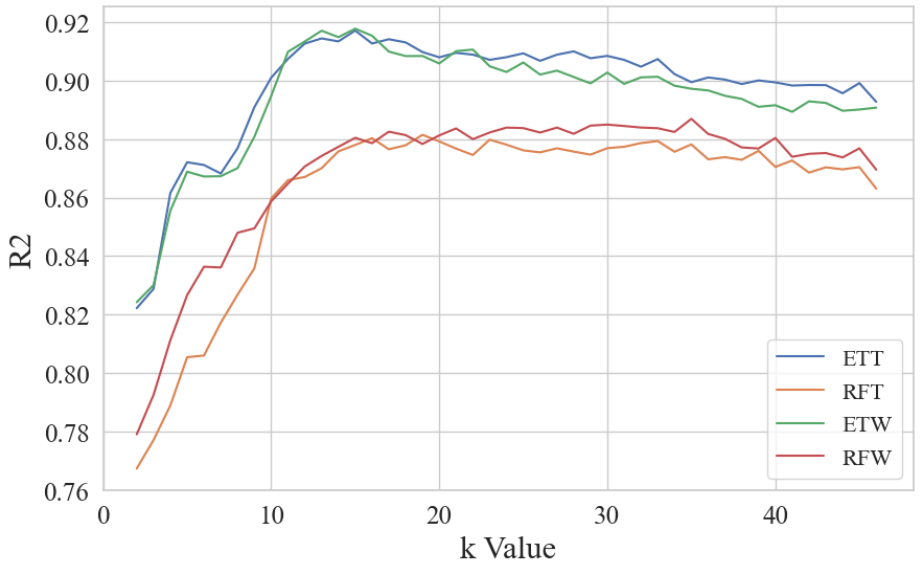


Figure 16. Graph. R^2 for different numbers of features.

Figure 17 and figure 18 show the results of the ETT. The test MAPE and R^2 scores are plotted with the fitting time as *k* increases. The model training time shows an overall consistency in the increasing trend, whereas the MAPE and R^2 do not. The MAPE and R^2 show a decrease or stabilization in the model performance after a steep improvement. The result proves that the addition of features does not always guarantee a better performance but instead worsens the usability of the model by increasing the average training time. Therefore, a suitable feature selection procedure is necessary to achieve the best performance with the smallest number of features.

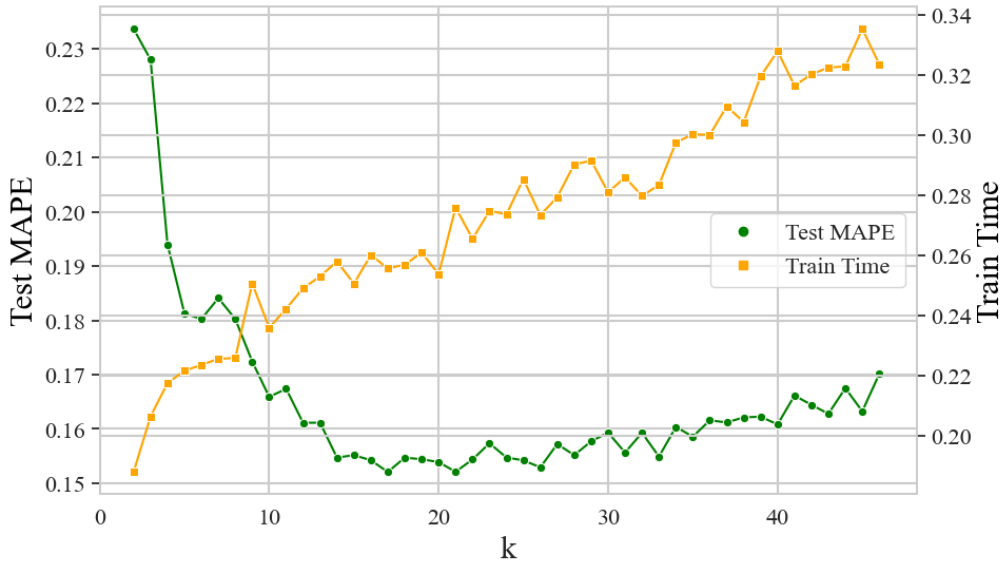


Figure 17. Graph. MAPE for ETT for different numbers of features.

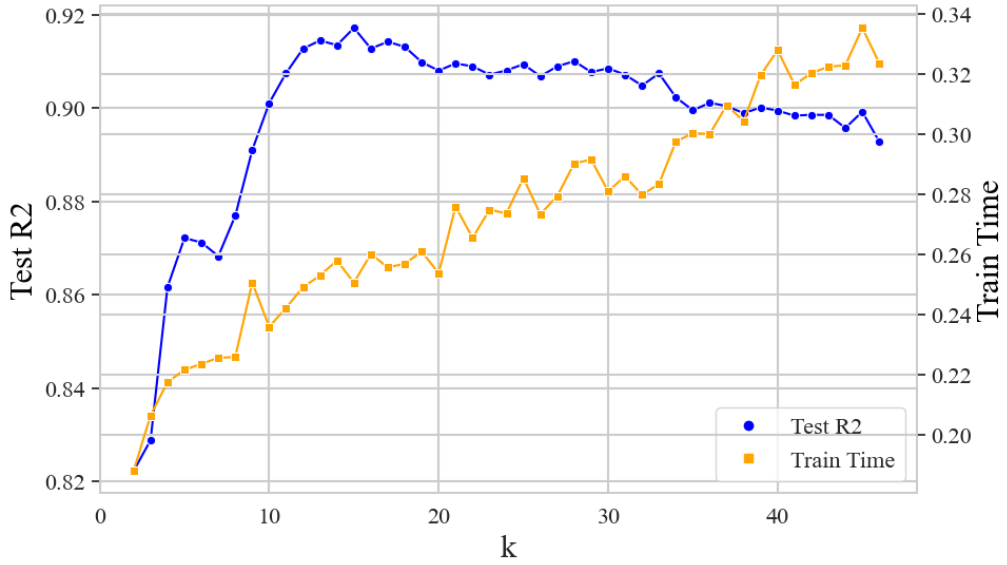


Figure 18. Graph. R^2 for ETT for different numbers of features.

Feature Analysis

Figure 19 is an example of the calculated MI and feature importance (FI) for selected features. The ETT model with the lowest MAPE is selected. The ET model has 18 features, sorted by the values of MI. The MI calculates the statistical dependency between features, and the FI shows

the reduction in model performance with and without a selected feature, using Gini impurity. MI focuses on the information gain and FI focuses on the impurity reduction, which is used for splitting the trees.

Total Miles, Duration, and Section Number had significantly high MI values exceeding 0.6, followed by features such as Year, NHS, SRP_H, and Number of Lanes. The top three features for both models with the highest FI are Total Miles, Number of Lanes, and AADT. Whereas the Duration and Section Number share more information between the target variable, the variables Number of Lanes and AADT show more substantial predictive power.

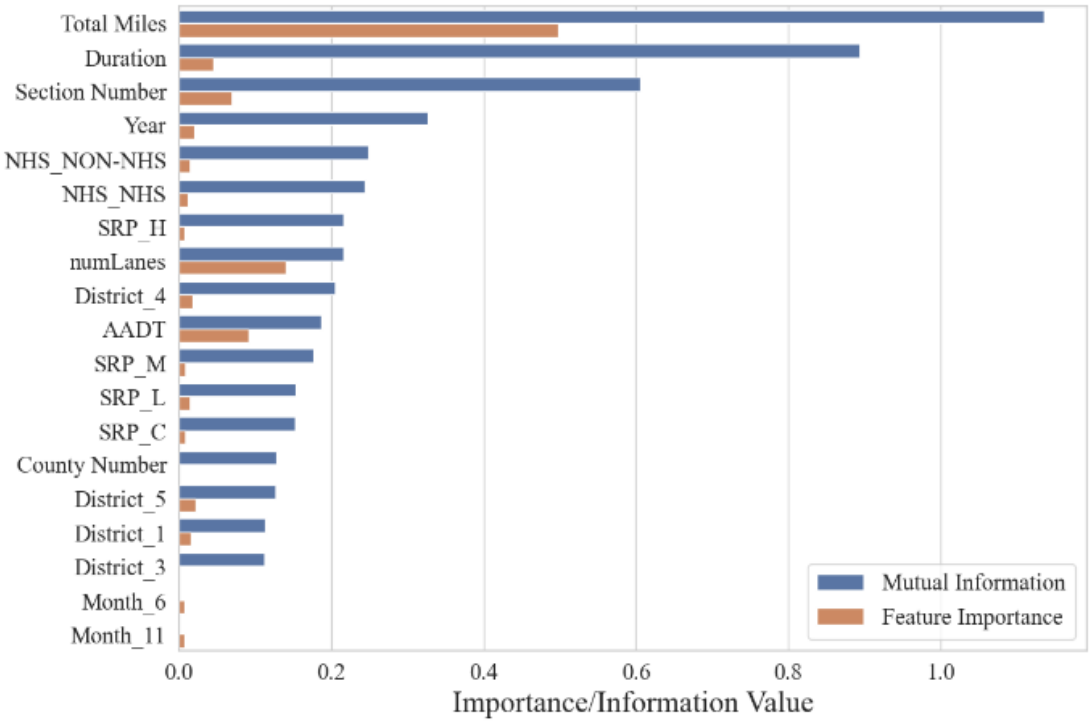


Figure 19. Graph. MI and FI for the best performing ETT model.

Predictive Result Analysis

Table 22 shows the five results for each model in terms of high accuracy or low test MAPE. Model performances, including MAPE, RMSE, R^2, and time, are recorded. Time is recorded in seconds and indicates the total time consumed for fitting and scoring. The time values will be interpreted based on the respective comparisons between the models.

Table 22. Top five results for each model (based on test MAPE).

	k	**Train MAPE (%)**	**Test MAPE (%)**	**Train RMSE**	**Test RMSE**	**R^2 (%)**	**Time (sec)**
Extra Trees (Transformed)	17	0.000	15.207	0.00	516135.48	91.420	0.266
	21	0.000	15.210	0.00	521011.67	90.955	0.286
	26	0.000	15.292	0.00	531839.02	90.681	0.284
	20	0.000	15.387	0.00	527737.94	90.801	0.267
	25	0.000	15.420	0.00	526439.65	90.938	0.293
Extra Trees	19	0.000	17.986	0.00	531546.84	90.848	0.280
	16	0.000	18.104	0.00	520993.69	91.540	0.273
	22	0.000	18.114	0.00	524246.84	91.067	0.280
	25	0.000	18.131	0.00	536772.76	90.625	0.294
	21	0.000	18.144	0.00	526132.76	91.012	0.283
Random Forest (Transformed)	23	6.832	19.318	288116.38	642229.57	87.984	0.366
	35	6.815	19.329	290043.20	645876.64	87.822	0.388
	33	6.839	19.379	294655.51	648436.00	87.936	0.382
	31	6.880	19.433	294207.91	652802.99	87.741	0.383
	30	6.836	19.438	289545.26	650781.04	87.694	0.395
Random Forest	35	9.037	23.110	253774.23	621141.96	88.697	0.386
	30	8.924	23.155	256994.44	625853.64	88.500	0.456
	29	9.012	23.243	259143.52	626340.52	88.465	0.478
	24	9.032	23.254	258034.19	630092.42	88.397	0.424
	21	9.089	23.315	258476.57	628016.74	88.369	0.348

Note: The values for the MAPE, R^2, and Time columns are rounded to the third decimal place, and the RMSE values are rounded to the second decimal place.

As illustrated in Figure 15, both ET model cases showed higher accuracy than the RF models. The best MAPE of each ETT, ETW, RFT, and RFW was about 15.2, 18, 19.3, and 23.1 percent, respectively. Log transformation of the target variable had a 15.6 percent reduction in the error for the ET model and a 16.5 percent reduction in the error for the RF model. Also,

the average MAPE of the ET models is approximately 4.6 percent smaller than that of the RF models. Such a decrease in MAPE indicates that the distribution of target variables significantly impacts the estimation and that ET is superior in cost prediction for the preliminary stages of maintenance projects. The impact of extreme target variable values is reduced due to the log transformation. The test RMSE of the best-performing model is 516,135.48, which implies that the outlier data will also exhibit an extreme value. Such considerable variation of values can have a profound impact on the model, although they may not have an equivalent impact in terms of the exploratory data itself. Another notable result is the computation time required to achieve the best MAPE. The average computation time of the top five best cases is short in the order of ETT, ETW, RFT, and RFW. The average *k* value follows the order of ETW, ETT, RFW, and RFT. In summary, the randomness of the ET algorithm has a significant influence on the model performance in terms of computation time. Also, despite the validation in Figure 17 and Figure 18 that the computation time increases as the *k* increases, the log transformation of the target variable is expected to have contributed to reducing the overall time, regardless of the number of features.

CHAPTER 4. CONCLUSIONS

Variability of cost and regional effects and the trade-off between alternative treatment strategies represent significant challenges for the GDOT's Office of Maintenance in making the right investment decisions. The regional effects on pavement deterioration rate and effectiveness of treatment methods were empirically examined using the actual data retrieved from the GDOT's asset management software to develop network-level, cost-effective maintenance plans. The updated decision trees provide a framework for tradeoff analysis between various treatment options and consider life cycle cost variability for enhanced maintenance decision-making. The updated decision tree for maintenance and rehabilitation projects can help the GDOT better plan investments to improve the conditions of its strategic assets and, thus, allocate its limited resources more efficiently and effectively to various project types in different districts. Substantial cost savings are anticipated as the result of updating the decision tree and utilizing optimal strategies for intervention.

The outcome of this research is anticipated to provide the GDOT's Office of Maintenance with systematic and comprehensive methods for developing network-level, cost-effective maintenance, and rehabilitation projects over the mid- and long-term investment horizon. The proposed research aims to help the GDOT's Office of Maintenance better allocate funding and improve maintenance strategies at the network level through updating the decision trees considering geographical differences and temporal variation of improvement cost. The actual deterioration rate of pavement systems was empirically examined, and the actual effectiveness of various treatment options was researched by using collected data from the GDOT's asset management software program.

The updated decision tree recommending an optimal course of action will be considered by the GDOT's Office of Maintenance for implementation, in order to better consider regional effects, cost variation, and trade-offs among various treatments. The enhanced maintenance and rehabilitation strategy will provide the GDOT with comprehensive tools to better consider regional effects and cost variations in developing cost-effective treatment plans for statewide maintenance and rehabilitation projects. The research deliverables contain detailed information about how to use the developed decision tree to enhance the state of mid- and long-term cost estimating for maintenance and rehabilitation projects.

PAVEMENT DETERIORATION MODEL

The researchers constructed Markov chain pavement deterioration models using empirical pavement inspection data and performed chi-squared tests to analyze dependency between pavement condition transitions and multiple factors. The results demonstrate that the deterioration rate of roadways is dependent on variables including Interstate, AADT, NHS, Number of Lanes, District, and Annual Average Temperature. It is advisable to consider grouping pavement assets based on the variables influencing deterioration probabilities when developing probabilistic models to accurately capture the deterioration behaviors of each group.

The analysis results reveal that the degrading probability of pavement is higher for Interstate roadways compared to Non-Interstate roadways. Additionally, a positive relationship is observed between degrading probabilities and AADT, indicating that as AADT increases, the likelihood of pavement deterioration also increases. For Non-Interstate roadways, the degrading probability shows a decreasing trend as average temperature increased. However, the underlying causation behind the effects of average temperature on degrading probabilities requires further investigation.

Furthermore, the researchers developed classification models and evaluated their capabilities in predicting pavement condition classes. Although individual ML models bring advantages in recognizing patterns and relationships within data, their integration via ensemble techniques such as voting and stacking elevate performance in predicting pavement condition levels. This combination mitigates the errors of individual models, leading to more consistent and accurate pavement condition predictions. The cross-validation results show that ensemble models generally achieve higher accuracy with lower variance than individual ML models for predicting pavement condition classes. Based on the findings of this research, the authors suggest utilizing ensemble approaches in pavement condition prediction models to generate more reliable and consistent predictions.

The findings contribute to the body of knowledge in finding impacts of various factors on pavement deterioration and offering valuable insights for improving asset management plans. The prediction of pavement condition levels through ML classification algorithms and ensemble approaches can help transportation asset management practitioners improve their practices in accurately predicting pavement condition levels. Accurate prediction of pavement conditions will facilitate better planning and timely maintenance interventions, which can lead to substantial savings in terms of maintenance costs. Further research aimed at identifying additional influential variables for pavement deterioration and establishing optimal thresholds for grouping roadways with similar degradation behavior would improve the effectiveness of pavement management and aid in prioritizing maintenance efforts for road networks.

COST ESTIMATION MODEL

Various ML algorithms were employed and compared to estimate the cost per lane mile in highway maintenance projects. The results satisfied state DOTs' three tasks for cost estimation in

asset management planning that were proposed in Chapter 3. Maintenance Cost Estimation Model.

1. The lowest MAPE was 18.513 percent (5-fold) and 16.891 percent (10-fold) from the ET model, below the proposed error of 25 percent in the early stage of project planning.
2. The pavement condition, which in the GDOT's case is OCI grade, was input in the model.
3. The ET model exhibited the potential for seamless and relatively swift integration into long-term planning, enabling the alignment between financial and asset management plans.

The initial stage information and asset information were extracted into three groups of variables, utilized to develop regression models, and performance metrics were recorded and compared across the models. Four models—RF, ET, GB, and XGB—demonstrated potential for practical application in transportation asset management field practices.

Moreover, the researchers aimed to establish a practical model with reasonable, or relatively high, accuracy through data preprocessing, feature selection, and application of tree-based ML algorithms. Four cases of regression models were developed, which consist of RF and ET regression, with both the raw value of the target variable and the log-transformed value of the target variable. The total length, duration, and number of involved maintenance sections of the project had the highest statistical dependence with the project cost. The total length, number of lanes in the section, and AADT had the most potent predictive power in building the model. The performance metrics indicate the superiority of the ET algorithm with MAPE of 15.02 percent, coefficient of the determinant of 91.4 percent, and computation time of 0.266 s. The predicted results demonstrate the potential of the tree-based algorithms in the early-stage highway

maintenance cost forecasting models and long-term cost management in transportation asset management field practices.

The cost estimation in the early stages of project development is used for the initial feasibility and funding studies of long-range programs (Anderson et al. 2006). Although the proposed model in this study satisfies the suggested standards, the competence of the ET algorithm can be further studied with advanced feature selection or its utilization in advanced model structure. Furthermore, this model provides synergy with existing pavement condition prediction models for performing cost-benefit and optimization analysis for highway asset management.

ACKNOWLEDGMENTS

The research reported herein was sponsored by the Georgia Department of Transportation (GDOT) through Research Project 21-03. The authors acknowledge and appreciate the help of Mr. Andy Doyle, GDOT State Maintenance Engineer; Ms. Ernay Robinson, GDOT Assistant State Maintenance Engineer; Mr. Brennan Roney, Research Program/Project Manager; and the many subject matter experts who contributed their time and knowledge.

REFERENCES

Adeli, H. and Wu, M. (1998). "Regularization Neural Network for Construction Cost Estimation." *Journal of Construction Engineering and Management*, *124*(1), pp. 18–24. Available online: https://doi.org/10.1061/%28ASCE%290733-9364%281998%29124%3A1%2818%29.

Ahmad, M.W., Reynolds, J., and Rezgui, Y. (2018). "Predictive Modelling for Solar Thermal Energy Systems: A Comparison of Support Vector Regression, Random Forest, Extra Trees and Regression Trees." *Journal of Cleaner Production*, *203*, pp. 810–821. Available online: https://doi.org/10.1016/j.jclepro.2018.08.207.

Alavi, S., Tavares, M.P., and Nv, R. (2009). *Highway Project Cost Estimating and Management*. Final Report, FHWA/MT-08-007/8189. The State of Montana Department of Transportation, Helena, MT. Available online: https://www.mdt.mt.gov/other/webdata/external/research/docs/research_proj/project_cost/final_report.pdf.

Ali, J., Khan, R., Ahmad, N., and Maqsood, I. (2012). "Random Forests and Decision Trees." *International Journal of Computer Science*, *9*(5), No. 3, pp. 272–278. Available online: https://www.ijcsi.org/papers/IJCSI-9-5-3-272-278.pdf.

Allen, B.W., Zimmerman, K.A., Duncan, G.M., Zilay, R., and Holabaugh, J. (2023). *Incorporating Maintenance Costs into a Transportation Asset Management Plan*. Transportation Research Board, Washington, D.C. Available online: https://trid.trb.org/View/2265767.

Alsharef, A., Banerjee, S., Uddin, S.M.J., Albert, A., and Jaselskis, E. (2021). "Early Impacts of the COVID-19 Pandemic on the United States Construction Industry." *International Journal of Environmental Research and Public Health*, *18*(4), pp. 1–21. Available online: https://doi.org/10.3390/ijerph18041559.

Anderson, S., Molenaar, K., and Schexnayder, C. (2006). *Final Report for NCHRP Report 574: Guidance for Cost Estimation and Management for Highway Projects During Planning, Programming, and Preconstruction*. Project 8-49, National Cooperative Highway Research Program, Transportation Research Board, Washington, DC. Available online: https://doi.org/10.17226/22045.

Ashuri, B., Baek, M., and Li, M. (2022). *Enhancing the Accuracy of Construction Cost Estimates for Major Lump Sum (LS) Pay Items and Generating a More-Accurate List of Pay Items Throughout the Design Development Process*. Final Report, Research Project 20-17, Georgia Department of Transportation, Atlanta, GA. Available online: https://g92018.eos-intl.net/eLibSQL14_G92018_Documents/20-17.pdf.

Asmar, M. El, Hanna, A.S., and Whited, G.C. (2011). "New Approach to Developing Conceptual Cost Estimates for Highway Projects. *Journal of Construction Engineering and Management*, *137*(11), pp. 942–949. Available online: https://doi.org/10.1061/(asce)co.1943-7862.0000355.

Assaad, R., Asce, S.M. and El-Adaway, I.H.. (2021). "Guidelines for Responding to COVID-19 Pandemic: Best Practices, Impacts, and Future Research Directions." *Journal of Management in Engineering*, *37*(3). Available online: https://doi.org/10.1061/(ASCE)ME.1943-5479.0000906.

Attoh-Okine, N.O. (1994). Predicting Roughness Progression in Flexible Pavements Using Artificial Neural Networks. *3rd International Conference on Managing Pavements*, *1*(1), pp. 55–62. Available online: https://onlinepubs.trb.org/Onlinepubs/conf/1994/cp1/cp1-v1-007.pdf.

Bashar, M.Z. and Torres-Machi, C. (2021). "Performance of Machine Learning Algorithms in Predicting the Pavement International Roughness Index." *Transportation Research Record: Journal of the Transportation Research Board*, *2675*(5), pp. 226–237. Available online: https://doi.org/10.1177/0361198120986171.

Battiti, R. (1994). "Using Mutual Information for Selecting Features in Supervised Neural Net Learning. IEEE Trans Neural Network." *IEEE Transactions on Neural Networks*, *5*(4), pp. 537–550. Available online: https://doi.org/10.1109/72.298224.

Berrar, D. (2019). "Cross-Validation." *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, *1*, pp. 542–545. Available online: https://doi.org/10.1016/B978-0-12-809633-8.20349-X.

Bewick, V., Cheek, L., and Ball, J. (2003). "Statistics Review 8: Qualitative Data – Tests of Association. *Critical Care*, *8*(46). Available online: https://doi.org/10.1186/cc2428.

Biörck, J., Blanco, J.L., Mischke, J., Strube, G., Ribeirinho, M.J., and Rockhill, D. (2020). "How construction can emerge stronger after coronavirus." McKinsey & Company, May 8. (website) Available online: https://www.mckinsey.com/capabilities/operations/our-insights/how-construction-can-emerge-stronger-after-coronavirus.

Black, M., Brint, A.T., and Brailsford, J.R. (2005). "Comparing Probabilistic Methods for the Asset Management of Distributed Items." *Journal of Infrastructure Systems*, *11*(2), Available online: https://doi.org/10.1061/(ASCE)1076-0342(2005)11:2(102).

Bock, M., Cardazzi, A., and Humphreys, B.R. (2021). *Where the Rubber Meets the Road: Pavement Damage Reduces Traffic Safety and Speed*. Working Paper 29176, National Bureau of Economic Research, Cambridge, MA. Available online: http://www.nber.org/papers/w29176.

Breiman, L. (2001). "Random Forests." *Machine Learning*, *45*, pp. 5–32. Available online: https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth & Brooks/Cole Advanced Books & Software, New York. Available online: https://doi.org/10.1201/9781315139470.

Cano-Ortiz, S., Pascual-Muñoz, P., and Castro-Fresno, D. (2022). "Machine Learning Algorithms for Monitoring Pavement Performance. *Automation in Construction*, *139*, 104309. Available online: https://doi.org/10.1016/J.AUTCON.2022.104309.

Cao, Y. and Ashuri, B. (2020). "Predicting the Volatility of Highway Construction Cost Index Using Long Short-Term Memory." *Journal of Management in Engineering*, *36*(4). Available online: https://doi.org/10.1061/(asce)me.1943-5479.0000784.

Cao, Y., Ashuri, B., and Baek, M. (2018). "Prediction of Unit Price Bids of Resurfacing Highway Projects through Ensemble Machine Learning." *Journal of Computing in Civil Engineering*, *32*(5), 04018043. Available online: https://doi.org/10.1061/(ASCE)CP.1943-5487.0000788.

Carer, P. (2006). "Probabilistic Methods Used in Asset Management for MV Electrical Equipment at EDF." *2006 9th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, p. 15. Available online: https://doi.org/10.1109/PMAPS.2006.360284.

Chen, T. and He, T. (2024). *xgboost: eXtreme Gradient Boosting*. Version 1.7.7.1. Available online: https://cran.r-project.org/web//packages/xgboost/vignettes/xgboost.pdf.

Cheng, M.Y., Tsai, H.C., and Sudjono, E. (2010). "Conceptual Cost Estimates Using Evolutionary Fuzzy Hybrid Neural Network for Projects in Construction Industry." *Expert Systems with Applications*, *37*(6), pp. 4224–4231. Available online: https://doi.org/10.1016/j.eswa.2009.11.080.

Chou, J.S. (2011). "Cost Simulation in an Item-based Project Involving Construction Engineering and Management." *International Journal of Project Management*, *29*(6), pp. 706–717. Available online: https://doi.org/10.1016/j.ijproman.2010.07.010.

Chou, J.S., Lin, C.W., Pham, A.D., and Shao, J.Y. (2015). "Optimized Artificial Intelligence Models for Predicting Project Award Price." *Automation in Construction*, *54*, pp. 106–115. Available online: https://doi.org/10.1016/j.autcon.2015.02.006.

Dietterich, T.G. (2000). "Ensemble Methods in Machine Learning." International Workshop on Multiple Classifier Systems, *Lecture Notes in Computer Science Series*, *1857*, pp. 1–15. Available online: https://doi.org/10.1007/3-540-45014-9_1.

Fernández, A., García, S., Herrera, F., and Chawla, N.V. (2018). "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary." *Journal of Artificial Intelligence Research*, *61*, pp. 863–905. Available online: https://doi.org/10.1613/JAIR.1.11192.

Fickler, R., Cheim, L., and Lam, B. (2009). "Application of Probabilistic Techniques in Asset Management." *2009 IEEE/PES Power Systems Conference and Exposition, PSCE 2009*, Seattle, WA, pp1–7. Available online: https://doi.org/10.1109/PSCE.2009.4839938.

Flintsch, G.W. and Chen, C. (2004). “Soft Computing Applications in Infrastructure Management.” *Journal of Infrastructure Systems*, *10*(4), pp. 157–166. Available online: https://doi.org/10.1061/(ASCE)1076-0342(2004)10:4(157).

Friedman, J.H. (2001). “Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*. *29*(5), pp. 1189–1232. Available online: https://doi.org/10.1214/AOS/1013203451.

Gandhi, I. and Pandey, M. (2016). “Hybrid Ensemble of Classifiers Using Voting.” *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, Greater Noida, India, pp. 399–404. Available online: https://doi.org/10.1109/ICGCIOT.2015.7380496.

Georgia Department of Transportation. (2023). *Transportation Asset Management (TAM) – Georgia DOT*. (website) Atlanta, GA. Available online: https://www.dot.ga.gov/GDOT/Pages/TAM.aspx, last accessed November 6, 2023.

Georgiou, P., Plati, C., and Loizos, A. (2018). “Soft Computing Models to Predict Pavement Roughness: A Comparative Study.” *Advances in Civil Engineering*, *2018*, 5939806. Available online: https://doi.org/10.1155/2018/5939806.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). “Extremely Randomized Trees.” *Machine Learning*, *63*(1), pp. 3–42. Available online: https://doi.org/10.1007/s10994-006-6226-1.

Guyon, I., and Elisseeff, A. 2003. “An Introduction to Variable and Feature Selection.” *Journal of Machine Learning Research*, *3*(2003), pp. 1157–1182. Available online: https://jmlr.csail.mit.edu/papers/volume3/guyon03a/guyon03a.pdf.

Hegazy, T. and Ayed, A. (1998). “Neural Network Model for Parametric Cost Estimation of Highway Projects.” *Journal of Construction Engineering and Management*, *124*(3), pp. 210–218. Available online: https://doi.org/10.1061/(ASCE)0733-9364(1998)124:3(210).

Hegazy, T., Fazio, P., and Moselhi, O. (1994). “Developing Practical Neural Network Applications Using Back-Propagation.” *Computer-Aided Civil and Infrastructure Engineering*, *9*(2) pp. 145–159. Available online: https://doi.org/10.1111/j.1467-8667.1994.tb00369.x.

Herbsman, Z. (1983). “Long-range Forecasting Highway Construction Costs.” *Journal of Construction Engineering and Management*, *109*(4), pp. 423–434. Available online: https://doi.org/10.1061/(ASCE)0733-9364(1983)109:4(423).

Hu, A., Bai, Q., Chen, L., Meng, S., Li, Q., and Xu, Z. (2022). “A Review on Empirical Methods of Pavement Performance Modeling.” *Construction and Building Materials*, *342*, part B. Available online: https://doi.org/10.1016/j.conbuildmat.2022.127968.

Ilbeigi, M., Ashuri, B., and Joukar, A. (2017). “Time-Series Analysis for Forecasting Asphalt-Cement Price. *Journal of Management in Engineering*, *33*(1). Available online: https://doi.org/10.1061/(asce)me.1943-5479.0000477.

Inkoom, S., Sobanjo, J., Barbu, A., and Niu, X. (2019). “Pavement Crack Rating Using Machine Learning Frameworks: Partitioning, Bootstrap Forest, Boosted Trees, Naïve Bayes, and K - Nearest Neighbors.” *Journal of Transportation Engineering, Part B: Pavements*, *145*(3), 04019031. Available online. https://doi.org/10.1061/JPEODX.0000126.

Justo-Silva, R., Ferreira, A., and Flintsch, G. (2021). “Review on Machine Learning Techniques for Developing Pavement Performance Prediction Models.” *Sustainability (Switzerland)*, *13*(9). Available online: https://doi.org/10.3390/su13095248.

Kaloop, M.R., El-Badawy, S.M., Ahn, J., Sim, H.B., Hu, J.W., and Abd El-Hakim, R.T. (2022). “A Hybrid Wavelet-optimally-pruned Extreme Learning Machine Model for the Estimation of International Roughness Index of Rigid Pavements.” *International Journal of Pavement Engineering*, *23*(3), pp. 862–876. Available online: https://doi.org/10.1080/10298436.2020.1776281.

Kargah-Ostadi, N. and Stoffels, S.M. (2015). “Framework for Development and Comprehensive Comparison of Empirical Pavement Performance Models.” *Journal of Transportation Engineering*, *141*(8), 04015012. Available online: https://doi.org/10.1061/(ASCE)TE.1943-5436.0000779.

Kazemitabar, S.J., Amini, A.A., Bloniarz, A., and Talwalkar, A. (2017). “Variable Importance Using Decision Trees.” *Advances in Neural Information Processing Systems*, 3st Conference, Long Beach, CA. Available online: https://papers.nips.cc/paper_files/paper/2017/file/5737c6ec2e0716f3d8a7a5c4e0de0d9a-Paper.pdf.

Kim, K. and Hong, J. (2017). “A Hybrid Decision Tree Algorithm for Mixed Numeric and Categorical Data in Regression Analysis.” *Pattern Recognition Letters*, *98*, pp. 39–45. Available online: https://doi.org/10.1016/J.PATREC.2017.08.011.

Kotsiantis, S., Kanellopoulos, D., and Pintelas, P. (2007). “Data Preprocessing for Supervised Leaning.” *International Journal of Computer and Information Engineering*, *1*(12), pp. 4104–4109. Available online: https://doi.org/10.5281/zenodo.1082415.

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). “Estimating Mutual Information.” *Physical Review E – Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *69*(6), p. 16. Available online: https://doi.org/10.1103/PhysRevE.69.066138.

Kumari, S., Kumar, D., and Mittal, M. (2021). “An Ensemble Approach for Classification and Prediction of Diabetes Mellitus Using Soft Voting Classifier.” *International Journal of Cognitive Computing in Engineering*, *2*, pp. 40–46. Available online: https://doi.org/10.1016/J.IJCCE.2021.01.001.

Li, M. and Ashuri, B. (2021). “Proportional Cox Hazards Model to Quantify the Likelihood of Underestimation in Transportation Projects.” *Journal of Construction Engineering and Management*, *147*(10). Available online: https://doi.org/10.1061/(asce)co.1943-7862.0002164.

Li, M., Baek, M., and Ashuri, B. (2021). "Forecasting Ratio of Low Bid to Owner's Estimate for Highway Construction." *Journal of Construction Engineering and Management*, *147*(1). American Society of Civil Engineers (ASCE). Available online: https://doi.org/10.1061/(asce)co.1943-7862.0001970.

Marcelino, P., de Lurdes Antunes, M., Fortunato, E., and Gomes, M.C. (2021). "Machine Learning Approach for Pavement Performance Prediction." *International Journal of Pavement Engineering*, *22*(3), pp. 341–354. Available online: https://doi.org/10.1080/10298436.2019.1609673.

Mazari, M. and Rodriguez, D.D. (2016). "Prediction of Pavement Roughness Using a Hybrid Gene Expression Programming-Neural Network Technique." *Journal of Traffic and Transportation Engineering (English Edition)*, *3*(5), pp. 448–455. Available online: https://doi.org/10.1016/J.JTTE.2016.09.007.

Merigó, J.M., Zhou, L., Yu, D., Alrajeh, N., and Alnowibet, K. (2018). "Probabilistic OWA Distances Applied to Asset Management." *Soft Computing*, *22*(15), pp. 4855–4878. Available online: https://doi.org/10.1007/s00500-018-3263-1.

Merrow, E.W., Phillips, K.E., and Myers, C.W. (1981). *Understanding Cost Growth and Performance Shortfalls in Pioneer Process Plants*. Contract No. DE-AC01-79PE70078, U.S. Department of Energy, The Rand Corporation. Available online: https://www.rand.org/content/dam/rand/pubs/reports/2006/R2569.pdf.

Nasteski, V. (2017). "An Overview of the Supervised Machine Learning Methods. *Horizons.B*, *4*, pp. 51–62. Available online: https://www.researchgate.net/publication/328146111_An_overview_of_the_supervised_machine_learning_methods.

Pérez-Acebo, H., Mindra, N., Railean, A., and Rojí, E. (2019). "Rigid Pavement Performance Models by Means of Markov Chains with Half-year Step Time." *International Journal of Pavement Engineering*, *20*(7), pp. 830–843. Available online: https://doi.org/10.1080/10298436.2017.1353390.

Peterson, L. (2009). "K-Nearest Neighbor." *Scholarpedia*, *4*(2), 1883. Available online: https://doi.org/10.4249/scholarpedia.1883.

Polikar, R. (2006). "Ensemble Based Systems in Decision Making." *IEEE Circuits and Systems Magazine*, *6*(3), pp. 21–44. Available online: https://doi.org/10.1109/MCAS.2006.1688199.

Rodríguez, J.D., Pérez, A., and Lozano, J.A. (2010). "Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation." *IEEE Trans Pattern Anal Mach Intell*, *32*(3), pp. 569–575. Available online: https://doi.org/10.1109/TPAMI.2009.187.

Sagi, O. and Rokach, L. (2018). "Ensemble Learning: A Survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1249. Available online: https://doi.org/10.1002/WIDM.1249.

Santero, N.J. and Horvath, A. (2009). "Global Warming Potential of Pavements." *Environmental Research Letters*, *4*(3). Institute of Physics Publishing. Available online: https://doi.org/10.1088/1748-9326/4/3/034011.

Schexnayder, C.J. and Mayo, R.E. (2004). *Construction Management Fundamentals*. McGraw-Hill Professional, Boston, MA.

Shannon, C.E. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Chicago. Available online: https://web.archive.org/web/19980715013250/http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf.

Shirley, C. (2023). *Testimony on the Status of the Highway Trust Fund: 2023 Update*. Chad Shirley, Principal Analyst Microeconomic Studies Division, Before the Subcommittee on Highways and Transit Committee on Transportation and Infrastructure, U.S. House of Representatives, Washington, DC. Available online: https://docs.house.gov/meetings/PW/PW12/20231018/116425/HHRG-118-PW12-Wstate-ShirleyC-20231018.pdf.

Shtayat, A., Moridpour, S., Best, B., and Rumi, S. (2022). "An Overview of Pavement Degradation Prediction Models." *Journal of Advanced Transportation*, *2022*, 7783588. Hindawi Limited. Available online: https://doi.org/10.1155/2022/7783588.

Small, K.A. and Winston, C. (1988). "Optimal Highway Durability." *The American Economic Review*, *78*(3), pp. 560–569. Available online: https://www.jstor.org/stable/1809153.

Sonmez, R. and Ontepeli, B. (2009). "Predesign Cost Estimation of Urban Railway Projects with Parametric Modeling." *Journal of Civil Engineering and Management*, *15*(4), pp. 405–409. Available online: https://doi.org/10.3846/1392-3730.2009.15.405-409.

Teicholz, P. (1993). "Forecasting Final Cost and Budget of Construction Projects." *Journal of Computing in Civil Engineering*, *7*(4), pp. 511–529. Available online: http://dx.doi.org/10.1061/(ASCE)0887-3801(1993)7:4(511).

Titus-Glover, L., Darter, M., and Quintus, H.V. (2019). "Impact of Environmental Factors on Pavement Performance in the Absence of Heavy Loads." Publication No. FHWAHRT-16-078. Federal Highway Administration, McLean, VA. Available online: https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/16084/16084.pdf.

Tsai, J., Wang, Z., and Purcell, R. (2010). *Improving GDOT's Highway Pavement Preservation*. Final Report, Research Project 05-19, Georgia Department of Transportation, Atlanta, GA. Available online: https://www.dot.ga.gov/BuildSmart/research/Documents/05-19b.pdf.

Tu, J.V. (1996). "Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes." *Journal of Clinical Epidemiology*, *49*(11). Available online: https://doi.org/10.1016/s0895-4356(96)00002-9.

Urban Institute. (2022). “Highways and Roads Expenditures.” State and Local Backgrounders Project. (website) Available online: https://www.urban.org/policy-centers/cross-center-initiatives/state-and-local-finance-initiative/state-and-local-backgrounders/highway-and-road-expenditures, last accessed April 21, 2024.

Valles-Valles, D. and Torres-Machi, C. (2023). “Deterioration of Flexible Pavements Induced by Flooding: Case Study Using Stochastic Monte Carlo Simulations in Discrete-time Markov Chains.” *Journal of Infrastructure Systems*, *29*(1). Available online: https://doi.org/10.1061/jitse4.iseng-2109.

Vergara, J.R. and Estévez, P.A. (2014). “A Review of Feature Selection Methods Based on Mutual Information.” *Neural Computing and Applications*, *24*(1), pp. 175–186. Available online: https://doi.org/10.1007/s00521-013-1368-0.

Wang, T., Harvey, J., Lea, J., and Kim, C. (2014). “Impact of Pavement Roughness on Vehicle Free-flow Speed.” *Journal Transportation Engineering*, *140* (9). American Society of Civil Engineers (ASCE). Available online: https://doi.org/10.1061/(ASCE)TE.1943-5436.0000689.

Washington, S., Karlaftis, M., and Mannering, F. (2011). *Statistical and Econometric Methods for Transportation Data Analysis*. (2nd ed.). Chapman and Hall/CRC, New York. Available online: https://doi.org/10.1201/9781420082869.

Wilmot, C.G. and Mei, B. (2005). “Neural Network Modeling of Highway Construction Costs.” *Journal of Construction Engineering and Management*, *131*(7). Available online: https://doi.org/10.1061/(ASCE)0733-9364(2005)131:7(765).

Wolpert, D.H. (1992). “Stacked Generalization.” *Neural Networks*, *5*(2), pp. 241–259. Available online: https://doi.org/10.1016/S0893-6080(05)80023-1.

Wright, S. (1921). “Correlation and Causation.” *Journal of Agricultural Research*, *20*(7), pp. 557–585.

Yang, J., Lu, J.J., Gunaratne, M., and Xiang, Q. (2003). “Forecasting Overall Pavement Condition with Neural Networks: Application on Florida Highway Network.” *Transportation Research Record: Journal of the Transportation Research Record*, *1853*(1), pp. 3–12. Available online: https://doi.org/10.3141/1853-01.

Zhang, M., Gong, H., Jia, X., Xiao, R., Jiang, X., Ma, Y., and Huang, B. (2020). “Analysis of Critical Factors to Asphalt Overlay Performance Using Gradient Boosted Models.” *Construction and Building Materials*, *262*, 120083. Available online: https://doi.org/10.1016/J.CONBUILDMAT.2020.120083.

Zhang, Y., Minchin, R.E., Jr., and Agdas, D. (2017). “Forecasting Completed Cost of Highway Construction Projects Using LASSO Regularized Regression.” *Journal of Construction Engineering and Management*, *143*(10) Available online: https://doi.org/10.1061/(ASCE)CO.1943-7862.0001378.

Zheng, G., Chai, W.K., and Katos, V. (2019). “An Ensemble Model for Short-term Traffic Prediction in Smart City Transportation System.” *2019 IEEE Global Communications Conference, GLOBECOM 2019 – Proceedings*. Available online: https://doi.org/10.1109/GLOBECOM38437.2019.9014061.

Ziari, H., Maghrebi, M., Ayoubinejad, J., and Waller, S. T. (2016). “Prediction of Pavement Performance: Application of Support Vector Regression with Different Kernels.” *Transportation Research Record: Journal of the Transportation Research Board, 2589*(1), pp. 135–145. Available online: https://doi.org/10.3141/2589-15.