IDAHO TRANSPORTATION DEPARTMENT

RESEARCH REPORT

Developing Enhanced Performance Curves of ITD Asphalt Pavements by Mining the Historical Data

RP 293

By

Yong Deng, Xianming Shi

Washington State University

Prepared for

Idaho Transportation Department

ITD Research Program, Planning Services

Highways Development

June 2023

YOUR Safety ••• YOUR Mobility ••• YOUR Economic Opportunity

Disclaimer

Technical Report Documentation Page

1. Report No. FHWA-ID-23-293	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Developing Enhanced Performance Curves of ITD Asphalt Pavements by Mining the Historical Data		5. Report Date June 2023 6. Performing Organization Code
7. Author(s) List all authors. Include ORCID numbers with hyperlinks if known; if so, separate each author with a line break Yong Deng, https://orcid.org/0000-0002-4299-1936 . Xianming Shi, https://orcid.org/0000-0003-3576-8952		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Civil & Environmental Engineering Washington State University 101 Sloan Hall, Pullman WA 99164-2910		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Idaho Transportation Department (SPR) Highways Construction and Operations, Contracting Services, Research Program PO Box 7129, Boise, ID 83707-7129		13. Type of Report and Period Covered Final Report 09/01/20020 – 03/31/2023 14. Sponsoring Agency Code RP-293

15. Supplementary Notes

Project performed in cooperation with the Idaho Transportation Department and Federal Highway Administration and the National Center for Transportation Infrastructure Durability and Life-Extension (TriDurLE).

16. Abstract

Current ITD approach to predict asphalt pavement performance over-simplifies the dynamic and nonlinear nature of pavement deterioration and fails to account for complexities in the pavement condition. As such, this project aimed to develop reliable, realistic and enhanced performance curves for ITD asphalt pavements by mining the historical data. The project consisted of literature review, practitioner survey, assessment of the current ITD pavement performance curves, data processing and modeling for both new and rehabilitated asphalt pavements. Neural networks calibrated with particle swarm optimization achieved desirable prediction performance for asphalt pavement rutting in Idaho using the AASHTOWare Pavement ME Design™ (PMED) data. Gene expression programming models achieved better prediction performance than linear regression and mechanistic-empirical models for four typical distresses – rutting, longitudinal cracking, thermal cracking and roughness of asphalt pavements using PMED data. Deep learning models achieved better prediction performance than statistical models for short-term rutting development of a field asphalt pavement with ITD PMS data than piece-wise regression models for overall condition index of sampled field asphalt pavements. These models can extend the application in terms of the distress type, pavement type and areas of interest. To avoid overfitting and ensure basic rationality of predictive models, statistical methods are necessary to check the stability, robustness, sensitivity, etc. of constructed models before application.

17. Key Words	18. Distribution Statement
Asphalt pavement; distress; rutting; maintenance; neural networks; particle swarm optimization; statistical method; gene expression programming; deep learning; CNN; LSTM; predictive model; reproducibility; robustness; accuracy; performance curve	Copies available from the ITD Research Program Data available on the Washington State University Research Exchange - https://doi.org/10.7273/000005133

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	136	

Acknowledgments

This work is financially supported by the ITD (Project Number 293) and National Center for Transportation Infrastructure Durability & Life-Extension (TriDurLE). The authors would also like to thank the technical support from the ITD Technical Advisory Committee and the Research Office, especially that from the initial project champion Jim Poorbaugh and final project champion Riley Bender, and from the project manager Ned Parrish and senior research analyst Amanda Laib.

Technical Advisory Committee

Each research project is overseen by a Technical Advisory Committee (TAC), which is led by an ITD project sponsor and project manager. The TAC is responsible for monitoring project progress, reviewing deliverables, ensuring that study objectives are met, and facilitating implementation of research recommendations, as appropriate. ITD's Research Program Manager appreciates the work of the following TAC members in guiding this research study.

- Project Sponsor: [Name]
- Project Manager: [Name]
- TAC Members: [List]
- FHWA-Idaho Advisor: [Name]

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

AADT .... Average annual daily traffic

AADTT .... Average annual daily truck traffic

AASHTO .... American Association of State Highway and Transportation Officials

Adam .... Adaptive moment estimation

AI .... Artificial intelligence

AMPT .... Asphalt Mixture Performance Tester

ANN .... Artificial neural network

APA .... Asphalt Pavement Analyzer

ARIMA .... Autoregressive integrated moving average

CNN .... Convolutional neural network

DEM .... Discrete element model

DFR .... Draft final report

DL .... Deep learning

DOT .... Departments of transportation

ES .... Exponential smoothing

FE .... Finite element

FHWA .... Federal Highway Administration

GA .... Genetic algorithm

GEP .... Gene expression programming

GP .... Genetic programming

HWTD .... Hamburg Wheel Tracking Device

IDI .... Individual distress indices

IRI .... International roughness index

ITD .... Idaho Transportation Department

LOWESS .... Locally weighted scatterplot smoothing

LR .... Linear regression

LSTM .... Long short-term memory

LTPP .... Long-Term Pavement Performance

MCP .... Markov Chain Process

MEPDG .... Mechanistic-empirical pavement design guide

ML .... Machine learning

M&R .... Maintenance and rehabilitation

MSE .... Mean squared error

NCAT .... National Center for Asphalt Technology

NCHRP .... National Cooperative Highway Research Program

NMSE .... Normalized mean square error

NN .... Neural network

OCI .... Overall condition index

PCA .... Principal component analysis

PCR .... Pavement condition rating

PMED .... AASHTOWare Pavement ME Design™

PMS .... Pavement management system

PSO .... Particle swarm optimization

Q-Q .... Quantile-quantile

RMSE .... Root mean square error

RNN .... Recurrent neural network

SHRP .... Strategic Highway Research Program

SPLOM .... Scatter plot of matrices

SST .... Superpave Shear Tester

TMS .... Traffic monitoring system

TPM .... Transition Probability Matrix

TRB .... Transportation Research Board

TRID .... Transportation Research Integrated Database

UTC .... University Transportation Center

VLWT .... Vertically Loaded Wheel Tester

Executive Summary

Introduction

Current Idaho Transportation Department (ITD) approach to predict asphalt pavement performance over-simplifies the dynamic and nonlinear nature of pavement deterioration and fails to account for complexities in the pavement condition. Therefore, more accurate and reliable models are needed to predict the damage of ITD asphalt pavements at the project level and enable more accurate and robust optimization of pavement preservation and maintenance and rehabilitation (M&R) decisions at the network level. The overarching goal of this project is to develop reliable and realistic and enhanced performance curves for ITD asphalt pavements by mining the historical data. To this end, the project identifies the appropriate model types, model parameters and additional criteria to use in the enhanced asphalt performance curves. Besides, the project develops, calibrates and validates distress-specific models for forecasting future pavement conditions and compare with current models.

Methodology

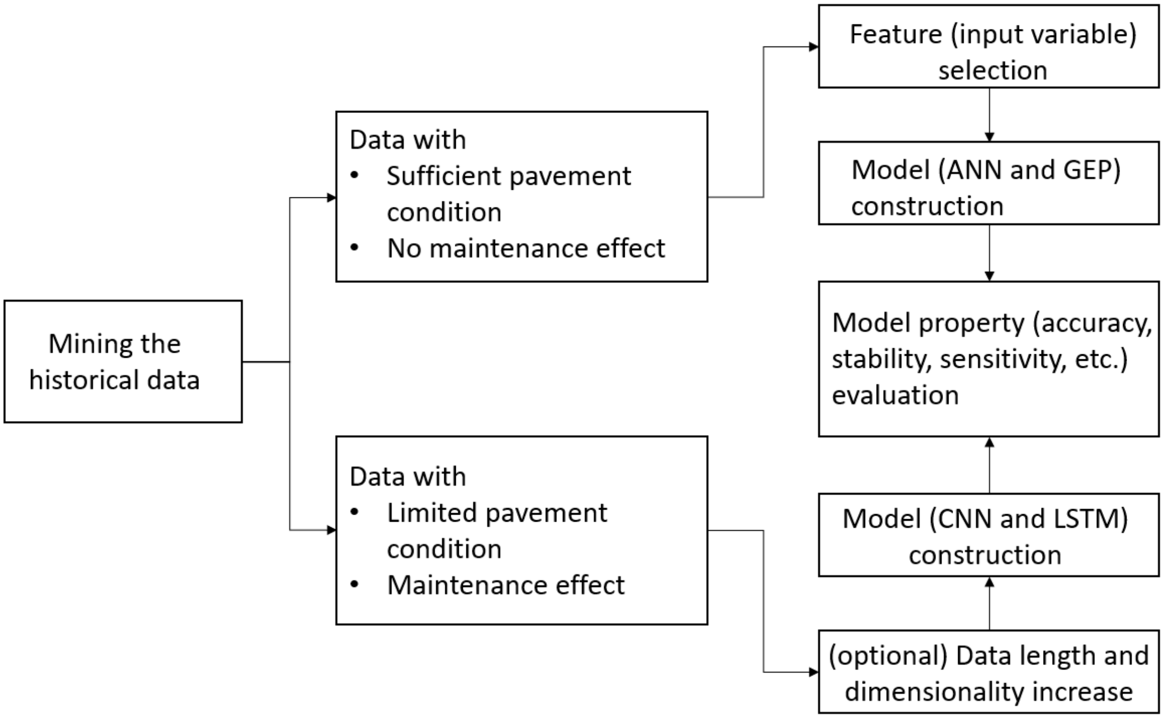
First, the research team conducted a thorough and targeted reviews of literature on asphalt pavement deterioration models. The focus was placed on promising approaches to modeling the performance of asphalt pavements that can be leveraged by this specific study. Meanwhile, a practitioner survey was also conducted to identify performance deterioration models used by various highway agencies for asphalt pavements and to capture the insights and experiences of users on the existing models.

Second, the research team assessed the effectiveness and identify issues with the current ITD pavement performance curves, which helps shape the scope of subsequent tasks.

Third, the research team processed data to screen, reduce, and transform the existing condition and road inventory data in the AASHTOWare Pavement ME Design™ (PMED) and ITD pavement management system (PMS), for both new and rehabilitated asphalt pavements. Potential input (or output) variables were gathered and processed for different model types. Some statistical analyses were conducted to identify significant input factors and their significant interactions for pavement condition, such that these factors were included for further data mining.

Fourth, the research team applied different predictive models on the processed data and entailed the comparison between the actual ITD data against the predictions by the enhanced performance curves of ITD asphalt pavements. Other model properties such as stability, sensitivity, etc. were examined as well to ensure the applicability of applied models.

In general, the data mining and modelling procedures in this project can be summarized as in the following flow.



Main Findings

- Predictive models for field asphalt pavement should consider the characteristics of applied data and have basic accuracy, rationality and simplicity. Characteristics of mechanical, numerical, empirical and machine learning models can be leveraged for higher prediction performance.

- Neural networks (NNs) calibrated with particle swarm optimization have desirable prediction performance for the asphalt pavement rutting in Idaho using the PMED data. Model with proper variable selection and number of hidden neurons can achieve a balance between accuracy, reproducibility and rationality.

- Gene expression programming models have better prediction performance than linear regression and mechanistic-empirical models for four typical distresses – rutting, longitudinal cracking, thermal cracking and roughness of asphalt pavements in Idaho using PMED data.

- Deep learning models achieved better prediction performance than statistical models for short-term rutting development of a field asphalt pavement with ITD PMS data and than piece-wise regression models for overall condition index of sampled field asphalt pavements. Increasing data quantity and dimensionality and model complexity can further improve prediction performance. They can be utilized for data with limited quantity, lacking pavement condition and maintenance effect.

Recommendations

- Characteristics of data should be investigated before applying predictive models. As for pavement performance, the focuses can be put on the availability of material, structure, traffic and environment conditions as well as maintenance effect. Different model types have different applicability and performance for different data.
- Artificial intelligence model (both ML and DL models) and algorithm are promising in predicting pavement performance for the accuracy, efficiency and automation. AI can be applied in model form selection, model calibration, etc. in different ways and to different degrees.
- To avoid overfitting and ensure basic rationality of predictive models, statistical methods are necessary to check the stability, robustness, sensitivity, etc. of constructed models before application.
- Models introduced in this project can extend the application in terms of the distress type, pavement type and areas of interest, with associated codes and instructions provided as deliverables.

1. Introduction

1.1 Problem Statement

The Idaho Transportation Department (ITD) is committed to enhancing its Pavement Management System (PMS) for better management of the state pavement assets. While ITD has implemented the AgileAssets Pavement Analyst software since 2011 (ITD 2019), ITD has recently identified an urgent need to refine the performance curves (i.e., pavement deterioration models). This is because the current practices in the ITD PMS predicts the pavement performance based on fitting piecewise linear function to historic data. Such a time-series forecasting approach often fails to account for the highly nonlinear, dynamic nature of pavement deterioration mechanisms (Prozzi and Madanat 2002, Prozzi and Madanat 2003, Khraibani, Lorino et al. 2012), with the viscoelastic multi-layer pavement materials change with age, temperature, moisture, and traffic loadings. In addition, sometimes multiple mechanisms are at work (Khraibani, Lorino et al. 2012, Xue and Liu 2013, Xue, Liu et al. 2013).

The ITD Asset Management office has found that the current ITD approach over-simplifies the reality and fails to account for complexities in the design, materials selection, construction features, maintenance and rehabilitation (M&R) history, etc. More mechanistic and reliable models are needed to predict the damage of ITD asphalt pavements at the project level and enable more accurate and robust optimization of pavement preservation and M&R decisions at the network level.

ITD has been negatively impacted by the problem of inaccurate performance curves, which has prevented ITD pavement engineers from better understanding the current and future pavement conditions and how to quantify the progression of each distress mechanism (and the remaining pavement life) as a function of design, materials selection, construction, M&R, etc. at the project level. This, in turn, has compromised their ability to identify best practices in these stages of the pavements and to optimize the asset management at the network level.

1.2 Research Background

ITD has collected an abundance of historical data related to the condition of asphalt pavements (a.k.a., flexible pavements), which presents a golden opportunity for data mining and development of enhanced performance curves. Approximately 20 years of pavement condition data is available, with approximately 125,000 records per year for the state PMS. Such datasets have been collected on 1/10-mile segments (both directions), and then aggregated to about 3,500 segments (with length ranging from 0.1 to 3 miles). ITD has reported the pavement condition by common distresses, e.g., rutting, longitudinal roughness, cracking, and raveling (KEI 2015).

PMS is of great importance in meeting the challenges of maintaining a state’s pavement network at serviceable condition (Hill, Cheetham et al. 1991, Gulen, Zhu et al. 2001, Lea and Harvey 2002, Yuan and Mooney 2003, Abaza, Khatib et al. 2004, Li, Cheetham et al. 2006, Hong, Perrone et al. 2017). It embraces the asset management approach and emphasizes the timely preservation, maintenance, and upgrading of pavement assets through structured planning and cost-effective resource allocation. When properly implemented, PMS provides an objective and systematic approach to predict the future condition of the pavement network or a specific section, to illustrate the long-term consequences of different funding levels and M&R policies/activities, and to assist daily decision-making and selecting the optimum M&R strategies.

Pavements deteriorate over time upon construction, due to their exposure to harsh environment and destructive forces. A modeling study of long-term pavement performance (Turtschy and Sweere 1999) indicated that the four common types of distresses causing pavement damage (cracking, rutting, raveling, and longitudinal unevenness) exhibited a gradual increase in the level of distress with time. Such progression of distresses over time coincides with the generic pavement performance curve (Figure 1.1), in which pavement conditions follow a pattern of deterioration exhibiting minimal deterioration at first, and as it ages and experiences increased loading and environmental factors, deterioration continues at an increased rate until pavement reconstruction becomes necessary. As shown in Figure 1.1, various rehabilitation strategies are preferred for different levels of pavement deterioration. Proactive maintenance, in the form of preventive maintenance and minor rehabilitation, is performed during the initiation and early propagation stages of distresses. Reactive maintenance, in the form of major rehabilitation and reconstruction, is often performed on a failing pavement where the distresses have greatly propagated. As a stopgap measure to keep the roadway at an acceptable serviceability, reactive maintenance is seldom cost-effective. Studies (Abaza 2002, Mandapaka, Basheer et al. 2012) have indicated that if the M&R activities took place while the pavement was still in the slow deterioration phase instead of deferring them until the sharp deterioration phase, life-cycle cost might be greatly reduced.

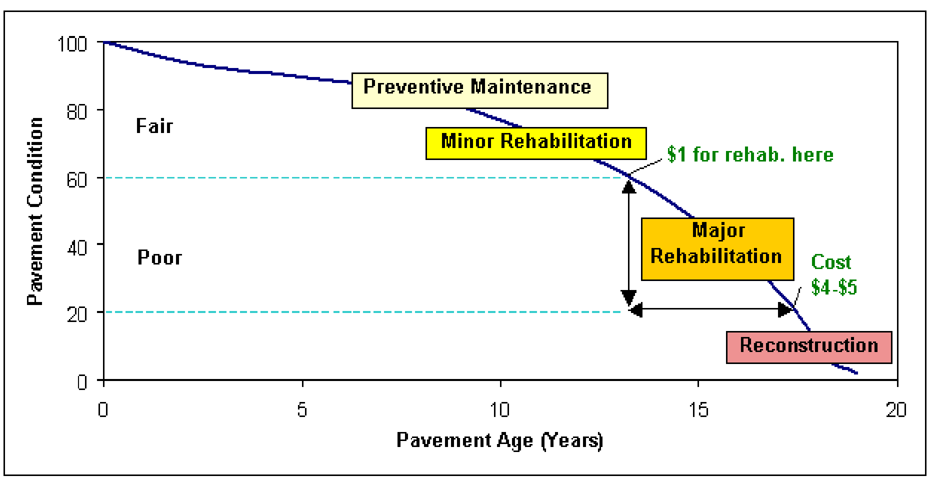


Figure 1.1 A generic pavement deterioration curve illustrating timing for rehabilitation efforts

To optimize pavement M&R plans in terms of life-cycle cost, pavement performance curves should be established to permit the appropriate intervention occur at the right time. As an essential part of PMS, these curves (i.e., deterioration models) predict future pavement conditions based on present condition under a user-defined range of future traffic loading and maintenance scenarios. Figure 1.2 illustrates how performance curves are used to predict future deterioration of an existing pavement, with or without rehabilitation actions. The actual shape of the pavement deterioration curve is a function of many factors including quality of construction, pavement type (rigid, flexible, or composite), pavement material (bituminous or concrete), base course, subgrade soil, deterioration factors (traffic, climate and environmental effects), and M&R history. The forecasted information of life expectancy and benefit, along unit cost of rehabilitation actions, can be further used to perform life-cycle cost analysis, by which the optimum rehabilitation alternative can be identified.

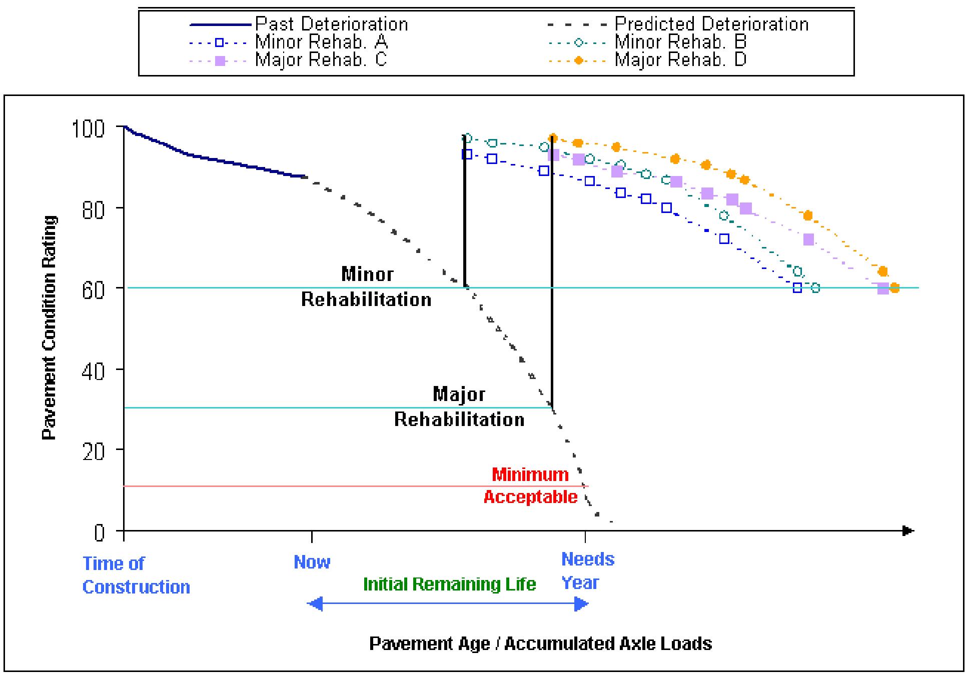


Figure 1.2 Illustration of how performance curves are used to predict future deterioration of an existing pavement, with or without rehabilitation actions

Albeit the great progress made in developing pavement performance curves, current practices still leave ample room for further improvement (Haas 1998, Ningyuan, Kazmierowski et al. 2001, Abaza 2004, Gupta, Kumar et al. 2014). At the network level, many pavement deterioration models attempt to predict the overall performance of pavement sections merely based on their ages, where all sections are grouped into pavement families by pavement type and rehabilitation history, by geographical region and roadway system type, or by more family factors (Kim and Kim 2006). Such models risk over-simplifying the reality and have limitations in accuracy, and thus are not applicable for robust pavement management at the project level. At the project level, most pavement deterioration models focus on specific modes of pavement damage and are developed based on an empirical interpretation of performance data, a mechanistic theory (Collop and Cebon 1995), or a combination of both (Queiroz 1983, Doré, Konrad et al. 1999, Ullidtz 1999, Prozzi 2001, Park and Kim 2003, Li, Pierce et al. 2009). Among them, the mechanistic-empirical (M-E) modeling approach (Bigl and Berg 1996, Timm, Birgisson et al. 1998, Shoukry and William 1999, Li, Pierce et al. 2009) has gained favor as it pools together the advantages of both the mechanistic and empirical modeling, i.e., incorporating the structural response models relating pavement responses (stress, strain, and deflection) to environmental loading conditions and distress models relating actual pavement performance to pavement responses (Bigl and Berg 1996, Timm, Birgisson et al. 1998, Shoukry and William 1999, El-Badawy, Awed et al. 2011, El-Badawy, Bayomy et al. 2012, Seitllari, Lanotte et al. 2019). For instance, El-Badawy et al. (El-Badawy, Bayomy et al. 2012) conducted the local calibration of the mechanistic-empirical pavement design guide (MEPDG) models for dynamic modulus (E*) predictions of asphalt pavements, using 27 HMA mixtures commonly used in the State of Idaho.

Pavement performance can be characterized by individual indicators or aggregate indicators. The most common individual indicators include cracking, roughness, rutting, and skid resistance. Among them, cracking gauges the pavement serviceability from a structural point of view, roughness signals the pavement serviceability in terms of riding quality, whereas rutting and skid resistance implicate safety (Prozzi 2001). As an aggregate indicator, the pavement condition rating (PCR) takes into account both severity and extent ratings of distresses and thus is expected to provide a strong basis for selection and timing of M&R actions and decision-making both at the network level and at the project level.

1.3 Project Goal and Objectives

The overarching goal of this project is to develop reliable and realistic and enhanced performance curves for ITD asphalt pavements by mining the historical data. To this end, the project has the following objectives:

1. identify the appropriate model types to use in the enhanced asphalt performance curves according to the historical data types;
2. identify the appropriate parameters and additional criteria to use in the enhanced asphalt performance curves;
3. develop and calibrate distress-specific models for forecasting future pavement conditions, for both new and rehabilitated asphalt pavements;
4. validate existing and enhanced curves using historical performance data.

1.4 Work Plan

This research project consists of five sequential tasks.

Task 1: Literature review and survey of current practices

This task started with the review of literature relevant to asphalt pavement deterioration models (i.e., performance curves), either for specific distress or for overall condition rating, with the focus on their limitations and input factors. The research team performed a thorough and targeted reviews of literature to identify sources available to gather information pertinent to this study. The focus was placed on promising approaches to modeling the performance of asphalt pavements that can be leveraged by this specific study.

A practitioner survey was also conducted to identify performance deterioration models used by various highway agencies for asphalt pavements and to capture the insights and experiences of users on the existing models, in terms of reliability, precision, input and output parameters, consideration of M&R history, implementation considerations, etc.

This task also includes the review of historical condition data collected by ITD on asphalt pavements, via individual distress indices (IDIs) such as fatigue cracking, transverse cracking, rutting, and International Roughness Index (IRI). Historical data of other variables that may affect pavement performance, i.e., model input factors, was also reviewed, such as the pavement age, design parameters (e.g., pavement thickness), construction features, material properties, climate zone (e.g., temperature and moisture), traffic exposure, maintenance district, and treatment and repair history.

Task 2: Assessing existing curves using historical performance data

In light of the findings from Task 1, the research team assessed the effectiveness and identify issues with the current ITD pavement performance curves (for new and rehabilitated asphalt pavements, respectively), which helps shape the scope of subsequent tasks (e.g., additional criteria needed for the performance curves).

While ITD has implemented the AgileAssets Pavement Analyst software since 2011 (ITD 2019), ITD has recently identified an urgent need to refine the performance curves. This is because the current practices in the ITD PMS predicts the pavement performance based on fitting piecewise linear function to historic data. Such a time-series forecasting approach often fails to account for the highly nonlinear, dynamic nature of pavement deterioration mechanisms (Prozzi and Madanat 2002, Prozzi and Madanat 2003, Khraibani, Lorino et al. 2012); even though the underlying assumption is to learn and model the dynamics of distress evolution using the time-series data of condition ratings (or IDIs).

The ITD Asset Management office has found that the current ITD approach over-simplifies the reality and fails to account for complexities in the design, materials selection, construction features, maintenance and rehabilitation (M&R) history, etc. Nonetheless, in this task, the research team examined the effectiveness and weaknesses of the current ITD performance curves (of asphalt pavements) more comprehensively.

Meanwhile, the AASHTOWare Pavement ME Design™ (PMED), as a comprehensive tool for the analysis and design of new and rehabilitated flexible and rigid pavement structures based on mechanistic-empirical (ME) principles, is applied by most national and local transportation agencies, including ITD

(Bayomy, Muftah et al. 2018). Pavement information considered in PMED include materials, structures, traffic and environment conditions, which is not coupled with corresponding pavement sections in the AgileAssets Pavement Analyst software. Therefore, this task also sought for appropriate model types taking advantages of sufficient pavement information as PMED and improved current ITD performance curves.

Task 3: Data processing and data mining for enhanced performance curves

This task started with a data processing procedure to screen, reduce, and transform the existing condition and road inventory data in the PMED and ITD PMS, for both new and rehabilitated asphalt pavements. Potential input (or output) variables were gathered and processed for different model types. Some statistical analyses were conducted to identify significant input factors and their significant interactions for pavement condition, such that these factors were included for further data mining.

In light of the findings from Task 2, the existing ITD records relevant to the performance of asphalt pavements were sorted by each individual IDI along with the aforementioned model input factors. Depending on the completeness and quality of road inventory and condition data, a sample dataset of ITD asphalt pavement sections were identified for further analyses in this task. Subsequently, statistical techniques were applied to extract useful information from the dataset and to test relevant hypotheses. To utilize the limited historical data effectively, the research team investigated all the explanatory factors, qualitatively evaluated the significance of each explanatory factor, and identified those influencing pavement performance most significantly. The interactions between various explanatory factors were also evaluated. These efforts aimed to better understand and organize the historical performance data and to determine the right type of data for inclusion for further analyses.

The basic requirements for any pavement deterioration model include the following:

- an adequate data base;
- inclusion of significant variables affecting deterioration;
- careful selection of the functional form of the model to represent the physical, real-world relationship;
- careful selection of the calibration algorithm of the model to ensure the model accuracy and construction efficiency;
- criteria to assess the accuracy, stability and rationality of the model.

Based on the inventory and condition data available in PMED and ITD PMS and the findings from the data processing procedure, the research team developed pavement IDI forecast models (i.e., performance curves) of different types. Cutting-edge techniques such as artificial intelligence (AI) algorithms, machine learning (ML) and deep learning (DL) models were utilized in the model development.

Task 4: Validating enhanced curves using historical performance data

This task entailed the comparison between the actual ITD data against the predictions by the enhanced performance curves of ITD asphalt pavements. Once the predictive models were trained and tested (in Task 3), these models were then validated before they can be employed to predict the evolution of IDIs of a specific pavement class in future years (under a given set of input variables). This validation step is needed to ensure that the developed performance curves are reliable and consistent. Specifically, residuals and their distributions were examined via series of performance indicators and hypothesis testing.

Furthermore, a sensitivity analysis was conducted to check how robust the models are to variabilities or errors in the input factor(s). Sensitivity analysis is used to determine how sensitive a model is to changes in the value of the parameters of the model and to changes in the structure of the model (Iooss and Saltelli 2017). In this task, the objective of sensitivity analysis is to find the subset of input variables that are most responsible for variation in each model output and explore their influence trends, etc.

Task 5: Final report, final presentation and training materials

The research team prepared Draft Final Report (DFR) and draft final presentation to the ITD project panel. The DRF documents the entire research effort, results and discussion, key findings/conclusions, and recommendations for further research or implementation. The DFR incorporated findings from Tasks 1 to 4 and was prepared in accordance with ITD guidelines in a format conducive to distribution and adoption (e.g., being Section 503-compliant).

A TriDurLE-sponsored webinar was prepared upon project completion and project panel review, in the effort to facilitate the implementation of the final deliverables of this research. The webinar also serves to engage the stakeholder community (particularly ITD pavement professionals) and benefit the training activities (e.g., how ITD asphalt pavements deteriorate by common modes of distress and how various data mining and modelling techniques can be utilized in their performance curves).

1.5 Report Organization

This report presents the research work completed for developing enhanced performance models of ITD asphalt pavements by mining the historical data. The report is organized into five chapters as described below:

Chapter 1: provides the introduction of this research project, presents the problem statement, research background, project goal and objectives, work plan and report organization;

Chapter 2: presents a literature review of the current pavement performance prediction models for flexible pavements, and summarizes the results of a practitioner survey;

Chapter 3: presents artificial neural network (ANN) models for PMED data, from data collection and processing, model construction and validation, to model comparison with current applied models;

Chapter 4: presents gene expression programming (GEP) models for PMED data, from data collection and processing, model construction and validation, to model comparison with current applied models;

Chapter 5: presents DL models for ITD PMS data, from data collection and processing, model construction and validation, to model comparison with current applied models;

Chapter 6: summarizes the key findings from this project and presents recommendations for future work for ITD consideration.

2. Predictive Models for Asphalt Pavement Performance: State of the Knowledge

2.1 Introduction

Asphalt pavement, or flexible pavement has been widely applied since the 1920s. It is named for its surface layer, which is mainly constructed with aggregates and liquid asphalt. Currently, over 90 percent of pavements in the US are flexible pavements considering their durability, resilience, cost efficiency and eco-friendliness (Huang 2004). Compared with the rigid pavement, the viscous nature of contained asphalt provides the asphalt pavement with more flexibility. Partial energy from the traffic load can be dissipated in the deformation to resist fatigue damage to the pavement (Lytton 2000). A properly designed and constructed flexible pavement can typically last 15-20 years without total replacement. Besides, the construction time and raw material cost of the flexible pavement are lower than the rigid pavement. Moreover, the flexible pavement can be largely recycled and has become an emerging additive to improve the stiffness of the virgin pavement (Luo, Gu et al. 2018).

The current situation is that more than 1/3 of the annual highway budget is spent on the preventative maintenance and rehabilitation of national pavement networks (Juang and Amirkhanian 1992). For longer service time and cost-effective decisions, the performance evaluation and prediction of the flexible pavement should be focused on (Deng, Luo et al. 2019). Additionally, the safety and riding quality of drivers and technicians can benefit significantly from a pavement in good condition. The flexible pavement suffers from synthetic effects of the environment and traffic load (Deng, Luo et al. 2020). Besides, as a multilayer structure made of composite materials, the distress mode and degree can vary with the material composition, structural configuration, and absolutely the environmental and loading conditions, as shown in Table 2.1(a). All these factors make the deterioration of the flexible pavement a complex and highly dynamic process (Khraibani, Lorino et al. 2012).

Continuous efforts have been made on characterizing deteriorations in flexible pavement materials and structures. Accordingly, various models have been proposed for the deterioration evaluation and prediction in terms of individual distress modes or comprehensive performance of the flexible pavement. Table 2.1(b) lists representative national highway research programs in the US, in which performance models of the flexible pavement were proposed, modified, calibrated and/or validated. These major projects are either funded by the Federal Highway Administration (FHWA) or belong to the National Cooperative Highway Research Program (NCHRP) and the Strategic Highway Research Program (SHRP). These projects typically include comprehensive information such as fundamental mechanisms of flexible pavement distress modes, laboratory characterizations from the deterioration initiation, propagation to material failure, field calibrations of deterioration development models and recommendations for the pavement design, maintenance and rehabilitation. Meanwhile, methodologies and models recorded in reports by the local DOTs, papers and standards are also showing ideas, experience and concerns of people in the academia and industry on this topic.

Table 2.1 Representative Distresses and National Projects on Flexible Pavement Performance Models

(a) Representative Distresses

Distress	Description	Causes
rutting	permanent deformation at pavement surface in the wheelpath	repeated traffic load/intermediate and high temperatures/insufficient compaction of asphalt mixtures/shear flow and crack propagation in asphalt mixtures
alligator cracking (bottom-up fatigue cracking)	interconnected cracks at pavement surface	load-induced fatigue damage of asphalt mixtures/weak supporting layers
longitudinal cracking (surface-down fatigue cracking)	cracks parallel to the pavement's centerline or laydown direction at pavement surface	load-induced tensile stresses and strains at layer surface/ load-induced shearing of asphalt mixtures/aging of asphalt mixtures
transverse cracking (thermal cracking)	cracks perpendicular to the pavement's centerline or laydown direction	low temperatures/temperature cycling
roughness*	irregularities at the pavement surface	-

*roughness is not typically treated as an individual distress and depends on rutting, surface-down and bottom-up cracking, and thermal cracking of asphalt pavement.

(b) Representative National Projects

Program	Year	Project Number	Distress Mode
FHWA	1984	FHWA RD-84-018	fatigue damage/rutting
FHWA	1998	FHWA RD-98-132	Roughness
FHWA	2012	FHWA HRT-11-045	rutting/fatigue cracking
NCHRP	1986	NCHRP 01-10	rutting/fatigue cracking
NCHRP	1989	NCHRP 10-26	roughness/rutting/cracking
NCHRP	1996	NCHRP 01-31	Roughness
NCHRP	1998	NCHRP 01-36	fatigue damage
NCHRP	2000	NCHRP 09-20	roughness/rutting/fatigue cracking
NCHRP	2000	NCHRP 10-48	fatigue damage

NCHRP	2003	NCHRP 09-17	Rutting
NCHRP	2004	NCHRP 01-37	bottom-up fatigue (or alligator) cracking/surface-down fatigue (or longitudinal) cracking/rutting/thermal cracking
NCHRP	2005	NCHRP 04-19(2)	rutting/cracking (no model was built)
NCHRP	2006	NCHRP 09-19	Rutting
NCHRP	2007	NCHRP 09-34	moisture damage (rutting and fatigue cracking served as indirect indicators)
NCHRP	2009	NCHRP 01-42	top-down fatigue cracking
NCHRP	2009	NCHRP 09-38	fatigue cracking
NCHRP	2010	NCHRP 01-41	reflection cracking
NCHRP	2011	NCHRP 09-22	rutting/fatigue cracking/thermal cracking
NCHRP	2011	NCHRP 09-33A	rutting/fatigue cracking/thermal cracking
NCHRP	2012	NCHRP 09-30A	Rutting
NCHRP	2013	NCHRP 09-44A	fatigue damage
NCHRP	2016	NCHRP 09-49A	transverse cracking/longitudinal cracking/rutting
NCHRP	2018	NCHRP 01-52	top-down cracking
SHRP	1993	SHRP A-357	fatigue cracking/rutting/thermal cracking
SHRP	1994	SHRP A-404	fatigue damage
SHRP	1994	SHRP A-415	rutting

The main contents of this review include the descriptions of current models for the flexible pavement deterioration evaluation and prediction. Their applications, advantages and limitations are mentioned as well.

2.2 Methodology

The research team conducted a review of several databases to gather relevant information, including: TRID, Google Scholar, ISI Web of Science, etc. In light of the available literature, any gaps in current approaches and potential solutions were identified. The information gathered in this task was used directly as the foundation of the other tasks in the project. For instance, the Mechanistic-Empirical Pavement Design Guide (MEPDG) has continually incorporated the up-to-date understanding by the pavement industry in terms of how each common mode of pavement distress (cracking, rutting, raveling, etc.) initiates and propagates in both new and rehabilitated asphalt pavements, as a function of a wide variety of factors.

Pertinent published literature to each of these topics was searched in the following:

- Publications and ongoing studies in the Transportation Research Integrated Database (TRID), or by the Transportation Research Board (TRB)/National Cooperative Highway Research Program (NCHRP);
- Publications, standards, technical reports, guides, handbooks, and manuals by ITD, Federal Highway Administration (FHWA), and American Association of State Highway and Transportation Officials (AASHTO);
- Publications by academic institutions, such as the University Transportation Centers (UTCs) and National Center for Asphalt Technology (NCAT);
- A review of documents and research in Canada, Europe and other international sources (e.g., World Road Association, i.e., PIARC);
- Other scholarly journal articles, proceedings, technical reports, etc.

A practitioner survey was also conducted to identify performance deterioration models used by various highway agencies for asphalt pavements and to capture the insights and experiences of users on the existing models, in terms of reliability, precision, input and output parameters, consideration of M&R history, implementation considerations, etc. The survey instrument was distributed to list serves such as Pav_Net and TriDurLE_Communications as well as selected state departments of transportation (DOTs). To ensure that the survey instrument receives a sufficient response rate in a timely manner, a user-friendly online survey questionnaire was made and send via email. Along with the information gathered in literature search, the results of the survey helped modify the details of subsequent tasks.

2.3 Mechanical Models and Numerical Models

Mechanical models treat asphalt mixture – the material of flexible pavement surface as a time- and rate-dependent material. It displays responses within four fundamental categories under external excitations – viscoelasticity, viscoplasticity, viscodamage and micro-damage healing (Al-Rub and Darabi 2012). All distress modes are representations of damages in the macro scale, which initiate from micro damages within the material (Lytton 2000). Evidences are the tertiary creep in the rutting test and post-peak behavior of the stress–strain response in the compressive strength test (Lytton 2000, Al-Rub and Darabi 2012, Zhang, Gu et al. 2017). Test results can be better matched by introducing viscodamage models, which take actions from the initiation and propagation of micro-cracks in the previous stages. Mechanical models mainly require material properties and model coefficient values measured and calibrated from laboratory tests, respectively. Calibrated mechanical models can have desirable predictions over new sets of experimental data if applied theories are generalized and advanced enough (Al-Rub and Darabi 2012, Darabi, Al‐Rub et al. 2012, Zhang, Gu et al. 2017).

The major disadvantage of mechanical models is the complexity. The stress state and environmental condition of a field pavement vary with time and location, which results in dynamic analysis and process. Timely decision-making in the pavement maintenance and rehabilitation can hardly be achieved with such a time-consuming method. Currently, pure mechanical models are mainly applied in laboratory tests on asphalt mixture samples in which the environmental and loading conditions are simple and uniform.

Numerical models play an important role in mechanical models and mechanistic-empirical models. For example, finite element (FE) model is a numerical model using the FE method. The FE method provides numerical solutions of partial differential equations, which can describe most engineering problems. The FE method solves the engineering problem of a complex system by dividing the system into finite elements. By solving the equation system assembled by all element equations to the original problem, the solutions at all element points can be obtained. For mechanical models, the FE model is typically built and analyzed in commercial packages such as ABAQUS, ANSYS and COMSOL (Darabi, Al‐Rub et al. 2012, Zhang, Gu et al. 2017). These packages provide a platform to couple multiple material models and solve complex equation systems. Obtained numerical solutions are compared with test measurements to validate mechanical models. Currently, mechanical models are rarely implemented into pavement FE models to predict long-term performance of asphalt pavement considering the computational time and storage space.

For mechanistic-empirical models, the pavement FE model can be built in packages introduced above and those aimed for pavement analysis such as ELLIPAVE, MICHPAVE and EverStressFE. First, pavement responses required in mechanistic-empirical models are elastic or viscoelastic responses. Second, packages such as ELLIPAVE and MICHPAVE simplify the pavement FE model in terms of the structure dimension and/or load configuration. In general, representative pavement responses rather than true pavement responses are applied in mechanistic-empirical models. Other numerical models such as the discrete element model (DEM) are currently limited to simulating laboratory and field tests on small-scale asphalt mixture specimens due to the model assumption, computational time and storage space (Peng 2014, Ma, Zhang et al. 2016, Ma, Zhang et al. 2016, Zhang, Ma et al. 2018).

2.4 Empirical Models and Mechanistic-Empirical Models

Empirical models are typically built without any material or structural responses. Pavement performance is associated with a given set of environmental, material, and loading conditions via regression analysis (Huang 2004). The advantages of empirical models, as opposed to mechanical models, are their simplicity of the model construction and implicit relations between pavement performance and these external factors. For example, Archilla and Madanat (Archilla and Madanat 2000) first identified several factors affecting the rutting development in flexible pavements from extensive literatures, which can be summarized as material properties, vehicle axles, thawing index and load numbers. Then they selected the exponential function from researches on the rutting development in pavements, unbound granular materials and natural soils. The exponential function can characterize the rutting development in the field road tests they studied. Finally, they specified values for model coefficients by performing statistical analysis. Recently, with the development of regression analysis, advanced model forms and regression approaches have been proposed. For example, a nonlinear mixed-effects model was applied in the evaluation and prediction of cracking progression in pavements (Khraibani, Lorino et al. 2012).

The major disadvantage of empirical models is the over-reliance of model coefficient values on the database for model calibration. The constructed empirical models can hardly characterize or predict performance of pavements of which any condition has changed. Therefore, empirical models have extremely limited applications, especially considering the global climate change.

Mechanistic-empirical models are the fast developing and widely applied models for pavement performance evaluation and prediction. They take advantages of mechanical models and empirical models - rationality and simplicity. Pavement responses, mechanical theories, external factors and statistical analysis are involved in mechanistic-empirical models at different degrees. The idea of the mechanistic-empirical approach can date back to the 1950s where the vertical compressive strain on the subgrade surface was used as an indicator for the pavement rutting (Kerkhoven and Dormon 1954, Huang 2004). This example presents the concept of “critical pavement response” which considers the failure criterion of a distress mode and are related to material properties, structural configuration and environmental and loading conditions of the pavement.

Current progress in mechanistic-empirical models are mainly recorded and implemented in the Mechanistic-Empirical Pavement Design Guide and the software AASHTOWare Pavement ME Design (ARA-ERES 2004). The procedures of using mechanistic-empirical models to evaluate and predict pavement performance are presented in Figure 2.1. Accordingly, required information to calibrate a mechanistic-empirical model are shown in Figure 2.1 as well. Inputs and outputs can be found in the laboratory and field test results and databases such as the Long-Term Pavement Performance (LTPP) Database. A pavement distress model typically includes three parts - the mathematical form characterizing the development of a distress mode; model parameters representing pavement responses, material properties, environmental and loading conditions; and model coefficients to be calibrated. As for pavement responses, either a layered elastic solution (JULEA) or the finite element (FE) approach can be used according to the design guide (ARA-ERES 2004). The previous one is a closed-form analytical solution predicting pavement responses at arbitrary locations. The other one is a numerical approach which is introduced in the previous section.

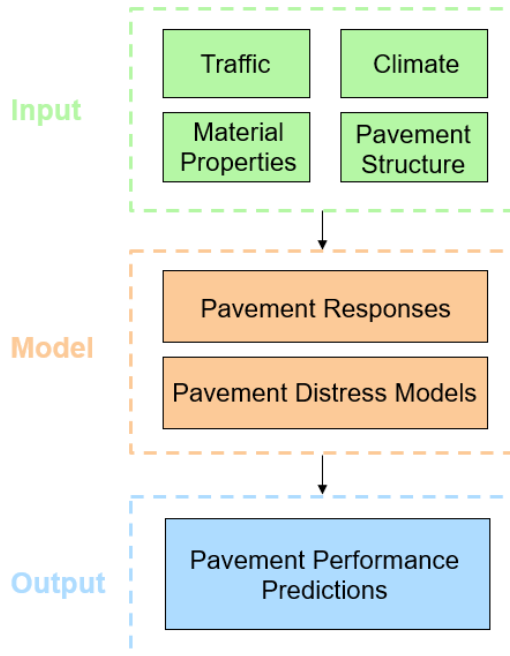


Figure 2.1 Flow of Pavement Performance Evaluation and Prediction Using Mechanistic-Empirical Models Revised from Reference (Lytton, Luo et al. 2018)

2.5 Machine Learning Models and Probabilistic Models

Compared with other model types, machine learning (ML) models are innovative models for pavement performance evaluation and prediction. They are constructed (or trained) by ML algorithms which can improve automatically through experience (Mitchell, Carbonell et al. 2012). ML models with different structures and operations are applied according to types of the problem and data for model training. One of the most applied ML models - artificial neural networks (ANNs) are shown in Figure 2.2. They capture the relationships between inputs and outputs as the biological nervous system. Figure 2.3 illustrates a feedforward NN model for predicting International Roughness Index (IRI) from climatic and traffic data (Hossain, Gopisetti et al. 2017). It is a three-layered architecture including the input layer, hidden layer and output layer. Each block or circle simulates a neuron in the human brain and each line represents the connection between neurons. The number of neurons in the input layer and output layer is determined by the specific problem. The neuron number in the hidden layer and the transfer function connecting neurons should be selected by users. The ANN model automatically adjusts the weight factor of each connection and the bias to the neuron in the model training and validation until the difference between the actual and predicted outputs drops below the threshold or the iteration number goes beyond the threshold. ANNs with different structures and learning algorithms and other ML models such as tree-based models are introduced in the following chapters.

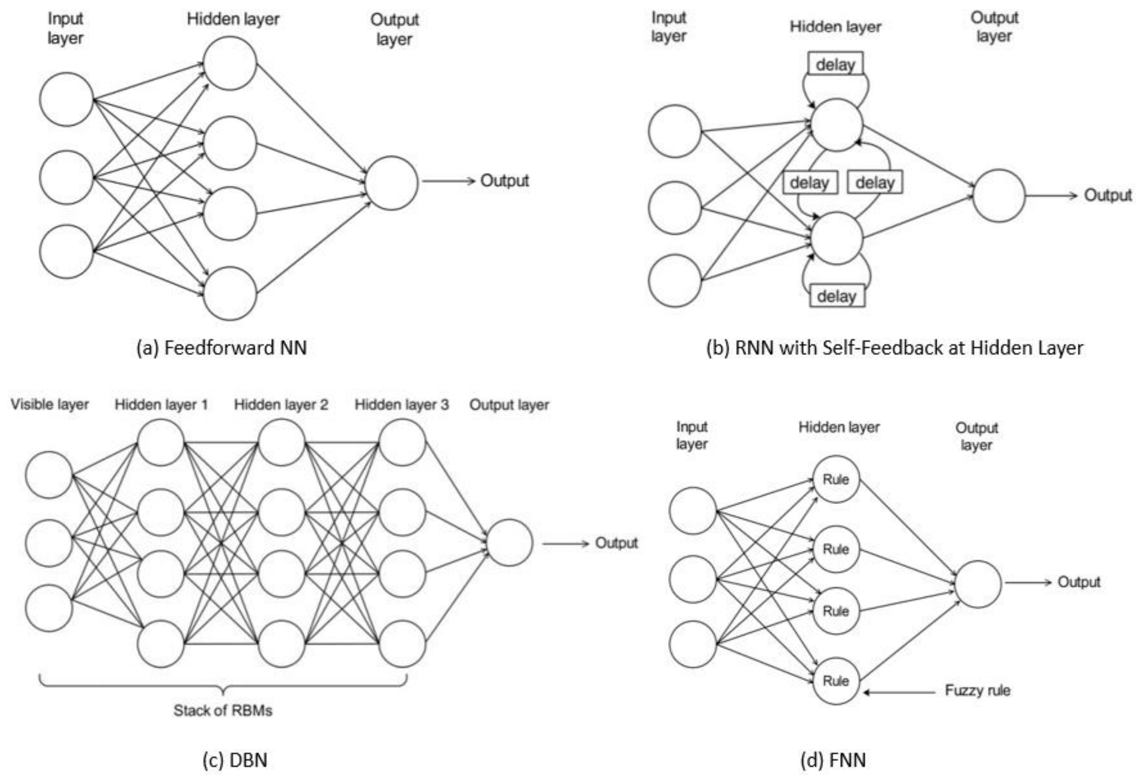


Figure 2.2 Structures of typical NN models (Do, Taherifar et al. 2019). RNN = recurrent neural network, DBN = deep belief network, and FNN = fuzzy neural network

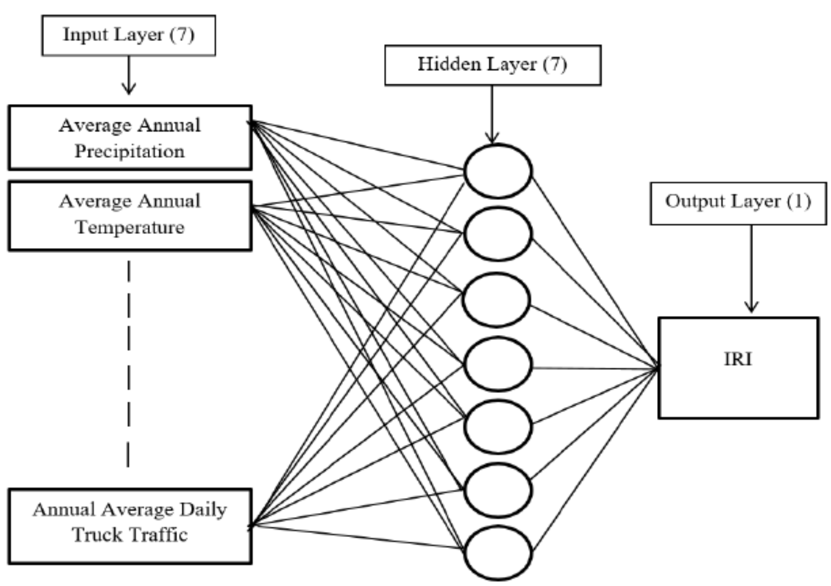


Figure 2.3 Schematic representation of an ANN model (Hossain, Gopisetti et al. 2017)

The main difficulties and limitations of ANN models can be summarized as follows (Ceylan, Bayrak et al. 2014, Hossain, Gopisetti et al. 2017):

- the design of an ANN model includes selections of the hidden layer number, the neuron number in hidden layer(s), the transfer function type, the number and type of inputs, etc. which vary with specific problems;
- ANN models cannot provide explicit relations between inputs and outputs as conventional models;
- the selection of datasets for model training and validation is random. Therefore, the constructed ANN model cannot guarantee the best one;
- as empirical models, a constructed ANN model has limited applications in predicting pavement performance with very different conditions;
- ANN models have possibilities of overfitting.

Models mentioned above can be categorized as deterministic models except that some ML models introduce the probabilistic framework to represent and manipulate uncertainty about models and predictions (Ghahramani 2015). As a comparison, probabilistic models provide a sequence of outputs with corresponding probabilities. The dynamic nature of pavements in terms of the deterioration, environmental and loading conditions and maintenance and rehabilitation (M&R) histories (Alimoradi, Golroo et al. 2020) is considered in such models. Therefore, they are widely applied in predicting comprehensive indices for the pavement condition such as the International Roughness Index (IRI). A representative probabilistic model in the pavement performance modelling is Markov Chain Process (MCP).

In MCP, the time history of the condition index is first divided into multiple condition states. The term transiting the condition index between condition states is called Transition Probability Matrix (TPM) expressed as the following equation,

$$\mathbf{P} = \begin{pmatrix} p_{11}^{t,t+1} & p_{12}^{t,t+1} & \cdots & p_{1(n-1)}^{t,t+1} & p_{1n}^{t,t+1} \\ 0 & p_{22}^{t,t+1} & \cdots & p_{2(n-1)}^{t,t+1} & p_{2n}^{t,t+1} \\ \vdots & 0 & \ddots & p_{3(n-1)}^{t,t+1} & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

$$\sum_{j=1}^{n} p_{ij}^{t,t+1} = 1$$

Where:

$p_{ij}^{t,t+1}$ = the probability that the condition from *i* at state *t* to *j* at state *t+1* which is defined and calculated by users from collected pavement performance data (Yang, Lu et al. 2006)

In MCP, the transition probabilities are assumed constant and the current condition is only relied on the previous one. For example, the IRI at state *t* can be expressed in terms of its initial value as the following equation (Alimoradi, Golroo et al. 2020).

$$\mathbf{IRI}_t = \mathbf{P} \times \mathbf{IRI}_{t-1} = \ldots = \mathbf{P}^t \times \mathbf{IRI}_0$$

MCP requires users to have clear perceptions of the data and pavement condition to deal with tasks such as defining condition indices and partitioning condition time histories. The major limitations of probabilistic models are that they cannot provide explicit forms predicting continuous pavement condition with associated model parameters and time; and those stationary transition probabilities oversimplify the problem and cause systematic error. Such error accumulates in the state transition and reduce the prediction accuracy progressively.

2.6 Comparison of Different Model Types

To better illustrate application and comparison of different model types in predicting pavement performance, rutting is utilized as an example in this section. Rutting or permanent deformation in flexible pavements occurs in both surface and supporting layers. This section introduces rutting in surface layers which are made of asphalt mixtures. Rutting typically accumulates at intermediate and high temperatures and under repetitive traffic loads (Tseng and Lytton 1989). Major laboratory test equipment characterizing rutting development in asphalt mixture samples (cylinders or slabs) include Asphalt Mixture Performance Tester (AMPT) (Dongré, D'Angelo et al. 2009), Hamburg Wheel Tracking Device (HWTD) (Lu and Harvey 2006), Asphalt Pavement Analyzer (APA) (Kandhal and Cooley 2003), Superpave Shear Tester (SST) (Shenoy and Romero 2001), French Pavement Rutting Tester (Romero and Stuart 1998), Georgia Loaded Wheel Tester (Shami, Lai et al. 1997), Vertically Loaded Wheel Tester (VLWT) (Hou, Shi et al. 2018), etc. In these tests, samples are under either repetitive wheel loads or continuous haversine compressive loads. Temperature and load speed/cycle are constant during each test. Test results show the rutting development in asphalt mixtures share a typical shape as shown in Figure 2.4. It can be divided into three distinctive stages based on the acceleration rate. Shape functions capturing the whole or partial curve were utilized in constructing empirical and mechanistic-empirical models. Physical interpretations or hypothesis on the mechanisms of three stages contributed to the theoretical model and parameter selections of mechanical, mechanistic-empirical and machine learning models.

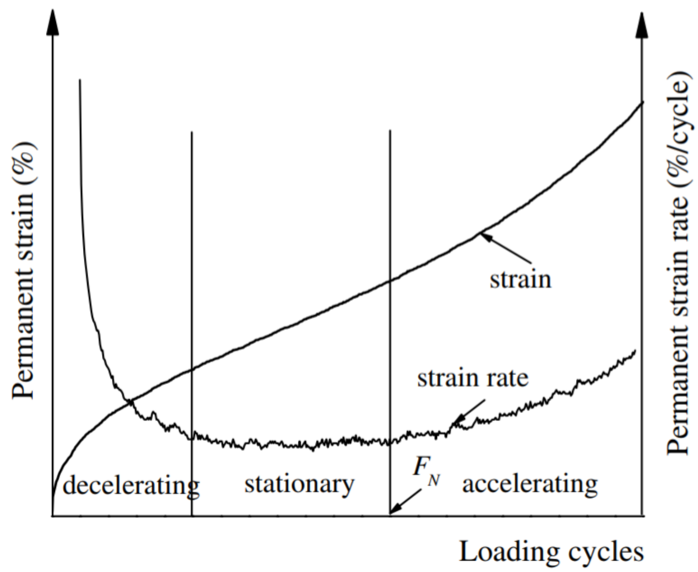


Figure 2.4 Permanent strain and strain rate versus the number of loading cycles (Zhang, Pei et al. 2012)

2.6.1 Mechanical Models and Numerical Models

According to mechanical models, the main contributor to the rutting development in asphalt mixtures is viscoplastic strain. Fundamental components determining the initiation and development of viscoplastic strain are the yield surface function, potential function and constitutive model (Al-Rub and Darabi 2012). The yield surface function, which is the same as potential function in associated viscoplastic models, determines the initiation, rate and direction of viscoplastic strain (Zienkiewicz, Humpheson et al. 1975). It is related to material inherent properties such as the strength and behaviors such as the work-hardening (Lytton 2000). Typical yield surface models for asphalt mixtures include von Mises (Khan and Huang 1995), Mohr–Coulomb (Fwa, Tan et al. 2004), Drucker–Prager (Tan, Low et al. 1994) and their modified versions (Zhang 2012). The constitutive model is responsible for predicting material responses under various environmental and loading conditions based on fundamental mechanics and theories such as thermodynamics (Schapery 1997, Al-Rub and Darabi 2012, Darabi, Al - Rub et al. 2012), energy balance (Zhang 2012, Zhang, Luo et al. 2013), arbitrary Lagrangian-Eulerian (Behnke, Wollny et al. 2019), etc.

As described before, current applications of mechanical models are limited to asphalt mixture samples. As for numerical models of flexible pavements which are implemented with mechanical models of asphalt mixtures, mechanical models are typically simplified. presents examples of flexible pavement numerical models. It can be seen that

- applied mechanical models of viscoplasticity include creep model which is included in the material library of ABAQUS and generalized Kelvin model which typically characterizes viscoelastic materials. Initiation and accumulation of permanent strain rely more on time rather than stress state of the material and exist in the entire service life of the pavement. Characterizations of viscoplasticity as a damage mode of the material are not reflected in these models;

- type, weight and speed variations of traffic vehicles were not considered, which proved to significantly affect the stress/strain state and rutting development (Deng, Luo et al. 2019, Deng, Zhang et al. 2022);
- applications of proposed numerical models in rutting development prediction at a network level are not practical due to their current performance and expenses of ABAQUS.

However, these numerical models proposed techniques in the pavement geometry simplification (Fang, Haddock et al. 2004), load equivalency (Huang, Mohammad et al. 2001, Ali, Sadek et al. 2009) and analysis acceleration (Wu, Chen et al. 2011), which improved the efficiency of computation and analysis, and provides references for related studies. Currently, comprehensive mechanical models described in Section 2.2 have been implemented into FE models of a slab in the wheel tracking test to compare different loading modes (Abu Al-Rub, Darabi et al. 2012) and a pavement structure to conduct the sensitivity analysis (Luo, Li et al. 2020).

Table 2.2 Representative Numerical Models for Rutting Development in Asphalt Pavements

Author	Numerical Models of Pavement and Mechanical Models of Material	Environmental and Loading Conditions	Results
Fang et al. (Fang, Haddock et al. 2004) and Huang et al. (Huang, Mohammad et al. 2001)	• A 2D plane strain FE model (Fang, Haddock et al. 2004) and a 3D FE model (Huang, Mohammad et al. 2001) were built in ABAQUS for pavements; • The creep strain rate was defined as $\dot{\varepsilon} = A\sigma^n t^m$ where σ is uniaxial equivalent deviatoric stress; t is loading time; and A, m and n are parameters obtained from creep tests.	• The load was modelled as quasi-static load with a transverse distribution. Non-uniform contact stress and transverse wheel wander were considered (Fang, Haddock et al. 2004); • A step load was applied which lasted total time of the test loads (Huang, Mohammad et al. 2001).	• A failure criterion was proposed based on the (deformed) pavement surface profile. Predicted failure mode of pavements matched field observations (Fang, Haddock et al. 2004); • Predicted rutting development had a reasonable degree of accuracy with measurements in an accelerated loading facility (ALF) test (Huang, Mohammad et al. 2001).
Ali et al. (Ali, Sadek et al. 2009)	• A 2D axisymmetric FE model was built in ABAQUS for the pavement; • The viscoplastic strain rate considering the time-temperature principle is $\dot{\varepsilon}_{vp} = A_T\sigma^n\left(\frac{t}{a_T}\right)^m$ where σ is deviatoric stress; a_T is temperature shift factor for the viscoplastic effect; t is loading time; and A_T, m and n are constitutive	• The time interval of a vehicular load was calculated by $T_p = \frac{a+b}{v_h}$ where a is the size of mesh element; b is the tire footprint; and v_h is the vehicle speed; • The temperature was measured in the test.	Predicted rutting development matched well with tests, in which the temperature and vehicle speed were constant.

	parameters obtained from full-scale tests.		
Wu et al. (Wu, Chen et al. 2011)	• A 2D axisymmetric FE model was built in ABAQUS for the pavement; • The material was modelled as elastoplastic in the first load cycle and linear elastic in the rest load cycles. The accumulated permanent strain at *n-th* cycle is $\varepsilon_p(N) = \frac{\sigma - \sigma_y}{h} + \sum_{n-1}^{n} \left(\frac{d_n - 1}{d_n} \right) \frac{\sigma}{E_L}$ where σ is cyclic deviatoric stress; σ_y is von Mises yield strength; h is hardening constant; E_L is loading modulus; and d_n is ratio of unloading modulus to loading modulus at *n-th* cycle.	• Constant tire pressures were used for different load levels. The field loading condition was modelled by an accelerated analysis. The permanent deformation after N load cycles is $PD(N) = PD(N_r)\left(\frac{N}{N_r}\right)^B$ where B is slope of the curve of permanent deformation (PD) against number of cycles in a log–log scale, which is obtained from laboratory tests; and N_r is the reference number of load cycles; • The average temperature was utilized for the asphalt layer to calibrate layer modulus. It was measured in the field and adjusted every 25,000 load cycles.	Predicted rutting development needed to be shifted to match field measurements. Shift factors ranged from 0.8 to 1.6.
Li et al. (Li, Huang et al. 2015)	• A 3D FE model was built in ABAQUS for the pavement; • A generalized Kelvin model was utilized for the asphalt mixture as $\varepsilon(t) = \frac{P_0}{20}\left(\sum_{i=1}^{n} \frac{1}{E_i}\left(1 - \exp\left(-\frac{t}{\tau_i}\right)\right) + \frac{t}{\eta_0} + \frac{1}{E_0} \right)$ where $\varepsilon(t)$ is total strain at time t; P_0 is load magnitude; and η_0, E_0, τ_i and E_i are model parameters determined from uniaxial cyclic compression test.	• The contact stress between the tire and pavement surface was decomposed into vertical and tangential stresses; and the movement of vehicular loads was simulated; • The temperature was measured in the field.	FE model provided an acceptable prediction of rutting depth in long and steep section of mountainous highway.

2.6.2 Empirical Models and Mechanistic-Empirical Models

Both empirical and mechanistic-empirical (ME) models include shape functions characterizing entire (Stage I+II+III) or partial (Stage I+II) curve of the rutting development such as polynomial, exponential and multi-staged functions. Compared with mechanical and numerical models, realistic and precise environmental and loading conditions are more convenient to be considered and implemented into empirical and ME models.

Table 2.3 introduces empirical and ME models with either representative forms, parameters or procedures to process field conditions. The fundamental discrepancy between empirical and ME models is that empirical models ignore the role of pavement structure as a system to respond and deteriorate according to external environmental and loading conditions. Asphalt layers of flexible pavements do not deteriorate as asphalt mixture samples in the laboratory. Therefore, material properties utilized for the empirical model calibration (Shell International Petroleum Company 1978, Khedr and Mikhail 1996, Ricardo Archilla and Madanat 2001, Epps 2002, Witczak 2007, Ji, Zheng et al. 2013) may have different effects on different pavement structures. Some structural parameters were considered in empirical models such as the layer thickness (Archilla and Madanat 2000), layer depth (Ji, Zheng et al. 2013) and stress state (Korkiala-Tanttu and Dawson 2007, Ji, Zheng et al. 2013). The first two are too general and the last one proved to be more affected by the loading condition (Deng, Zhang et al. 2022).

Pavement responses included in ME models were either measured (Kenis 1978) or calculated (Deacon, Harvey et al. 2002, Epps 2002, ARA-ERES 2004, Deng, Zhang et al. 2022). In fact, the introduction of the 'mechanistic' part contributes to the 'empirical' part as well. A recent study (Deng, Zhang et al. 2022) pointed out the introduction of pavement responses reduced the dependency of rest regression parameters since pavement responses changed accordingly with environmental and loading conditions. Therefore, a highly nondeterministic regression analysis for traditional empirical models can be simplified.

As for the dynamic nature of field temperature and traffic load, the service time of the pavement was partitioned. Temperatures were averaged (Ji, Zheng et al. 2013) or represented by extreme ones (Archilla and Madanat 2000); and traffic load was categorized (Kenis 1978, Archilla and Madanat 2000) or converted to the standard one (Khedr and Mikhail 1996, Epps 2002, Witczak 2007). Moreover, a statistical model for the wheel wander was considered for a more representative loading condition as the field (ARA-ERES 2004). Accumulated rut depth required transfer to the current time period, which is also a method considering the dynamic nature of field conditions (ARA-ERES 2004, Ji, Zheng et al. 2013).

Improvements for empirical and ME models can be made on modelling the variation of traffic load speed for the increasing consideration of viscoelastic models for the asphalt layer (ARA-ERES 2004, Deng, Zhang et al. 2022). Besides, deteriorating pavement models can be implemented into ME models for more representative pavement responses.

Table 2.3 Representative Empirical and Mechanistic-Empirical Models

Model Type	Shape	Author	Model Form	Loading and Environmental Conditions
Empirical	Two-Stage	Shell Method (Shell International Petroleum Company 1978)	$RD = kh\frac{\sigma_0}{S_{mix,v}}$ where k is the product of a dynamic factor and a configuration factor; h is the layer thickness; σ_0 is the constant stress of the standard wheel; and $S_{mix,v}$ is the mixture stiffness under rutting condition.	The mixture stiffness depends on the stiffness of its contained binder, $S_{bit,v} = \frac{3\eta_0}{W_{eq}t_w}$ where η_0 is the binder viscosity at the average paving temperature during pavement service life; W_{eq} is the number of standard wheel passes from the traffic spectrum; t_w is the wheel loading time related to the traffic speed.

Empirical	Two-Stage	Khedr and Mikhail (Khedr and Mikhail 1996)	$\varepsilon_p = AN^{-m}$ where ε_p is the permanent strain; *A* and *m* are model parameters.	• Environmental condition is reflected on the model parameter *A*, $A = J\left(\frac{\sigma}{E}\right)^S$ where σ is deviator stress; *J* and *S* are material constants; *E* is resilient modulus of asphalt mixture; • Traffic load was represented by the equivalent single axle load (ESAL) in the study.
Empirical	Two-Stage	Archilla and Madanat (Archilla and Madanat 2000, Archilla 2006)	$RD_{it} \approx \beta_{i10} + \sum_{s=1}^{t} a_i \exp\left[\beta_8\left(\frac{TI_s}{1000}\right)\right]\beta_9 \frac{\Delta N_{is}}{N_{is}^{1-\beta_9}}$ where RD_{it} is rut depth for section *i* at time *t*; N_{is} is the variable representing the cumulative number of load repetitions applied to pavement section *i* up to time period *s*; β_{i10} is rut depth immediately after construction for pavement section *i*; TI_s is the thawing index during time period *s*; β_8 and β_9 are the thawing index factor and N_{is} exponent; a_i is a correction factor.	• TI_t is related to mean minimum and maximum temperatures of the period; • ΔN_{is} is related to the loads in front axle of the vehicle, in single load axle(s) of the vehicle and in tandem load axle(s) of the vehicle, number of load axles and standard axle load； • This model was modified for the WesTrack Road Test in another study by the authors (Ricardo Archilla and Madanat 2001). Predicted rut depth accumulated with respect to the exponential of load repetitions. Material properties such as voids filled with asphalt (VFA) were involved in the model as well.
Empirical	Two-Stage	Epps (Epps 2002)	$\ln(rd) = -6.1651 + 0.30991\ln(ESAL) + 0.00294305V_{air}^2$ $+0.0688276P_{asp}^2 - 0.0657803P_{asp}P_{200} + 0.600498(fine-plus)$ $-1.59167(coarse) + 2.35276(replace)$ $+0.21327\ln(ESAL)(coarse) - 0.140386\ln(ESAL)(replace)$ where *rd* is rut depth; *ESAL* is equivalent single axle load; V_{air} is air void content; P_{asp} is asphalt content; P_{200} is percent aggregate finer than No. 200 sieve; *fine-plus*, *coarse* and *replace* are variables which take the value of unity in the fine plus, coarse, or replacement mixes.	The regression model derived from the WesTrack test and served as Level-1 model. The rut depth was related to the load repetition and material properties obtained from laboratory tests.
Empirical	Two-Stage	Witczak (Witczak 2007)	The field rut depth *Rut* was associated with the flow number F_n from the repeated load test and the permanent strain ε_p from the repeated load permanent deformation test: $\log(Rut) = \log(Rut_{1,000,000}) - 0.002(\log(ESAL))^2$ $+0.2815(\log(ESAL)) - 1.6079$ where *ESAL* is equivalent single axle load; $Rut_{1,000,000}$ is the rut depth at 1 million ESALs.	• The flow number F_n from the laboratory repeated load test should be converted to the temperature and traffic level in the field; • Materials and test results from FHWA-ALF and WesTrack tests were utilized to derive this regression model (Sullivan 2002).

Empirical	Two-Stage	Ji et al. (Ji, Zheng et al. 2013)	$RD = 6.714\times10^{-11}\left(N\right)^{0.6274}\left(T\right)^{5.2702}\left(d\right)^{0.5542}\times \left(\tau/\left[\tau\right]\right)^{1.9279}\left(v_d/20\right)^{-(m+1)}$ where *RD* is the rut depth after *N* number of load repetitions; *T* is pavement temperature; *d* is pavement depth; *τ* and *[τ]* are the maximum shear stress of asphalt layers and shear strength of asphalt mixtures; v_d is vehicle speed; *m* is the creep coefficient of asphalt mixtures obtained from laboratory tests.	• This regression model derived from an ALF test. The rut depth was calculated each month and the monthly load repetitions should be first adjusted according to the average monthly pavement temperature and added to the previous ones; • A similar model was proposed and calibrated in a previous study (Kim, Lee et al. 2017). In this study, the rut depth accumulation followed the method proposed by Deacon et al. (Deacon, Harvey et al. 2002) and was validated with independent field rutting performance data.
Empirical	Three-Stage	Zhou et al. (Zhou, Scullion et al. 2004)	A three-stage model was proposed for similar rutting development observed in (accelerated load facility) ALF tests: $\begin{cases}\varepsilon_p = aN^b, N < N_{PS}\\ \varepsilon_p = \varepsilon_{PS} + c(N - N_{PS}), N_{PS} \le N \le N_{ST}\\ \varepsilon_p = \varepsilon_{ST} + d(e^{f(N-N_{ST})} - 1), N \ge N_{ST}\end{cases}$ where ε_p is permanent strain; ε_{PS} and N_{PS} are the permanent strain and number of load repetitions corresponding to the initiation of the secondary stage; ε_{ST} and N_{ST} are the permanent strain and number of load repetitions corresponding to the initiation of the tertiary stage; *a*, *b*, *c*, *d* and *f* are model coefficients.	• ALF tests indicated possible occurrence of the third stage of rutting development in the field; • This proposed model was utilized in a laboratory repeated load test on field samples, in which the environmental and loading conditions were constant.
Empirical	Three-Stage	Korkiala-Tanttu and Dawson (Korkiala-Tanttu and Dawson 2007)	$\varepsilon_p = aN^b + A\left(\frac{N}{1000}\right)^B - C\left(e^{D(N/1000)} - 1\right)$ where ε_p is permanent strain; *N* is the number of load repetitions; *a*, *b*, *A*, *B*, *C* and *D* are regression parameters.	This model was utilized in a heavy vehicle simulator (HVS) test on a full-scale pavement. The environmental and loading conditions were kept constant.
Mechanistic-Empirical	Two-Stage	Kenis (Kenis 1978)	$R_p(N) = R_4(d/2)\mu_{sys}N^{-\alpha_{sys}}$ where $R_p(N)$ is the permanent deformation at load repetition *N*; $R_4(d/2)$ is the general deflection response of pavement surface as a function of load duration and temperature; μ_{sys} is a system rutting characteristic representing the fractional part of the general response that becomes permanent; α_{sys} is a system rutting	According to an application of VESYS model (Zhou and Scullion 2002), the rut depth in the asphalt layer can be calculated by $R_p(N) = \sum_{i=1}^{n}\int_{N_1}^{N_2}\left(U_i^+ - U_i^-\right)\mu_{sys}N^{-\alpha_{sys}}dN$ where U_i^+ and U_i^- are deflections at top and bottom of *i*-th finite layer due to axle group.

			characteristic representing the rate of change of permanent deformation.	
Mechanistic-Empirical	Two-Stage	Deacon et al. (Deacon, Harvey et al. 2002)	$RD = K\gamma_j^i$ where K is the model parameter; γ_j^i is the plastic strain at the j-th hour of trafficking.	• The accumulation of plastic strain is $\gamma_j^i = a_j\left[\left(\frac{\gamma_{j-1}^i}{a_j}\right)^{1/c} + \Delta n_j\right]^c$ $a_j = a\exp(b\tau)\gamma_j^e$ $\gamma_1^i = a_1\left[\Delta n_1\right]^c$ where γ_j^e is elastic shear strain at the j-th hour; Δn_1 and Δn_j are numbers of axle load repetitions applied during the first and j-th hour, respectively; a, b and c are model coefficients; • This model served as Level-2 model for rutting development in the WesTrack test. In the test, traffic loads were uniformly distributed throughout a 24-hr period and the yearly temperature environment was assumed to be the same for each year of the 10-year period (Epps 2002).
Mechanistic-Empirical	Two-Stage	ARA-ERES (ARA-ERES 2004)	• The relation between the permanent strain and resilient strain derived from the laboratory test and was modified for the field: $\frac{\varepsilon_p}{\varepsilon_r} = \beta_{r1}10^{-3.15552}T^{1.734\beta_{r2}}N^{0.39937\beta_{r3}}$ where ε_p is the permanent strain; ε_r is the resilient strain; T is the temperature; N is the number of load repetitions; β_{r1}, β_{r2}, and β_{r3} are calibration factors; • The total rut depth is the sum of ones in divided sublayers: $RD = \sum_{i=1}^{N}\varepsilon_{p_i}\Delta h_i$ where ε_{pi} is the permanent strain at i-th sublayer; Δh_i is the thickness of i-th sublayer.	• The resilient strain is calculated from the layered elastic solution: $\varepsilon_r = \frac{1}{\lvert E^*\rvert}\left(\sigma_z - \mu\sigma_x - \mu\sigma_y\right)$ where $\lvert E^*\rvert$ is the dynamic modulus magnitude of asphalt mixtures; μ is Poisson's ratio; σ_z, σ_x and σ_y are stresses at vertical and other two directions at a given depth; • Rutting development has different curves according to the temperature and stress/strain state of the evaluation period. Accumulated rut depth (permanent strain) should be transferred to the curve for the current period first.
Mechanistic-Empirical	Two-Stage	Deng et al. (Deng, Zhang et al. 2022)	The relation between the permanent strain and resilient strain was proposed for unbound granular materials (Lytton, Luo et al. 2019) and extended to asphalt mixtures: $\frac{\varepsilon_p(N)}{\varepsilon_r} = \gamma_\infty \times e^{-(\frac{\rho}{N-N_0})^\beta}(\frac{\sqrt{J_2}}{Pa})^m(\frac{\alpha I_1 + K}{Pa})^n$ where $\varepsilon_p(N)$ is the permanent strain at load repetition N; ε_r is the resilient strain; I_1 and J_2 are the first invariant of the stress tensor and the second invariant of the deviatoric	This model was utilized in a VLWT test on multi-layered asphalt mixture structures, in which the environmental and loading conditions were constant.

			stress tensor; Pa is atmosphere pressure; γ_∞, ρ, β, m, n, K and N_0 are model parameters.	
Mechanistic-Empirical	Three-Stage	Deng et al. (Deng, Zhang et al. 2022)	The relation between the permanent strain and resilient strain derived from a three-stage empirical model for the HWTT test (Yin, Arambula et al. 2014) and was modified by introducing the structural response ε_r: $$\frac{\varepsilon_p(N)}{\varepsilon_r} = \rho\left(\ln(\frac{N_\infty}{N - N_0})\right)^{-1/\beta}$$ where $\varepsilon_p(N)$ is the permanent strain at load repetition N; ε_r is the resilient strain; T is the temperature; ρ, β, N_0 and N_∞ are model parameters.	This model was utilized in a VLWT test on multi-layered asphalt mixture structures, in which the environmental and loading conditions were constant.

2.6.3 Machine Learning Models

Construction of a ML model for rutting development includes collection and organization of material, structure and pavement performance data for representative model inputs and outputs. Meanwhile, the selection of ML algorithms according to the requirements and characteristics of the problem and data is important.

Alharbi (Alharbi 2018) applied an ANN model with one hidden layer to predict rutting index from pavement age, thickness, average temperatures, etc. Compared with linear regression models, trained ANN improved the prediction accuracy (R^2) by 75.61%. Gong et al. (Gong, Sun et al. 2018) applied two ANN models to compare predicted total rut depth with the transfer function in the Pavement ME Design Guide (ARA-ERES 2004). The first ANN model applied one hidden layer and individual rut depth in the AC layer, base layer and subgrade as inputs. The second ANN model applied two hidden layers and additional 18 material, structural, environmental, traffic and time parameters as inputs. As a comparison, two linear regression models were built with identical inputs as ANN models to represent the transfer function in the Pavement ME Design Guide. Results showed two ANN models improved the prediction accuracy (R^2) by 22% and 88%. Moreover, by using the random forest algorithm, the relevancy of each input to the total rut depth was measured and ranked. Amin and Ajakaiye (Amin and Ajakaiye 2020) applied an ANN model with two hidden layers to predict maximum rut depth from the information of traffic, climate, time and pavement surface condition and profile. A total of 638 road segments were utilized and contributions of all inputs were evaluated by sensitivity analysis.

Examples introduced above are the feedforward ANN. As a descendant of feedforward ANN, each neuron in the hidden layer(s) of a recurrent neural network (RNN) can send produced output to itself. In the time scale, a neuron at each time step is triggered by the output from the previous step and the input for this step (Géron 2019). Obviously, RNN models are suitable for modeling time series data since they can remember and pass information through time (Okuda, Suzuki et al. 2018). Okuda et al. (Okuda, Suzuki et al. 2018), Choi and Do (Choi and Do 2020) trained RNN models to predict rut depth from time-series data of traffic, climate and inspection history. Good agreements were achieved between predicted and measured rut depths.

2.7 Practitioner Survey

This project designed a practitioner survey to identify performance deterioration models used by various highway agencies in the United States for asphalt pavements and to capture the insights and experiences of users on the existing models, in terms of reliability, precision, input and output parameters, consideration of M&R history, implementation considerations, etc. Approved by the project panel, the survey instrument was distributed to list serves such as Pav_Net and TriDurLE_Communications as well as selected state departments of transportation (DOTs). The complete version of the survey is posed in Appendix A. Table 2.4 presents a summary of the technical questions asked in the survey. The survey was delivered online via the platform Qualtrics® since March, 2021. As of the date of this report, a total of 43 effective responses were collected from 23 states of the United States.

Table 2.4 Summary of Technical Questions in the Survey

Questionnaire topics
• Specific distresses concerned in the applied models. (Q1) • Resources of applied models. (Q2) • Limitations of applied models. (Q3) • Inputs of applied models including the name, difficulties in the usage, etc. (Q4-Q10) • Purposes of the applied models. (Q11-Q13) • Performance of the applied models. (Q14-Q19) • Management of the applied models including the quality check, improvements, etc. (Q20-Q21) • Expectations of applied models. (Q22-Q23) • Opinions on the artificial intelligence models (Q24-Q25)

It was reported that rutting (15.66%), roughness (15.66%), transverse cracking (14.46%), longitudinal cracking (13.86%) and alligator cracking (13.86%) were the five distresses concerned most by the researchers and technicians in DOTs, according to a total of 166 choice counts. Therefore, data of these distresses and associated variables are focused and collected for the predictive models in this project.

Currently, the tools developed or purchased by individual agencies are the most popular choices for the pavement distress development prediction and management. Those tools include professional statistical packages such as R, business analytics services such as Power BI, and basic data visualization and analysis tools such as Excel spreadsheet, etc. The software built upon the AASHTO mechanistic-empirical (ME) pavement design guide - AASHTOWare Pavement ME Design ranked second in the survey. Considering the variety of applied tools, their limitations provided by the participants are quite scattered, from data quality to software update.

This survey paid attention to the data especially the model inputs and made corresponding questions. Following the mainstream predictive models such as the ME models, the model inputs were divided into four categories - traffic, climate, material and structure. Figure 2.5 shows their necessities in the predictive models and difficulties to be obtained according to the user experience of the participants. It was found that the traffic information was more difficult to be obtained than the climatic information resulting from the lack of the traffic monitoring system (TMS) in certain areas. As a comparison, climate data are more accessible from national weather databases and services. It was interesting to notice that the information on pavement structures and materials was believed to be important and necessary in the predictive models. However, a portion of the participants reported that variables of these two categorizes were not considered in the models or systems they currently applied. For those variables which are difficult to be obtained, the typical solutions include referring to recommended values in the systems, papers and reports, and using models without them.

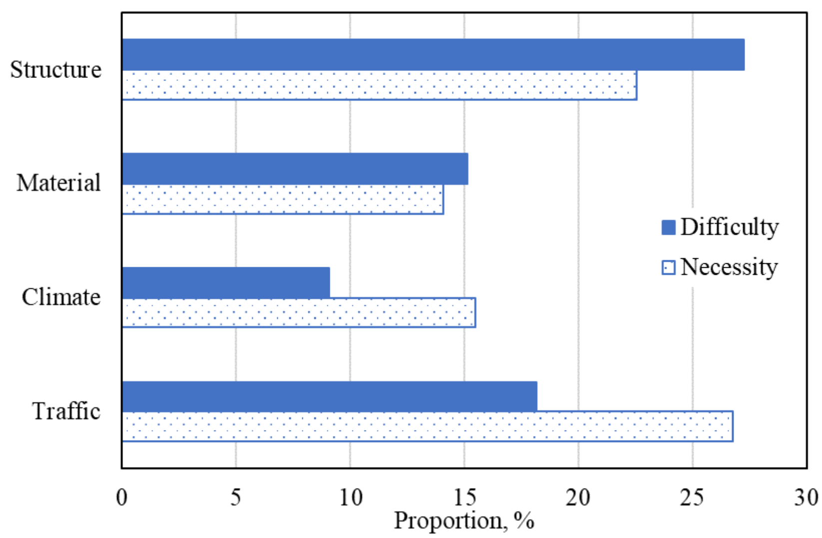


Figure 2.5 Information of four major model inputs

According to the survey responses, the main purpose of using these predictive models was to obtain distress indices for the pavement management. Therefore, the applied predictive models were expected with high qualities. Figure 2.6 shows the top five qualities of a good predictive model voted by the participants. It can be summarized that the accuracy, complexity and applicability are concerned most by those model users. Specifically, 80 percent of the participants expected the predictive models with the accuracy (R^2) over 0.80. Nearly 50 percent of the participants believed that the main factor causing the poor predictions was the limited data for the model calibration. The solutions they can think of include increasing data amount for the mode construction and the frequency of the model validation, performing outlier reviews, etc. As for the model reliability and ruggedness, they were proposed based on the experience of the participants in obtaining very different predictions in pavements with similar conditions.

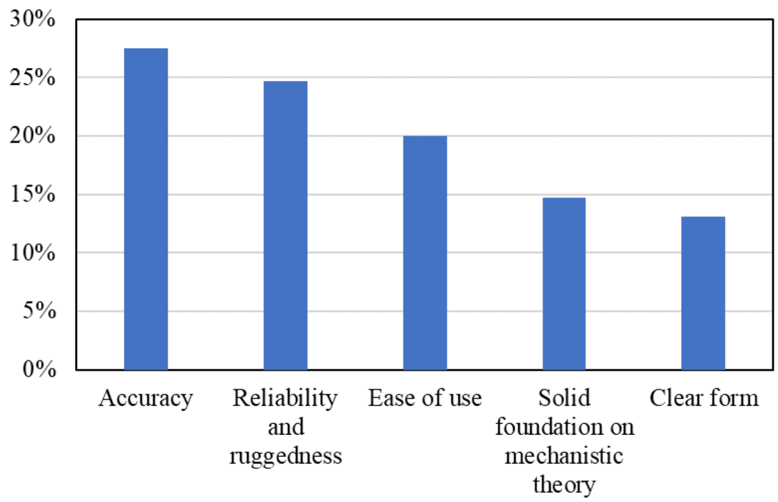
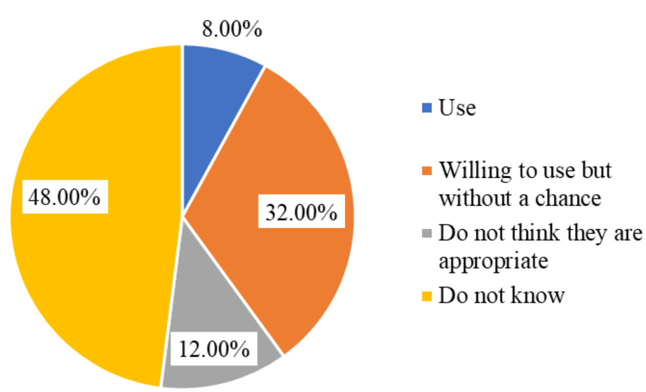
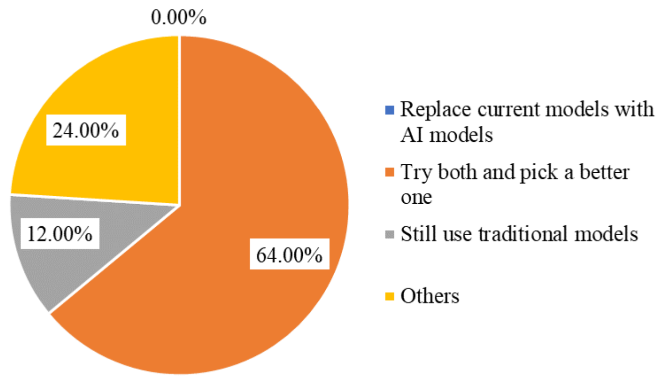


Figure 2.6 Top 5 model qualities

As the models to be developed in this project, questions on the artificial intelligence (AI) models were made in the survey. Responses of the participants on the knowledge of and attitude towards the AI models are presented in Figure 2.7. It was found that over 90 percent of the participants did not use AI models as the predictive models and half of the participants had no idea what the AI models were. However, it is promising that 32 percent of the participants showed their interests in using AI models as their predictive models and 64 percent the participants was willing to try after the comparison with traditional models. Therefore, it is worthwhile to develop and promote AI models in this project as the predictive models for pavement distresses.



(a)



(b)

Figure 2.7 Responses of the participants on the AI models (a) knowledge; and (b) attitude

2.8 Conclusions

In Chapter 2, a literature review on current predictive models of asphalt pavement performance was conducted. Specifically, rutting development was utilized as an example to compare different model types. A practitioner survey on the insights and experiences of users on the existing models by various highway agencies in the United States was analyzed. The major findings in this study can be summarized as follows.

- Mechanical model can have desirable predictions with applied theories generalized and advanced enough. However, for its complexity and time consumption, applications on field pavement sections for long time performance are limited;
- Empirical model has the advantages such as simplicity of the model construction and implicit relations between pavement performance and influencing factors. However, the over-reliance of model coefficient values on the applied training data restricts its applications for cases outside the database for model calibration;
- Mechanistic-empirical model takes advantages of mechanical model and empirical model with basic accuracy, rationality and simplicity. Pavement conditions and responses are involved in a mechanistic way;
- Machine learning model takes advantages of artificial intelligence and has sophisticated model structures and operations. Relations between pavement performance and influencing factors can be efficiently and automatically captured from data iteratively. However, it has potential problems of overfitting and limited applications with very different data as empirical model.

3. Developing Neural Network Models for Asphalt Pavement Performance in Idaho

3.1 Introduction

As described in Chapter 2, current popular predictive models include ME models and ML models, each having unique advantages and limitations. With the improving analyses of pavement deterioration mechanisms and computer-aided techniques in data mining, the construction and calibration of predictive models have gained continuous development. ME models take advantage of solid foundations of mechanical models and concise expressions of empirical models (Deng, Zhang et al. 2022). Mechanistic aspects are reflected on the involvement of material failure criteria and pavement responses. Meanwhile, the historical performance of the pavement is utilized in the selection of model forms and the calibration of model parameters as traditional empirical models. ME models have wide applications in the pavement evaluation and design, especially in the industry for their explicit model forms which feature ease of use and implementation. A representative example is the software AASHTOWare Pavement ME Design built upon the pavement ME design guide (ARA-ERES 2004). Table 2.3 shows the rutting model in the pavement ME design guide with brief descriptions of model forms and parameters. The main limitations of ME models (as shown in Table 2.3) are the difficulty to obtain the pavement responses from a comprehensive pavement system and the need for calibration of model parameters. The latter one is common in the application of ME models in local areas. Parameters provided in the guide were calibrated from records of nation-wide pavements. The model accuracy is supposed to weaken in individual states, and the typical solution is local calibration (Bayomy, Muftah et al. 2018).

The major advantage of ML models is their high accuracy. Thanks to the advances in ML algorithms and model structures, the pattern and trend of data can be accurately and efficiently captured and predicted (Murphy 2012), which cannot be achieved by traditional models such as regression model. For predictive models of pavement performance, current research focused on ML models enhances the model accuracy by optimizing the model structure (Gong, Sun et al. 2021) and comparing different algorithms (Mazari and Rodriguez 2016). However, the implicit or overly complicated model expressions limit the application of constructed models in those studies. For example, NNs with multiple hidden layers and decision trees can hardly be applied directly in practice. For better usability of the predictive models by practicing engineers, it is highly desirable to develop NN models with explicit forms, and these often feature a single hidden layer (Stathakis 2009). Another main limitation of ML models lies in the need to have a relatively large dataset, especially when there are a large number of parameters that could be used as input factors for model construction (Lee, Dernoncourt et al. 2017).

While NNs have been employed to establish pavement performance models, past studies have mainly focused on achieving high model accuracy and generally failed to examine other model properties such as reproducibility and robustness (Mazari and Rodriguez 2016, Gong, Sun et al. 2021). As a result, the relations between the number of hidden neurons and model reproducibility or between the number of hidden neurons and model robustness are poorly understood. This has presented a question whether the constructed models are robust enough and led to a potential issue that they provide unreliable predictions with new datasets. We hypothesize that the constructed models can be systematically and comprehensively evaluated via a series of statistical methods to prove their applicability.

Considering the ME models and associated tools are still applied by ITD for asphalt pavement performance prediction (Bayomy, Muftah et al. 2018), work in this section aims to take the advantages of both ME and ML models. Predictive models of pavement with concise formula were constructed and calibrated using ML algorithm. Moreover, this works aims to strike the right balance between the model performance and time expenses. Analyses of models with different structures, accuracy and computation time were conducted. Section 3.2 describes the preliminary work including the data collection and variable selection for the model construction; Section 3.3 introduces the applied model and algorithm and procedures of the model construction; Section 3.4 presents the performance of constructed models including the accuracy, reproducibility and robustness; Section 3.5 summarizes the conclusions from this study.

3.2 Data Processing

Most data for the model construction in this study was retrieved from a project on the local calibration of ME models for asphalt pavement in Idaho (Bayomy, Muftah et al. 2018). This project provided material properties, structural parameters and traffic condition of pavement, which were partially applied as model inputs in this study. Additionally, we extracted the environmental data recorded in the Long-term Pavement Performance (LTPP) database for a comprehensive description of the pavement system. Environmental data from the weather station closest to the pavement section was used as the indicators of the pavement's environmental condition. This study utilized a total of 117 data points of rut depth measured from 27 road segments in 6 districts of Idaho in 2010s. All candidate model inputs are listed and described in Table 3.1. Before model construction, they were further selected via statistical analyses to ensure model conciseness.

Table 3.1 Candidate Model Inputs

Category	Inputs
Material Properties	• Binder content (%) • Vbe (%) – volume of effective binder • Gb (1) – specific gravity of asphalt • Gse (1) – effective specific gravity of asphalt mixture • Gsb (1) – bulk specific gravity of aggregate

	• Gmb (1) – bulk specific gravity of asphalt mixture
Structural Parameters	• Thickness of asphalt layer (mm) • Thickness of base layer (mm)
Traffic Condition	• AADTT – average annual daily truck traffic
Environmental Condition	• Precipitation (mm) – water equivalent of total surface precipitation over year time period • Evaporation (mm) – surface evaporation over year time period • Average temperature (°C) – average of the daily air temperatures based on average hourly temperatures • Mean temperature (°C) – average of the daily mean air temperatures based on the daily maximum and minimum hourly air temperatures • Freeze index (°C) – summation of difference between 0 °C and mean daily air temperature, when mean daily air temperature is less than 0 °C • Freeze-thaw days – number of days in the year when the maximum air temperature is greater than 0 °C and minimum air temperature is less than 0 °C on the same day • Average wind velocity (m/s) – time averaged magnitude of hourly wind velocity for the year • Average relative humidity (%) – average daily average relative humidity for the year • Average cloud cover (%) – average hourly fraction of cloud cover • Shortwave surface average (W/s^2) – average surface incident shortwave radiation for time period

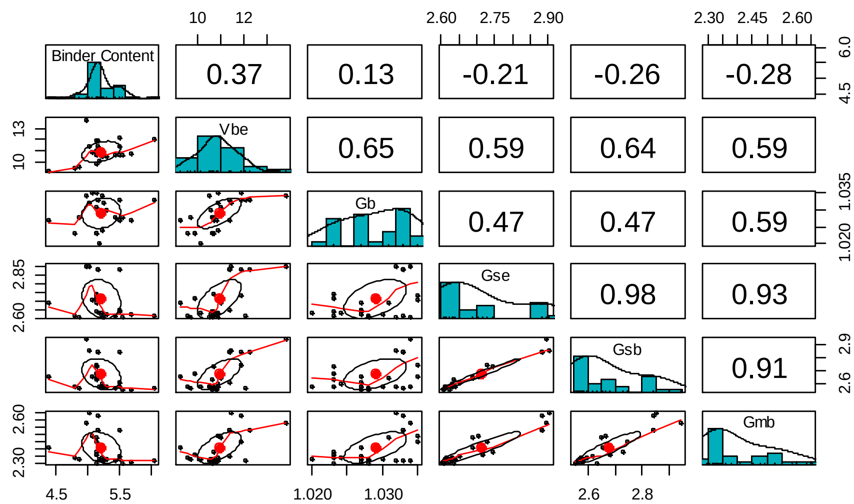
3.2.1 Correlation Analysis of Model Inputs

The first analysis is to assess the correlation between input factors so as to potentially reduce the redundant input factors (Saidi, Bouaguel et al. 2019). Figure 3.1 shows the scatter plot of matrices (SPLOM) of inputs in “Material Properties” and “Environmental Condition”. The histogram is on the diagonal showing the value and frequency of each input. The bivariate scatter plot is below the diagonal with a locally weighted scatterplot smoothing (LOWESS) line (Cleveland and Devlin 1988) and an ellipse showing the correlation between two inputs qualitatively. The flatter the ellipse is, the more related the two inputs are. The Pearson correlation coefficient is above the diagonal showing the correlation between two inputs quantitatively. It can be calculated from the following equation (Saidi, Bouaguel et al. 2019),

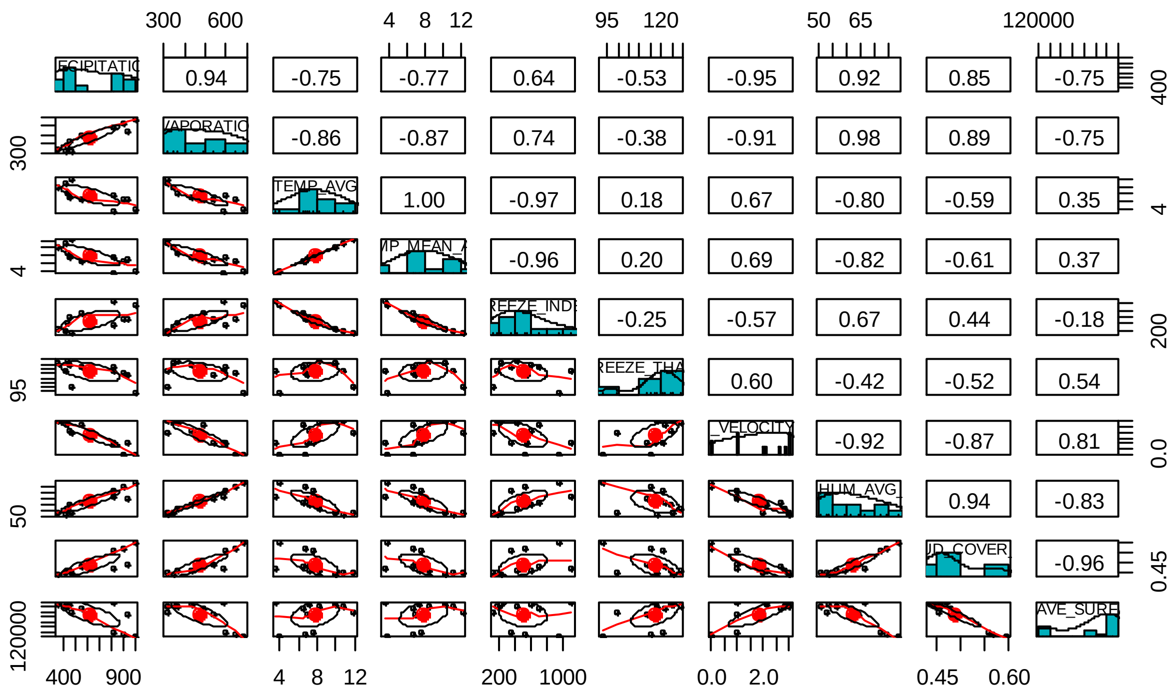
$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Where:

r = Pearson correlation coefficient
x_i, y_i = *i-th* values of two inputs
$\bar{x}$, $\bar{y}$= mean values of two inputs



(a)



(b)

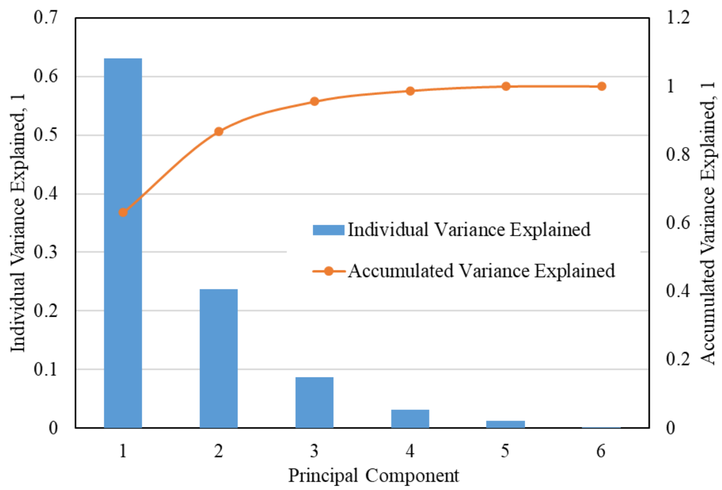
Figure 3.1 Scatter plot matrices of candidate model inputs including (a) inputs of Material Properties and (b) inputs of Environmental Condition

The inputs of material properties came from 28 asphalt mixtures in the project and the inputs of the environmental condition came from 12 weather stations in the LTPP database. To consider time effects on the environment, the average value during 2000-2020 was used for the inputs of the environmental condition in this step. From Figure 3.1, some obvious correlations can be observed, e.g., “Gse” and “Gsb”, and “Average temperature” and “Mean temperature”. The use of SPLOM simplified the correlation analysis of inputs since the exploration of their physical meanings was eliminated.

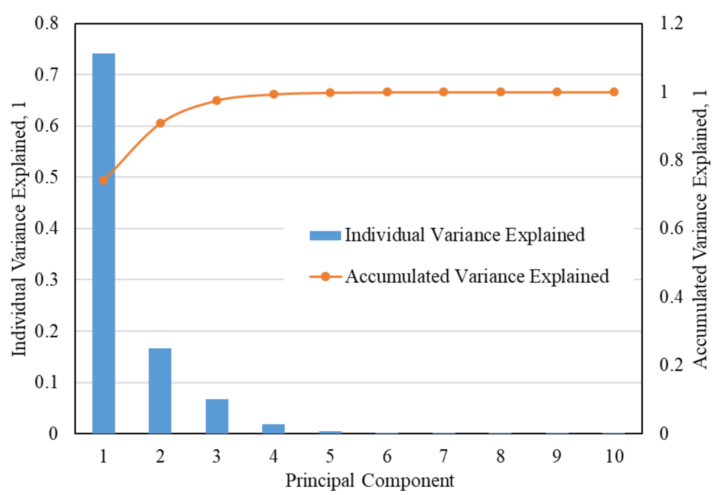
3.2.2 Principal Component Analysis of Model Inputs

The second analysis is the principal component analysis (PCA) of inputs. It is a common strategy to reduce the dimension of exploratory variables for making predictive models. The meaningful properties of original variables are retained and represented by the principal components, which are mutually orthogonal vectors constructed as linear combinations of original variables (Wold, Esbensen et al. 1987). To effectively compress the information expressed by original variables, the variance of original data is maximized in the directions of principal components. The contribution of each principal component to the total variance and its correlation with original variables can be used as indicators for the variable selection. Similar to the correlation analysis, PCA does not directly deal with physical meanings and mathematical forms of inputs, which are however reflected on the data.

Figure 3.2 shows the obtained principal components in both categories and their individual and accumulative explained variances. It can be seen from the plots that the number of required principal components to explain most of the total variance is fewer than the number of original variables. Absolute values of the loading value that indicates correlations between principal components and original variables are presented in Figure 3.3. Only the first six principal components are provided for a clearer presentation. These loading values were directly used in the variable selection described in the next section.



(a)



(b)

Figure 3.2 The properties of principal components including (a) inputs of Material Properties and (b) inputs of Environmental Condition

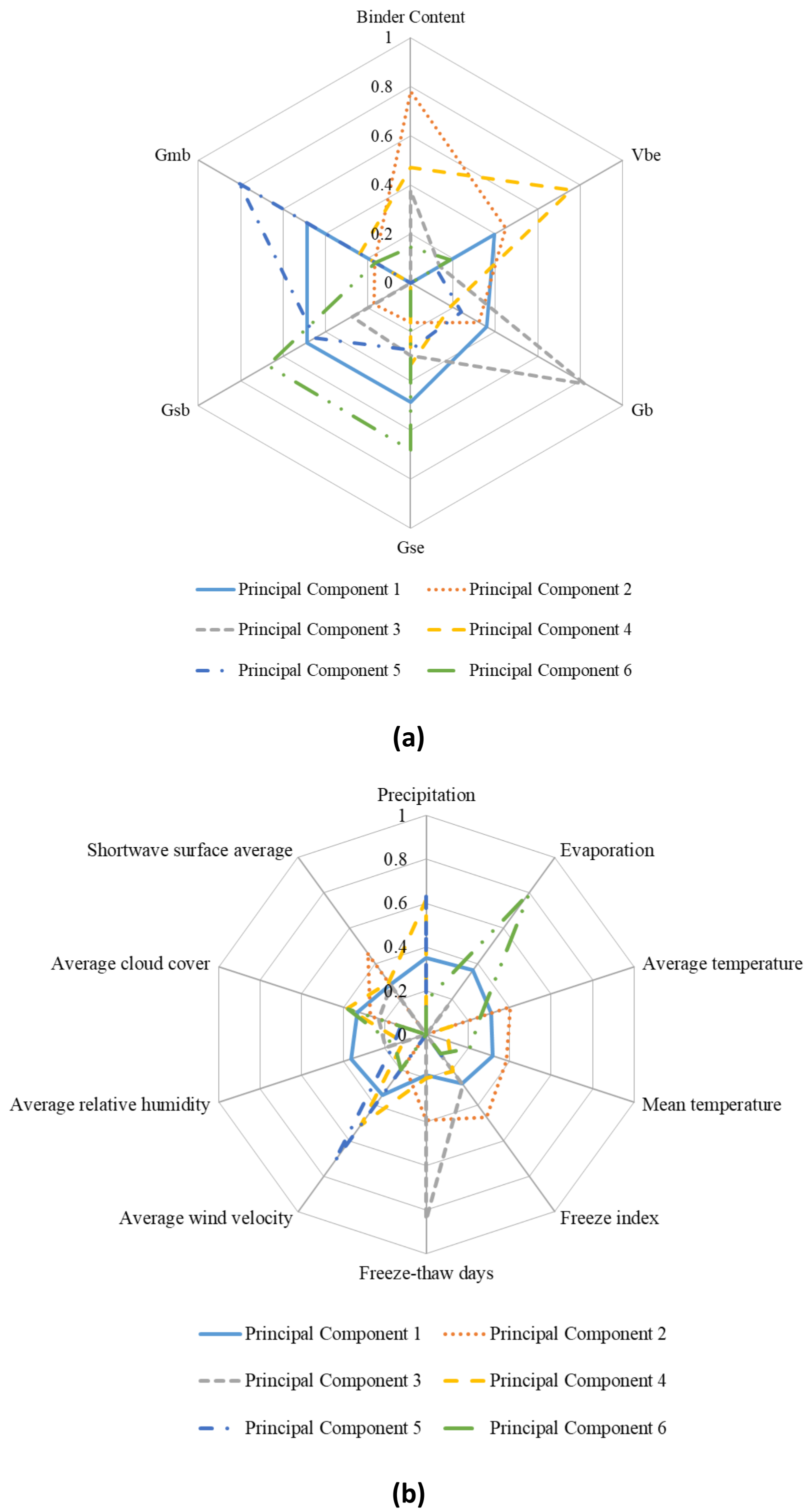


Figure 3.3 Correlations between principal components and original variables including (a) inputs of Material Properties and (b) inputs of Environmental Condition

3.2.3 Selection and Processing of Model Inputs

In the previous two sections, correlations of one input with others and the whole domain are generally clarified. Accordingly, this study developed a three-step variable selection method based on the correlation analysis and PCA, which is described in Figure 3.4. The first step is to select principal components with high explained variances in an inverse order. The second step is to select original variables with high correlations with those selected principal components in an inverse order. The last step is to eliminate some selected variables with high correlations with other selected variables in a reverse order. Specific values of the variance and correlation are mentioned in Figure 3.4. This novel method takes full advantage of the correlation analysis and PCA to reduce the model input dimensionality to as few variables as possible. Finally, “Binder content”, “Gb” and “Gsb” were selected to represent the material properties and “Evaporation”, “Freeze index”, “Freeze-thaw days” and “Shortwave surface average” were selected to represent the environmental condition. In the model construction described in the next section, all selected variables were transferred to the “accumulated” ones according to the date of rut depth measurement.

```
Input:  n: total number of principal components
        m: total number of original variables
        p: total number of selected variable for Round 1
        {PC(1), PC(2) ,…, PC(n)}: vector of principal components
        {MS(1), MS(2) ,…, MS(n)}: vector of maximum score of principal components
        {OV(1), OV(2) ,…, OV(m)}: vector of original variables
1. Generate vectors for selected variables Selected_Variable and principal components
   Selected_PC
2. Select eligible principle components
   for i = 1 to n do
      if accumulated explained variance of PC(i) < 0.95
         Selected_PC(end+1) = PC(i)
3. Select eligible variables for Round 1
         for j = 1 to m do
            if score of OV(j) >= 0.95* MS(i)
               Selected_Variable(end+1) = OV(j)
            end if
         end do
      end if
   end do
4. Eliminate variables for Round 2
   for k = p to 1 do
      for l = k-1 to 1 do
         if Pearson correlation coefficient of Selected_Variable(k) and
            Selected_Variable(l) > 0.90
             Selected_Variable(k) = [ ]
         end if
      end do
   end do
Output: Selected_Variable
```

Figure 3.4 Pseudo-code of variable selection

3.3 Proposed Predictive Models

Models proposed in this study aim to take advantage of the applicability of ME models and the accuracy of ML models. Accordingly, neural networks with one hidden layer calibrated with particle swarm optimization algorithm were utilized. The usability is reflected on their explicit expressions (in terms of connection weight and bias values), and the accuracy is reflected on the artificial intelligence-based model calibration.

3.3.1 Neural Networks

Neural network (NN), inspired by the biological nervous system, is widely applied in image recognition, language processing and predictive analysis, etc. to gain superior performance over conventional methods (Wang 2003, Shi, Schillings et al. 2004). The information reception, conversion and transmission are simulated in the artificial NN as in the real brain. Different types of NNs are suitable for addressing different problems and can be distinguished by the connection between neurons, complexity of the network and propagation mode of the information. The NN used in this study is a feed-forward NN, which was the first and simplest devised NN (Schmidhuber 2015). As its name implies, the information propagates in one direction from input layer, hidden layer to output layer as shown in Figure 3.5. In pavement engineering, it can serve as predictive models for material property characterization (Saha, Gu et al. 2018), structure condition evaluation (Gong, Sun et al. 2021), etc.

Figure 3.5 shows a typical three-layered feed-forward NN structure with one hidden layer and one output neuron. Neuron numbers in the input and output layers are determined by the numbers of inputs and outputs of the problem. The numbers of hidden layers and contained neurons can be adjusted based on the problem complexity and the NN performance. As shown in Figure 3.5, the information first propagates from the input layer to the hidden layer which is expressed by the following equations,

$$y'_j = f\left(net_j\right)$$

$$net_j = \sum_{i=1}^{N} w_{ij} x_i + b_j$$

Where:

y'_j = the value predicted by the *j*-th hidden neuron
f = transfer function, e.g., hyperbolic tangent function
w_{ij} = the weight of the *i*-th input neuron to the *j*-th hidden neuron
x_i = the value of the *i*-th input neuron
b_j = the bias added to the *j*-th hidden neuron
N = the number of input neurons

Subsequently, the information propagates to the output layer and is expressed by the output neuron as follows,

$$y_k = \sum_{j=1}^{M} w'_{jk} y'_k + b'_k$$

Where:

y_k = the value predicted by the k-th output neuron
w'_{jk} = the weight of the j-th hidden neuron to the k-th output neuron
b'_k = the bias added to the k-th output neuron
M = the number of hidden neurons

For a NN with one hidden layer, the total number of weights and biases can be calculated as follows,

$$N = m \times n + 2 \times n + o$$

Where:

N = the number of parameters to be calibrated
m = the number of inputs
n = the number of hidden neurons
o = the number of outputs

It should be noted that values of the input and output neurons are typically normalized values of original input and output variables. Benefiting from close neuron connections and nonlinear transfer functions illustrated in Figure 3.5, complex or fuzzy relations between inputs and outputs can be effectively captured. The major part of NN construction is basically the calibration of those weights and biases.

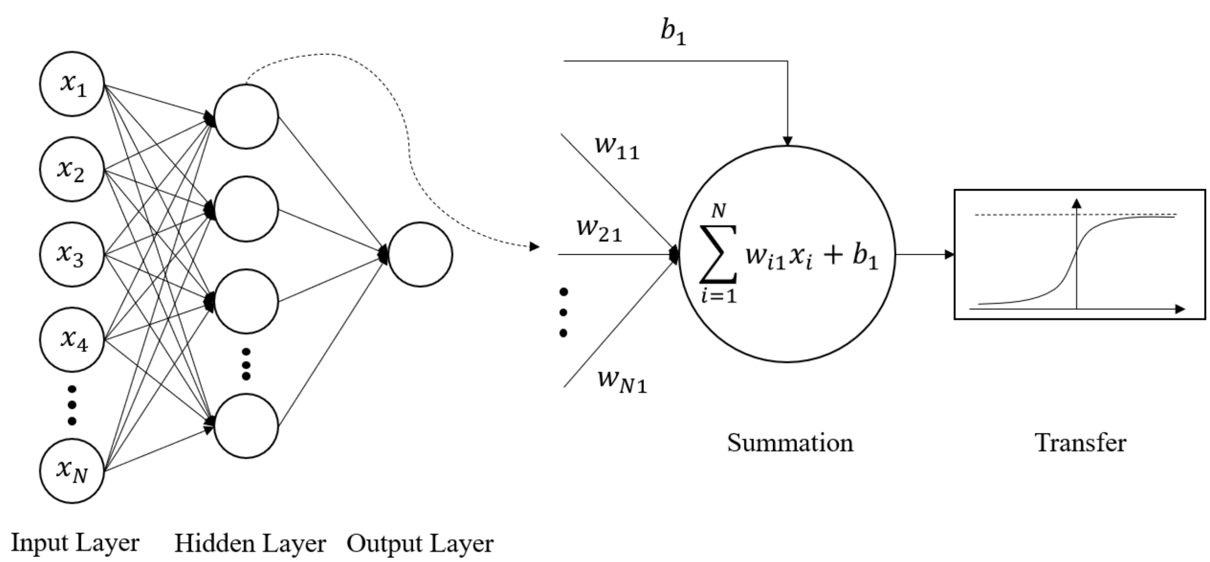


Figure 3.5 Architecture of a feed-forward neural network

Another reason underlying the popularity of NNs is that they are accessible in multiple programming platforms such as MATLAB and Python. For example, a feedforward NN with one hidden layer can be constructed using a built-in function “nftool” in MATLAB. Users only need to select the number of hidden neurons, proportions of data for model training, validation and testing, and training algorithm (e.g., Levenberg-Marquardt (Yu and Wilamowski 2018)). However, this study aimed to shed light on the relation between the numbers of hidden neurons and datasets and its effects on the accuracy, reproducibility and robustness of the constructed NN. Therefore, we applied an artificial intelligence (AI) algorithm to construct NNs to allow better comparisons.

3.3.2 Particle Swarm Optimization

Particle swarm optimization (PSO) algorithm, inspired by the social behavior of a swarm of organisms (Kennedy and Eberhart 1995), belongs to another branch – evolutionary algorithms of AI approaches. It commonly serves as a solution finder in engineering problems such as structure optimization (Cao, Qian et al. 2017), model equivalency (Deng, Luo et al. 2021) and parameter backcalculation (Deng, Luo et al. 2020, Deng, Luo et al. 2021). As shown in Figure 3.6, each potential solution holds a specific location as an individual organism carries information. The location is iteratively updated according to the current and historical locations of all solutions as the individuals exchange information with each other. The mathematical expressions of solution updating are presented in the following equations (Shabbir and Omenzetter 2015):

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1}$$

$$\mathbf{v}_i^{k+1} = w\mathbf{v}_i^k + c_1 r_{1,i}^k(\mathbf{p}_i^k - \mathbf{x}_i^k) + c_2 r_{2,i}^k(\mathbf{g}_i^k - \mathbf{x}_i^k)$$

Where:

$\boldsymbol{x}^i_k$, $\boldsymbol{v}^i_k$ = the position vector and velocity vector of the *i*-th solution in the *k*-th iteration
$\boldsymbol{p}^i_k$ = the best position of the i-th solution in the past k iterations
$\boldsymbol{g}^i_k$ = the best position of all solutions in the past k iterations
w = the inertial weight
c_1, c_2 = the cognition and social coefficients
$r^k_{1,i}$, $r^k_{2,i}$ = random numbers in the interval [0,1] for the *i*-th solution in the *k*-th iteration

In this study, the position vector of each solution contains the weights and biases of a NN. It is evaluated by the accuracy of the NN with the training group, which is introduced in the next section.

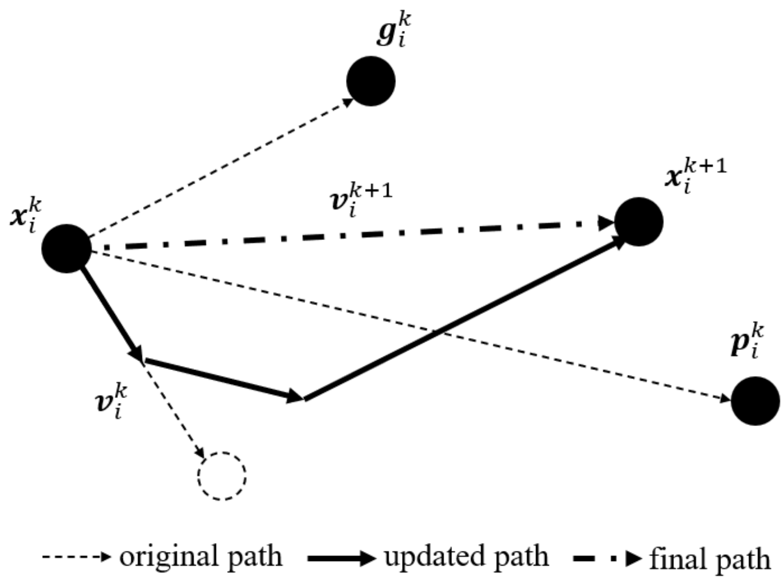


Figure 3.6 Mechanism of particle swarm optimization

3.4 Results and Discussion

This section presents the results and associated analyses of PSO-NN introduced in Section 3.3. Different from most applications of NNs which concern the model accuracy, the focuses of this study are the relations between the numbers of calibrated parameters and iterations and the updating efficiency, reproducibility and robustness of models. A total of 117 datasets were randomly separated into 80 percent for model training and 20 percent for model validation. Corresponding accuracies were named as training accuracy and validation accuracy. Considering the limited amount of datasets and the high complexity of applied model structures, the hyperparameters of PSO were not optimized in this study; instead, they followed previous research (Heris 2015, Deng, Zhang et al. 2022).

3.4.1 Updating Efficiency and Reproducibility

The updating performance of PSO can be validated from Figure 3.7(a), which shows the decreasing normalized mean square error (NMSE) of predicted outputs in the training group with increasing number of iterations. NMSE was utilized to evaluate the model accuracy in the model calibration. It has the calculation method as the following equation and the relation with the coefficient of determination as follows (Yu, Lai et al. 2007),

$$NMSE = \frac{\sum_i \left(y_i - \hat{y}_i\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2}$$

$$R^2 = 1 - NMSE$$

Where:

y_i = the i-th component of the actual value vector
$\hat{y}_i$ = the i-th component of the predicted value vector
$\bar{y}$= the mean of all actual values
R^2 = the coefficient of determination

Figure 3.7(a) illustrates the 95% confidence band from six repetitive runs of model calibration of four NNs. It can be seen that the updating patterns of PSO in NNs with different numbers of hidden neurons were similar. Besides, the model accuracy converged with the increase of hidden neurons even though their initial accuracies were different. Figure 3.7(b) illustrates the boxplot of the results after 10000 iterations, from which similar findings were observed in the variation of model accuracy. It should be noted that even with small numbers of hidden neurons (e.g., 3 & 5), PSO-ANN finally reached a desirable level of model accuracy with a certain level of computation time (e.g., 10000 iterations), which is much higher than those of nationally and locally calibrated ME models (Bayomy, Muftah et al. 2018).

A t-test was conducted to verify these findings statistically. It is a common strategy to determine the significance of differences between groups (Luo, Luo et al. 2013). Based on the assumption made above that the variance of model accuracy changed with the number of hidden neurons, the Welch's t-test (Welch 1947) was applied, which deals with groups with unequal variances. The main procedure is to calculate t-value as the following equation,

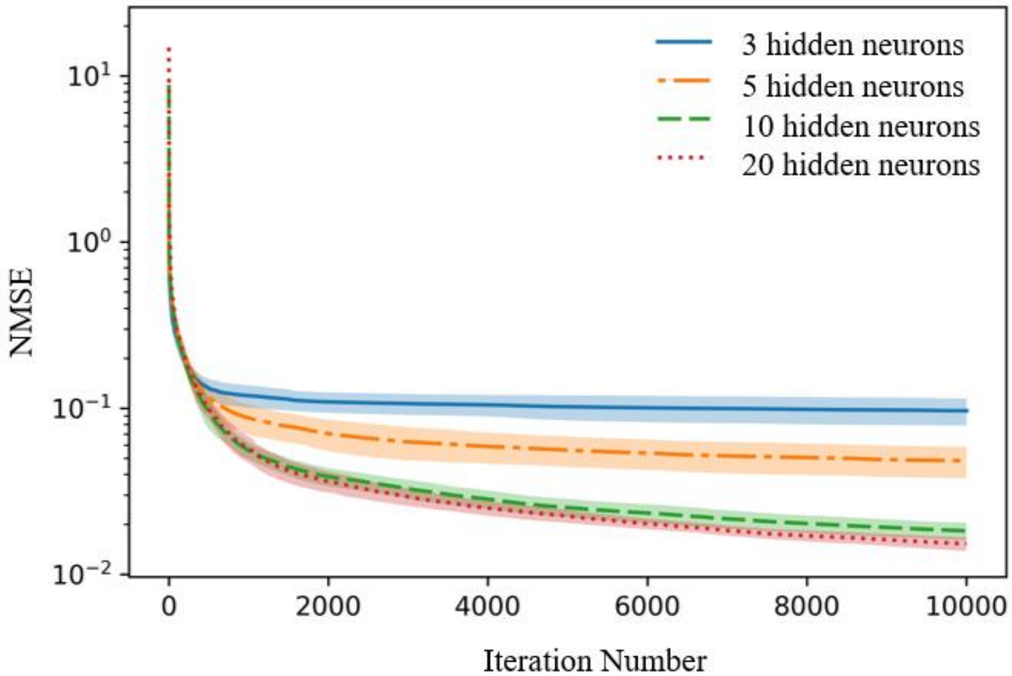
$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

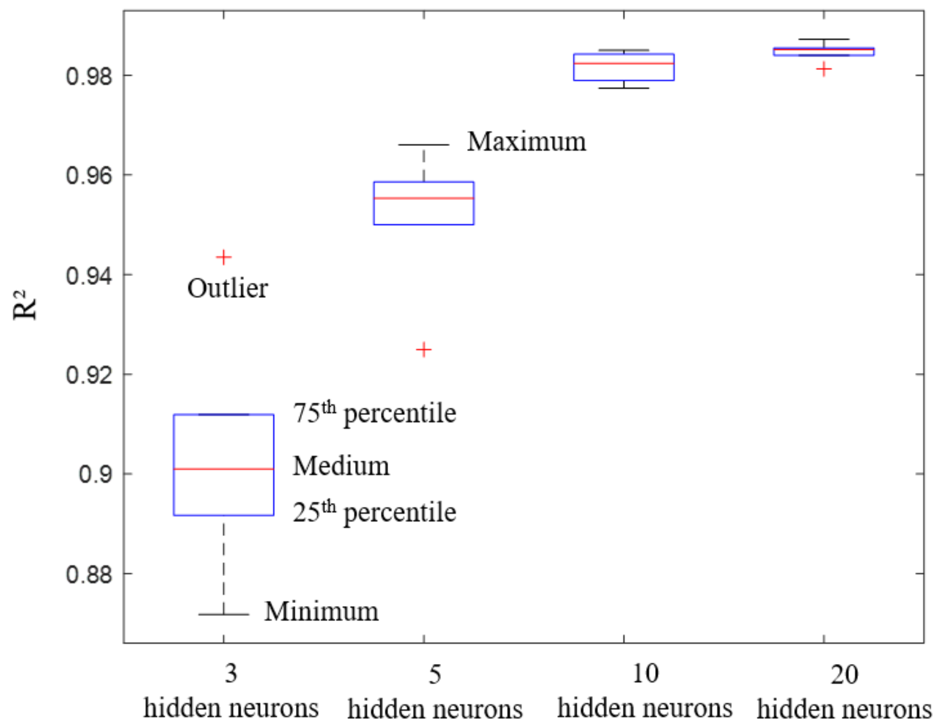
$\bar{X}_1$, $\bar{X}_2$ = means of two groups
s_1, s_2 = sample standard deviations of two groups
n_1, n_2 = sample sizes of two groups

The t-value is the ratio of differences between and within groups. It compares the condition of applied groups to the null hypothesis that there is no significant difference between their means. For each degree of freedom calculated as the following equation, t-value follows a specific normal distribution which indicates the probability of the null hypothesis being accepted or rejected.

$$DF = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$


3 hidden neurons
5 hidden neurons
10 hidden neurons
20 hidden neurons
NMSE
10^1
10^0
10^{-1}
10^{-2}
0
2000
4000
6000
8000
10000
Iteration Number

(a)



(b)

Figure 3.7 The Updating efficiency of PSO-NN including (a) NMSE with iterations and (b) Boxplot of model accuracy after 10,000 iterations

Table 3.2 reveals that the computation time and number of hidden neurons have interaction effects on the model accuracy. Table 3.2 shows the differences between the model accuracy (R^2) of NNs with different numbers of hidden neurons and iterations. NNs with small numbers of hidden neurons (e.g., 3 and 5) had similar accuracy only when the computation time (iteration number) was small (e.g., 1000). However, NNs with large numbers of hidden neurons (e.g., 10 and 20) had similar accuracy at any level of computation time. Although increasing the hidden neurons can mitigate the effect of limited computation time on the model accuracy, it leads to an increase of calibrated parameters and a potential unstable system, which is reflected on the decrease in the model reproducibility.

Table 3.2 Results of t-test

Iteration Number	Compared Groups (Hidden Neuron Numbers)	Degree of Freedom	t-value of Compared Groups	t-value at 95% Confidence Level	Result
1000	3 & 5	10.00	-2.21	[-2.23, 2.23]	Accept
1000	5 & 10	6.28	-2.98	[-2.42, 2.42]	Reject
1000	10 & 20	8.37	0.21	[-2.29, 2.29]	Accept

5000	3 & 5	8.63	-3.90	[-2.28, 2.28]	Reject
5000	5 & 10	5.77	-4.57	[-2.47, 2.47]	Reject
5000	10 & 20	8.42	-1.31	[-2.29, 2.29]	Accept
10000	3 & 5	8.02	-4.19	[-2.30, 2.30]	Reject
10000	5 & 10	5.45	-5.08	[-2.51, 2.51]	Reject
10000	10 & 20	8.64	-2.05	[-2.28, 2.28]	Accept

Model reproducibility can be examined via multiple runs of the model construction. Similar predicted outputs are expected from a stable model. The cosine similarity was utilized to compare two vectors as shown in the following equation,

$$\text{Similarity}(i, j) = \frac{\mathbf{A}_i \cdot \mathbf{A}_j}{\|\mathbf{A}_i\| \|\mathbf{A}_j\|}$$

Where:

Similarity(i,j) = the similarity between the *i*-th and *j*-th predicted output vectors $\boldsymbol{A}_i$ and $\boldsymbol{A}_j$, which individually contain 94 elements

For each NN,10 repetitive runs were conducted and a total of 45 similarity values were calculated accordingly. In this study, three different hidden neurons (5, 10 & 20) and two different model accuracies (0.90 & 0.95) were selected to investigate their relations with the model reproducibility. To eliminate the effect of initial values of calibrated parameters, each 10 repetitive runs shared the initial values of weights and biases of the NN.

As shown in Figure 3.8, cosine similarities of the selected six cases are similar and close to 1, which indicates good reproducibility of these calibrated NNs. However, two general trends can still be observed from the boxplot. First, with a fixed number of hidden neurons, the mean and variation of cosine similarity increases and decreases with the increase of training accuracy of the model, respectively. It results from the fact that more potential solutions exist at poorer training accuracy of the model. Second, with a fixed model accuracy, the mean of cosine similarity increases and then decreases with the increase of hidden neurons while the variation behaves oppositely. It can be explained by the weakness of increasing hidden neurons mentioned above that, an unstable system can be created with excessive calibrated parameters. More importantly, it reminds us an optimum number of hidden neurons may exist to balance the tradeoff between model accuracy and reproducibility with limited computation time.

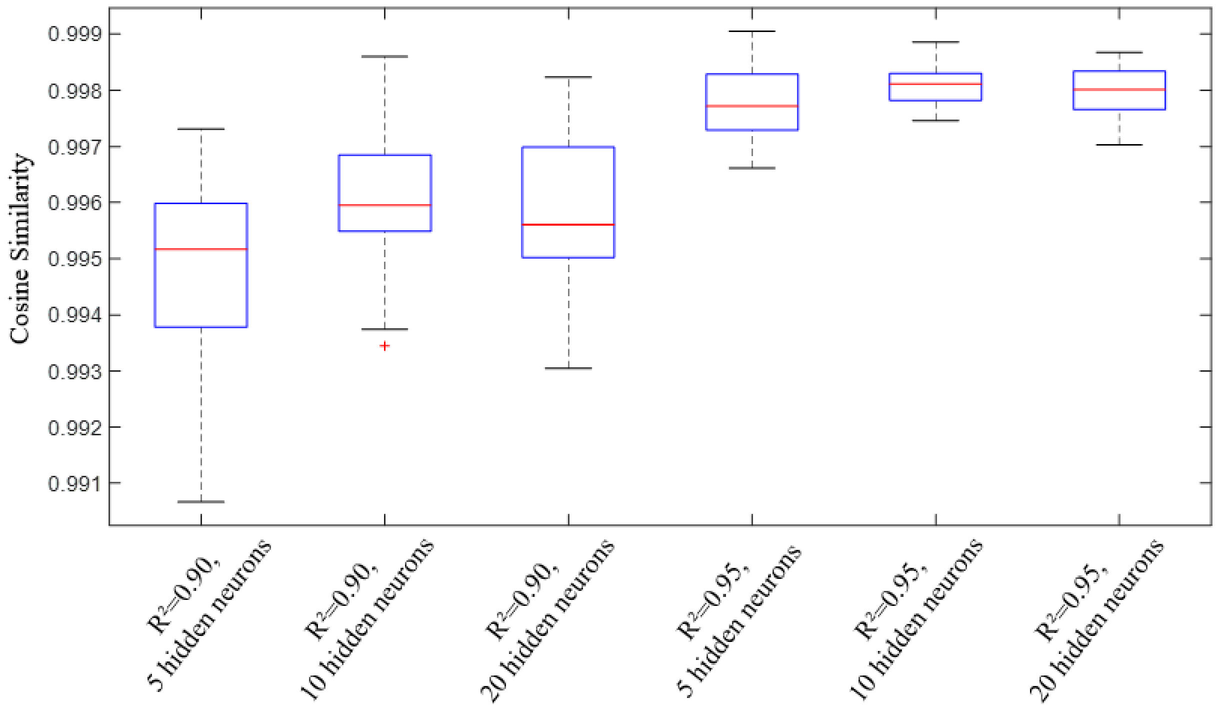


Figure 3.8 The boxplot of model reproducibility

3.4.2 Model Robustness

Robustness is an important quality widely applied in the post-analysis of statistical models. It requires constructed models not to be unduly sensitive to outliers. Model robustness can have different definitions in corresponding models. For example, Su et al. (Su, Zhang et al. 2018) treated the attack success rate, distortion, attack-agnostic robustness score and transferability as the indicators for the robustness of ImageNet models. Ringwood et al. (Ringwood, Mérigaud et al. 2019) treated the sensitivity of wave energy control systems to modelling errors as the robustness. Nisbet et al. (Nisbet, Elder et al. 2009) summarized several checking techniques for the post-analysis of ML models, in which two were selected for the model robustness in this study. The first one is the accuracy check of trained models using the validation group, i.e., examination of the performance of trained models to predict new data. The second one is the sensitivity check of model inputs, i.e., examination of the effects of random errors in model inputs on the model predictions.

The accuracy check in this study included evaluations of the variability by R^2 and the residual distribution by the Shapiro-Wilk test. The validation accuracy of models at different levels of training accuracy and with different numbers of hidden neurons is presented in a boxplot Figure 3.9. It can be seen from the plot that model accuracy with new datasets is not significantly affected by the two controlled conditions – training accuracy and hidden neuron number. Besides, compared with the training accuracy (in the range of 0.90 to 0.95), the level of validation accuracy is lower but tolerable (mostly in the range of 0.80 to 0.85), and this suggests that the overall accuracy of the PSO-NN models can be sufficiently high given a high training accuracy. Therefore, the focus can be put on the updating efficiency and reproducibility as described in Section 3.4.1 in the design of PSO-NN.

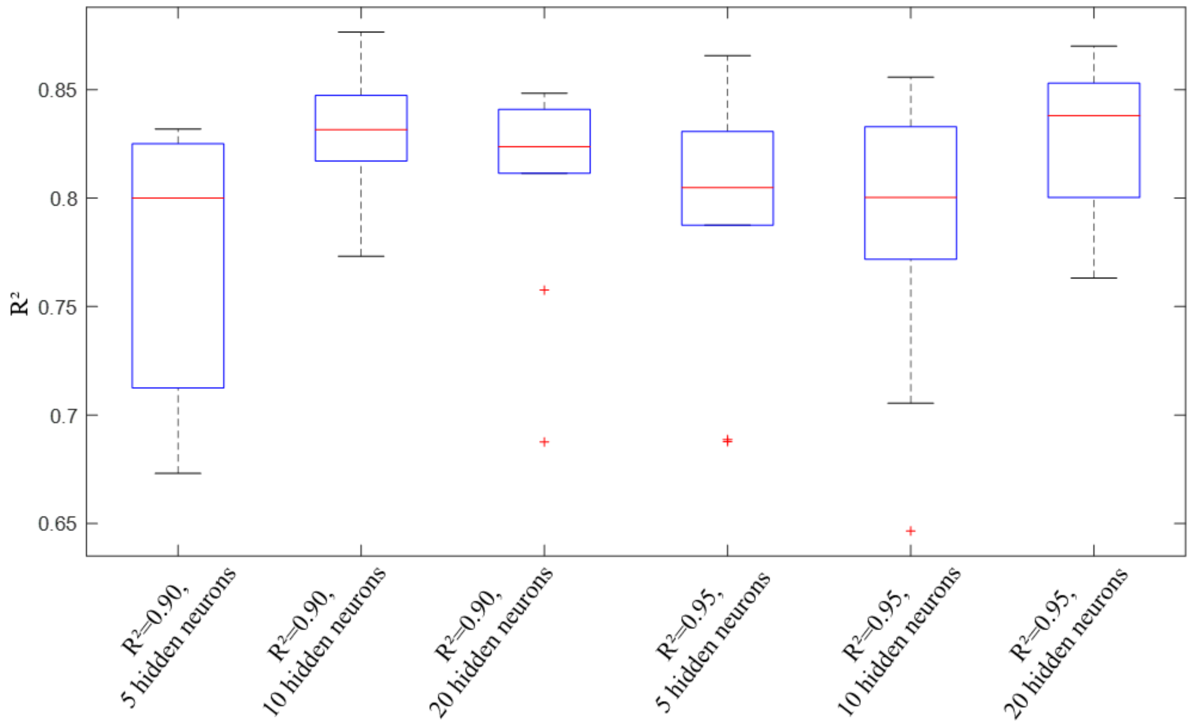


Figure 3.9 The boxplot of validation accuracy vs. training accuracy and number of hidden neurons

Shapiro-Wilk test is a typical test of normality in statistics (Shapiro and Wilk 1965). It tests the null hypothesis that samples follow a normal distribution using the statistic as the following equation,

$$W = \frac{\left(\sum_{i=1}^{N} a_i x_{(i)}\right)^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Where:

$x_{(i)}$ = the i-th ordered sample values
a_i = the i-th constant generated from the mean, variance and covariance of N order statistics from a normal distribution
x_i = the i-th sample value
$\bar{x}$ = the sample mean
N = the sample size

Similar to Welch's t-test described in Section 4.1, the null hypothesis can be accepted if the statistic of samples falls in the accepted range at a certain confidence level. Table 3.3 presents the Shapiro-Wilk test results of residuals from trained NNs with the validation group. While Table 3.3 does not list all the

W values of all 60 individual cases, the analysis revealed that the difference between predicted and measured rut depths in most cases follow a normal distribution, which indicates the model output was well explained by the selected model inputs. Combined with the accuracy presented in Figure 3.9, the residual distribution can be described by a standard normal distribution.

Table 3.3 Results of Shapiro-Wilk Test

Accuracy (R^2)	Number of Hidden Neurons	Proportion of Accepting the Hypothesis in 10 Repetitive Tests
0.90	5	80%
0.90	10	100%
0.90	20	80%
0.95	5	100%
0.95	10	80%
0.95	20	90%
Test Information	Sample Size: 23	Accepted range of the statistic W at a 95% confidence level: [0.9142, 1.0000]

The sensitivity check is to quantify the relative change in the response of model output corresponding to a small change in the model input (Chen, Ren et al. 2020). The typical method is the Morris method in which only one input variable is adjusted in each run (Morris 1991). In addition to showing the relative importance of model inputs in the constructed model (Shojaeefard, Akbari et al. 2013), the sensitivity of model inputs can be used to further reduce the original input dimensionality (Ye, Shi et al. 2009, Chen, Ren et al. 2020). Ye et al. (Ye, Shi et al. 2009) presented the step-by-step procedures of the Morris method for NNs as described in Figure 3.10, in which the model inputs and outputs were normalized to be equally compared.

Input: n: total number of data for the model training
m: total number of model inputs
x={ $x_1^1,\ x_2^1,\ ...,\ x_m^1;\ x_1^2,\ x_2^2,\ ...,\ x_m^2;\ ...;\ x_1^n,\ x_2^n,\ ...,\ x_m^n$ }: vector of model inputs
y: vector of model output

1. Normalize each model input and output; and calculate the mean ($\bar{X}$) and standard deviation (σ) of each normalized input
2. Train the NN of interest using normalized input and output vectors **X** and **Y**
3. Calculate the sensitivity of each model input

   for i = 1 to m **do**

   create input vector $\begin{array}{l}\{\bar{X}_1,\ \bar{X}_2,\ ...,\ \bar{X}_i-\sigma_i,\ ...,\ \bar{X}_m;\\ \bar{X}_1,\ \bar{X}_2,\ ...,\ \bar{X}_i-0.99\sigma_i,\ ...,\ \bar{X}_m;\ ...;\\ \bar{X}_1,\ \bar{X}_2,\ ...,\ \bar{X}_i+\sigma_i,\ ...,\ \bar{X}_m\}\end{array}$

   obtain the predicted output vector $\hat{Y}_i$ from the trained NN with the created input vector

   for j = 2 to 201 **do**

   $$S_i(j-1)=\frac{\hat{Y}_i(j)-\hat{Y}_i(j-1)}{0.01\sigma_i}$$

   end do

   calculate the averaged sensitivity

   $$\bar{S}_i=\left|\frac{1}{200}\sum_{j=1}^{200}S_i(j)\right|$$

   end do
4. Normalize the averaged sensitivity for all model inputs

   for k = 1 to m **do**

   $$\hat{S}_i=\frac{\bar{S}_i}{\sum_{i=1}^{m}\bar{S}_i}$$

   end do

Output: final sensitivity vector { $\hat{S}_1,\ \hat{S}_2,\ ...,\ \hat{S}_m$ }

Figure 3.10 Pseudo-code of sensitivity analysis (revised from (Ye, Shi et al. 2009))

The sensitivity of model inputs from the PSO-NN with 10 hidden neurons and the accuracy (R^2) 0.95 is presented in Figure 3.11. It can be seen that the variation of sensitivity in 10 repetitive runs is greater than the variations seen in the model accuracy (Figure 3.7(a)) or reproducibility (Figure 3.8), likely due to the fact that the inputs and outputs were normalized before the sensitivity analysis. Besides, the sensitivity of input variables is comparable with each other. No variables can be eliminated from the model for excessively low sensitivity. Considering that the sensitivity is significantly affected by the collected data and the various inputs were collected from different sources in this study, the calculated sensitivity data was not analyzed quantitatively.

Other model checks using a variable-by-variable method can be conducted on the development of predicted outputs. Trend and value change of the developed curve can be evaluated in the perspective of the dealt engineering problem. For example, Figure 3.12 shows the development of rut depth with 95% confidence bands corresponding to two major factors - AADTT and Freeze-thaw days respectively.

The PSO-NN with 10 hidden neurons and the accuracy (R^2) 0.95 was applied as in the sensitivity analysis. In each curve, the examined variable varies and the rest remain at their average values. It can be seen that the predicted rut depth increases with AADTT which is generally acknowledged. The predicted rut depth decreases with the number of freeze-thaw days per year, it can be explained by that the increasing temperature cycles fasten the aging of asphalt pavement materials, which leads to the stiffening of the asphalt layer and the decrease of the rutting accumulation (Stephens 1990, Deng, Luo et al. 2020).

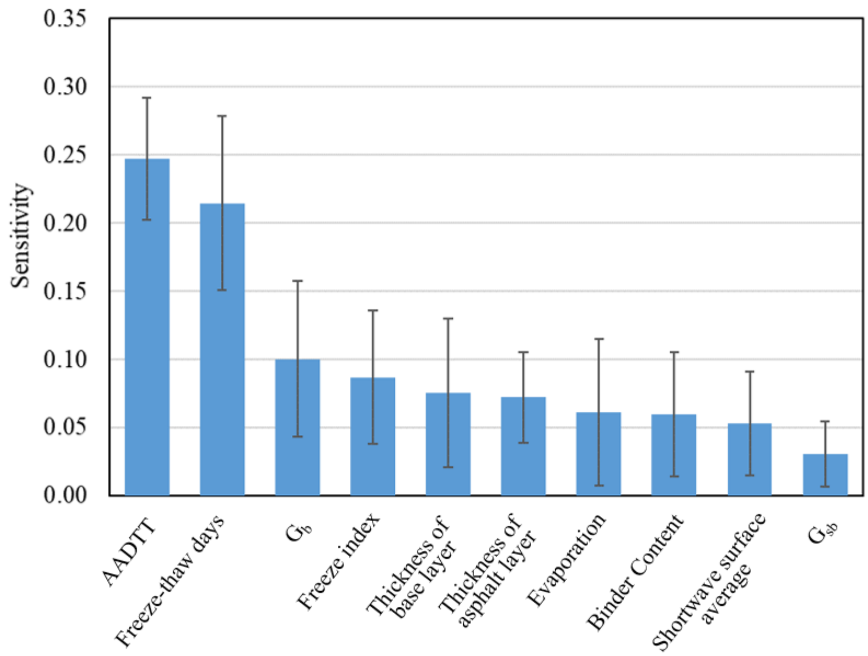


Figure 3.11 The sensitivity of model inputs

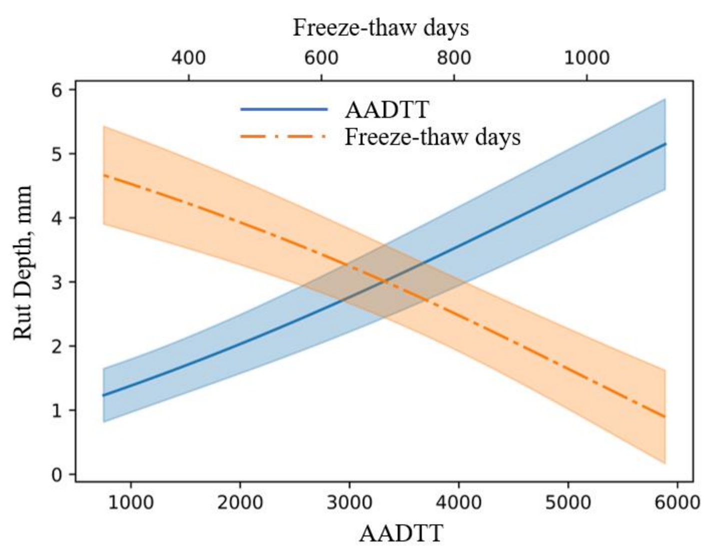


Figure 3.12 The predicted rutting development with AADTT and Freeze-thaw Days

3.5 Conclusions

In Chapter 3, a modeling approach was developed to predict the rutting of asphalt pavement rutting in Idaho using PMED data, from its pre-processing to post-analysis. The case study examined a total of 117 datapoints of rut depth measured from 27 road segments in 6 districts of Idaho in 2010s. Neuron networks with one hidden layer were calibrated using particle swarm optimization algorithm. Performance of models with different calibrated parameters, accuracy and calibration time was evaluated and compared. The major findings in this study can be summarized as follows.

- A three-step variable selection method with PCA and correlation analysis can effectively reduce the input dimensionality (from 19 to 10 inputs in this study), and this helps to mitigate the need for a relatively large dataset for ML models;
- Proper increase of hidden neurons (from 3 to 10) improves the model accuracy (R^2 from 0.90 to 0.98) while continuous increase of hidden neurons (from 10 to 20) has insignificant effects on the model accuracy (R^2 from 0.981 to 0.984);
- With a fixed number of hidden neurons (in the range of 5 to 20), the model reproducibility increases as the model accuracy (R^2) increases from 0.90 to 0.95. With a fixed model accuracy (R^2 in the range of 0.90 to 0.95), the model reproducibility shows a parabolic trend as the number of hidden neurons increases from 5 to 20. This finding strongly suggests that during the design of NN models one should consider an optimum number of hidden neurons to balance the model accuracy and reproducibility;
- When the training accuracy is sufficiently high (e.g., R^2 above 0.90), it as well as the number of hidden neurons have limited effects on the validation accuracy and normality of residuals;
- With proper variable selection, the sensitivity of model inputs is relatively comparable.

4. Developing Gene Expression Programming Models for Asphalt Pavement Performance in Idaho

4.1 Introduction

Chapter 3 presents the application of AI in constructing predictive models for asphalt pavement performance, specifically in the model structure and model calibration. The relationship between pavement distresses and influencing factors were described with weighted neurons and transfer function in NN. Meanwhile, it was pointed out that the accuracy can be improved with the increase of model complexity (i.e., number of hidden layers and neurons) and at the sacrifice of model stability. In this section, the advantage of AI algorithms in model calibration was kept and extended to the operation selection and combination. Explicit model form with ensured accuracy will benefit its practice in pavement performance prediction. The remainder of this chapter are organized as follows: Section 4.2 introduces the predictive model, including its background and mechanism; Section 4.3 describes the model construction from preparation to assessment; Section 4.4 presents the model performance for the four applied distresses, including comparisons with ME and regression models; Section 4.5 summarizes the conclusions from this study.

4.2 Proposed Predictive Models

The predictive pavement distress models were developed by the use of gene expression programming (GEP) in this study. GEP was proposed by Ferreira on the basis of the genetic algorithm (GA) and genetic programming (GP) (Ferreira 2001). In summary, linear strings of fixed length (characteristic of GA) are encoded to express nonlinear entities of different sizes and shapes (characteristic of GP) in GEP (Ferreira 2001). Variables and constants as well as arithmetic and logic operators can be included in the entity as elements of the gene. As a descendance of GA, GEP updates entities via mutation, crossover and selection, similar to what genes experience in the process of natural selection.

Unlike most applications of GA in optimization problems that numerical values were dealt with (Deng, Zhang et al. 2022), GEP was applied in this study for its capability of updating the predictive model on its function form. Explicit form of the predictive model can be guaranteed. Meanwhile, accuracy of the predictive model can be improved with the use of ML algorithm. K-expressions are applied in GEP to encode expression trees (e.g., Figure 4.1). They both represent the algebraic expression "$\sqrt{a \times (b + c)}$". For the convenience of genetic operations and valid programs, the length of the gene is determined by the length of the head which is set by users and the number of arguments of the function with the most arguments (e.g., 1 for " √ " and 2 for "+") (Deng, Zhang et al. 2022). It indicates the genes may contain elements which are not expressed but enables the genes to express trees of different sizes and shapes.

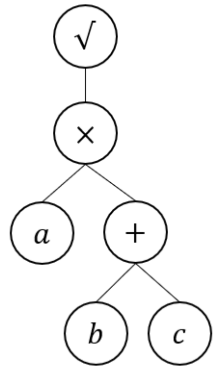


Figure 4.1 An example of expression tree

The procedure of solution (gene) updating in GEP is presented in Figure 4.2, in which the sizes of genes and chromosomes, fitness function, rates of genetic operations etc. affect the updating efficiency and performance. Details can be found in the follow-up research conducted by Ferreira (Ferreira 2006) and some of them will be discussed in the following sections. GEP has been applied in the areas of pavement materials and structures for property characterization (Aslam, Farooq et al. 2020) and performance prediction (Yao, Leng et al. 2021). However, currently there is the lack of a systematic application of GEP in constructing predictive models for individual distresses of field asphalt pavements as ME models, which is the motivation of conducting this section.

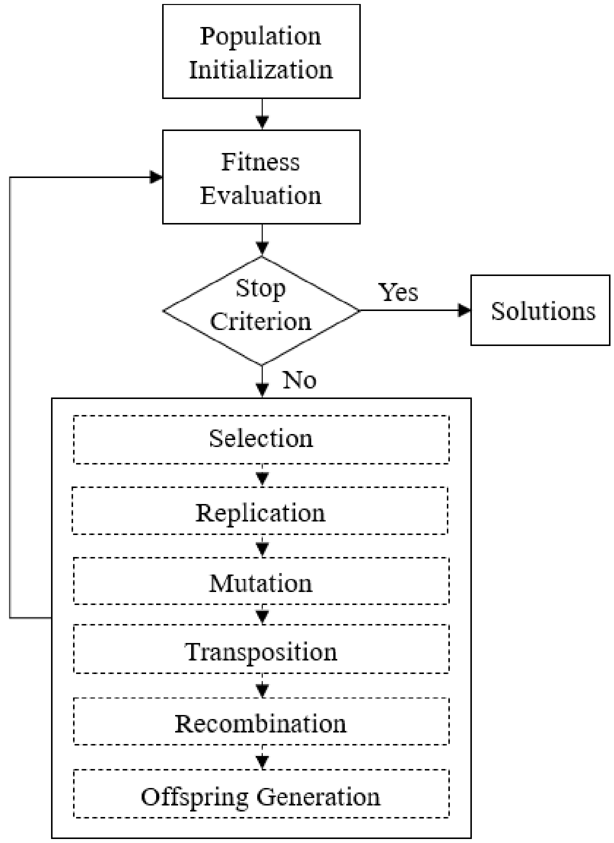


Figure 4.2 Flow chart of GEP

4.3 Model Construction

This section describes the complete process of model construction from the selection of model hyperparameters to the model performance evaluation and final form determination. As in Chapter 3, the rutting development of asphalt pavement segments in Idaho was utilized as an example and the input variables were selected by the method proposed in Section 3.2.

4.3.1 Hyperparameter Selection

For stable model construction and result analysis, the software GeneXproTools 5.0 (GEPSOFT 2014) was applied in this study to conduct the GEP. Three hyperparameters of GEP - the head size and numbers of chromosomes and genes were first selected based on a sensitivity analysis. These parameters determine the model complexity and calibration efficiency, which are essential for the application of the GEP in practice.

Since the effects of hyperparameters on the model construction were characterized, only the training dataset was utilized in this section. Basic settings of GEP are provided in **Error! Reference source not f ound.** and others such as the rates of mutation and transposition were set at values recommended in the software. In each case, one of the three hyperparameters was adjusted and others were set at their minimum values. As a general indicator for the model accuracy, the coefficient of determination (R^2) was used for the rest of this section.

Table 4.1 GEP Parameter Setting

GEP Parameter	Setting
Head Size	10, 15, 20
Chromosome Number	100, 300, 500
Gene Number	4, 5, 6
Function Set	+, −, ×, /, sqrt, exp, Inv, X2
Fitness Function	mean squared error (MSE)
Linking Function	addition

Figure 4.3 illustrates the evolution of model accuracy with generation number in three cases. Although the values of hyperparameters are different, a common plateau occurs when the generation number approaches to 100,000. Therefore, the 100,000-th generation was taken as the stop criterion in the model construction to balance the computation time and model accuracy. The model accuracy after 100,000 generations of six repetitive runs is presented in Figure 4.4. According to the mean and variance of R^2 values shown in the figure, the head size and numbers of chromosomes and genes were chosen to be 20, 500 and 4 in this section.

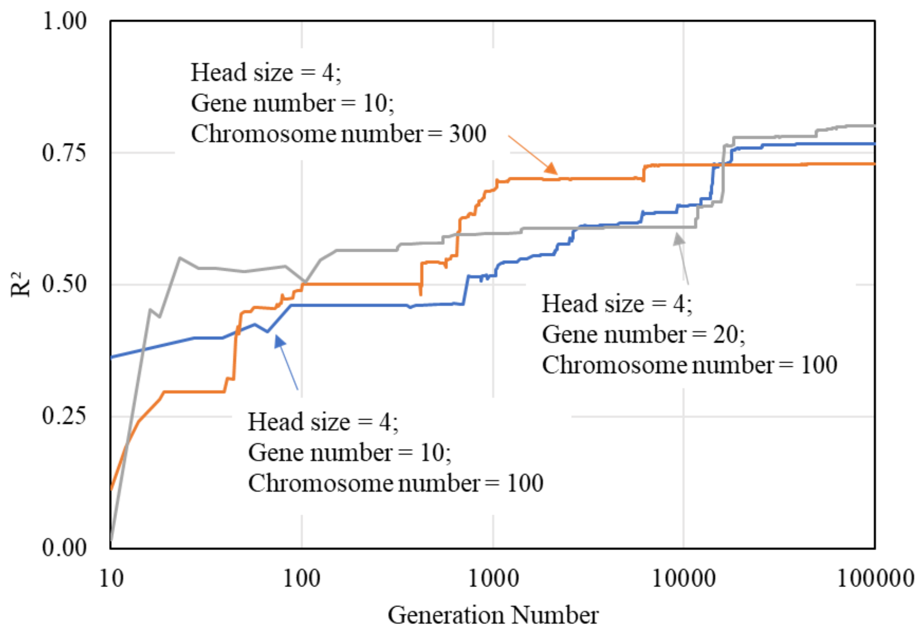
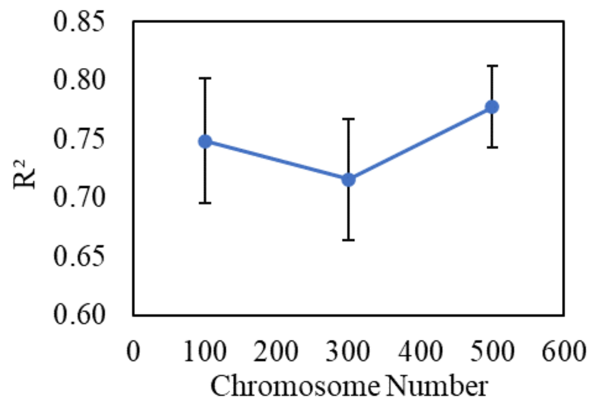
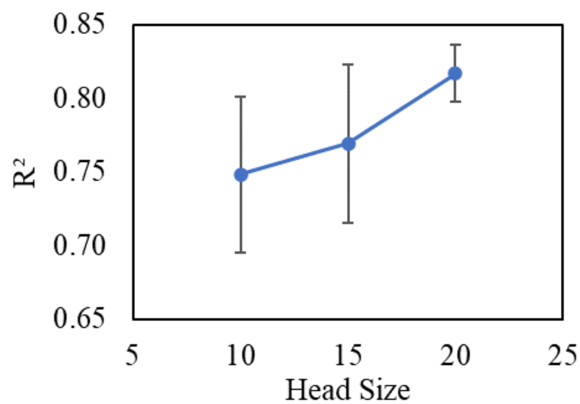


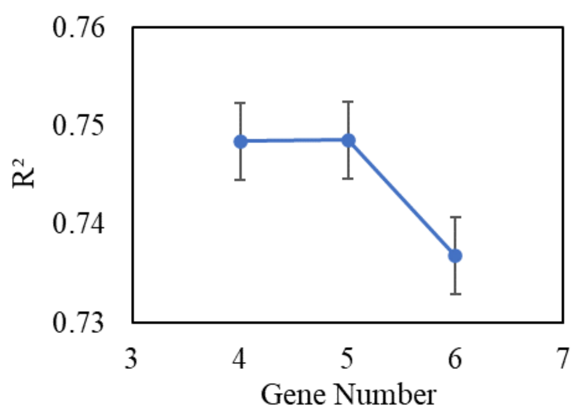
Figure 4.3 Model accuracy as a function of generation number



(a)



(b)



(c)

Figure 4.4 Model accuracy of cases with different (a) head sizes; (b) chromosome numbers; and (c) gene numbers

4.3.2 Model Evaluation

In addition to the model accuracy from the training dataset, several other properties should be paid equal attention for the model evaluation and final model determination. Especially in ML models, the obtained model forms and coefficients are likely to change during the repetitive model construction (Deng and Shi 2022). The first property is the model accuracy for the validation and test datasets. These two datasets were utilized in the total 42 models (seven cases X six repetitive runs) constructed in the previous section. As shown in the histograms on the diagonal of Figure 4.5, the training accuracy using R^2 as the indicator of all cases achieves the level 0.60-0.85. However, the corresponding validation and test accuracies dispersedly distribute across the range of 0 to 1. Furthermore, there is no significant correlation between model accuracies of the three datasets, which can be proven by the Pearson correlation coefficient (*r*), shown above the diagonal of Figure 4.5 (Saidi, Bouaguel et al. 2019). These two findings indicate that the training accuracy cannot guarantee the same performance on new datasets; and validation and test datasets are indispensable for a comprehensive evaluation of model accuracy. Accordingly, an overall accuracy considering the three datasets was applied in this study to select the final model form in repetitive runs.

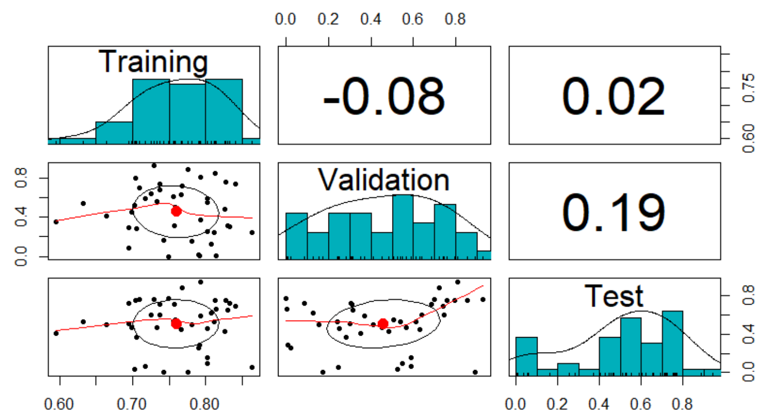


Figure 4.5 Scatter plot matrix of model accuracy

The second property is the normality of residuals. Model residual or error, is the difference between the actual and predicted values of model outputs. In a good model, the outputs are well described by the applied predictors (input variables) and constant terms. The remainder (i.e., residual) is supposed to be random and centered on zero (Cox and Snell 1968). In this study, the model residual (ε) was first controlled by the fitness function as shown in Table 4.1 and the following equation,

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)$$

Where:

N = the total number of data points

y_i, $\hat{y}_i$ = the actual and predicted values of the i-th data point

In the model evaluation, the residual distribution serves as a supplement to the overall accuracy. Two methods were utilized to examine the normality of residuals. The first one is the Shapiro-Wilk test as used in Section 3.4.2. The second one is the quantile-quantile (Q-Q) plot, which compares two probability distributions by plotting their quantiles (Gnanadesikan and Wilk 1968). Quantiles of two similar distributions are supposed to lie on the equality line. In this section, quantiles of the model residual distribution were plotted against ones of a standard normal distribution for the models failing the Shapiro-Wilk test.

The third property is the model sensitivity. Different from the one used in Section 3.4.2, the relative sensitivity of input variables was obtained using the following equation proposed in a previous study (Gandomi, Yun et al. 2013) for GEP models,

$$S_i = \frac{N_i}{\sum_{i=1}^{n} N_j} \times 100\%$$

$$N_i = f_{\max}\left(x_i\right) - f_{\min}\left(x_i\right)$$

Where:

$f_{max}(x_i)$, $f_{min}(x_i)$ = the maximum and minimum values of the predicted output over the *i*-th input domain while other inputs are taken their mean values

As mentioned previously, the relative sensitivity shows the contribution of model inputs (pavement properties) to outputs (pavement distresses) and is instructive to the pavement design and management. Meanwhile, the continuous changing pattern of the model output with individual inputs is a potential tool to an optimum pavement design (Deng, Shi et al. 2021).

4.4 Results and Discussion

This section presents the models constructed using the hyperparameters determined from Section 4.3.1 and the evaluation of their performance based on the properties introduced in Section 4.3.2. Results of all four distresses of asphalt pavement in Idaho are included, in which the roughness is represented by the international roughness index (IRI).

4.4.1 Model Accuracy

Starting from the model accuracy, Table 4.2 lists the three accuracies of all 24 models (4 distresses X 6 repetitive runs) constructed in this section using R^2 as the indicator. Considering the data of longitudinal cracking, thermal cracking and IRI is limited, the whole data in these three categories was divided into “Training” and “Validation & Test” datasets, of which the latter one contains the maximum and minimum values of input variables as the test dataset of rutting. The proportions of data in these two datasets are around 80% and 20% respectively. Similar to the finding from Figure 4.5, Table 4.2 indicates the limited relevancy of validation and test accuracies to the training accuracy. Therefore, as recommended in Section 4.3.2, the overall accuracy served as the indicator to select the final model form from repetitive model construction. The selected cases and their accuracies are marked in bold in Table 4.2. Equations of final models are presented in Appendix B.

Table 4.2 Accuracy of constructed models

Distress & Data Amount	Model No.	Training Accuracy	Validation & Test Accuracy	Overall Accuracy
Rutting	**1**	**0.82**	**0.61**	**0.68**
Rutting	2	0.76	0.20	0.10
Rutting	3	0.85	6.63E-04	5.94E-04
Rutting	4	0.83	0.21	0.58
Rutting	5	0.82	0.01	0.01
Rutting	6	0.81	0.17	0.43
Longitudinal Cracking	**1**	**0.94**	**0.69**	**0.89**
Longitudinal Cracking	2	0.96	0.03	0.70
Longitudinal Cracking	3	0.88	0.26	0.40
Longitudinal Cracking	4	0.91	0.64	0.78
Longitudinal Cracking	5	0.93	0.46	0.81
Longitudinal Cracking	6	0.93	0.20	0.57
Thermal Cracking	1	0.93	0.83	0.56
Thermal Cracking	2	0.95	0.25	0.60
Thermal Cracking	3	0.93	0.25	0.15
Thermal Cracking	**4**	**0.98**	**0.48**	**0.83**
Thermal Cracking	5	0.97	0.30	0.12
Thermal Cracking	6	0.95	0.30	0.69
IRI	1	0.84	0.58	0.79
IRI	2	0.92	0.50	0.85
IRI	3	0.82	0.55	0.70
IRI	4	0.89	0.80	0.84
IRI	**5**	**0.90**	**0.93**	**0.91**
IRI	6	0.86	0.66	0.81

For a clear idea of the accuracy achieved by the models constructed by GEP, Figure 4.6 shows the model accuracy of different models, including the linear regression (LR) model and local calibrated ME model (ME-Local). It can be seen from the plot that models constructed by GEP have higher accuracy, which results from the applied ML algorithm and the complexity of the model form. Compared with LR models shown as the following equation, models constructed by GEP have more optional functions and combinations to capture the relations between model inputs and outputs. However, they are not restricted by the specific models of material and structural responses and failures as ME models (Deng, Zhang et al. 2022).

$$y_i = \beta_0 + \sum_{j=1}^{N} \beta_j x_{ij} + \varepsilon_i$$

Where:

y_i = the i-th model output
x_{ij} = the i-th observation on the j-th model input
β_0, β_j = the model coefficients
ε_i = the i-th error term

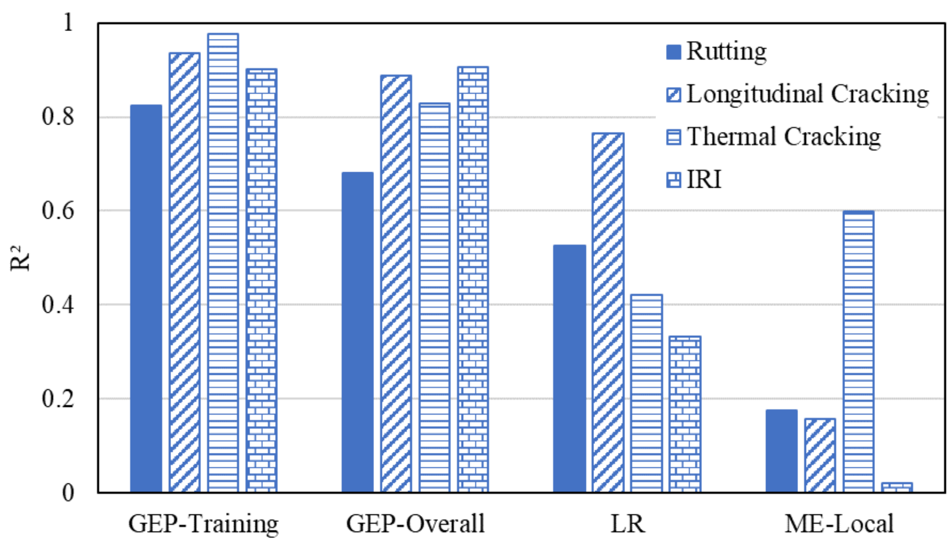


Figure 4.6 Comparison of model accuracy as a function of modeling approach and pavement distress type

4.4.2 Residual Normality

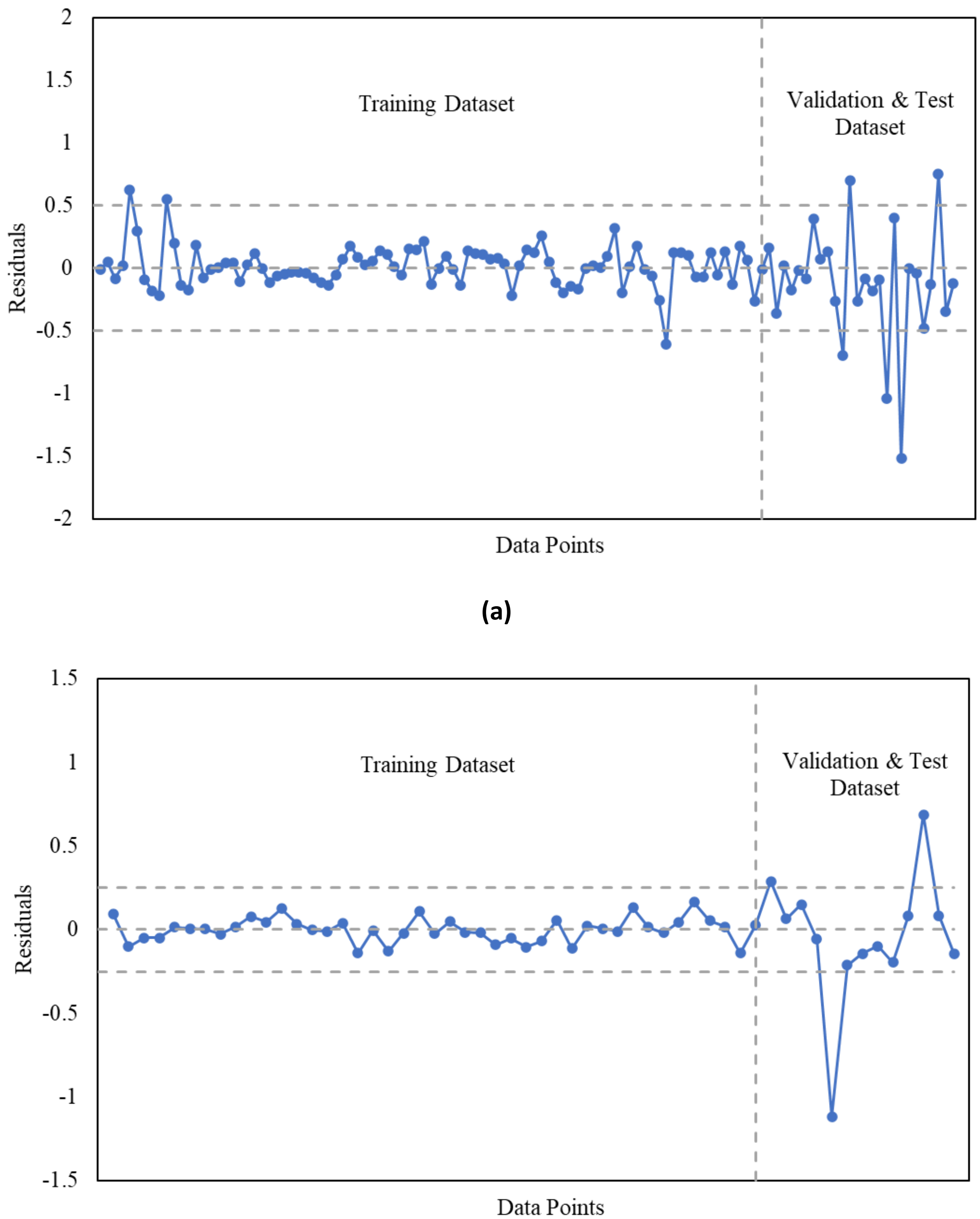
As described in Section 4.3.2, the distribution of model residuals was examined in addition to their values. Similar to the overall accuracy, both “Training” and “Validation and Test” datasets were considered in this section to check if residuals of constructed models with training and new data follow the normal distribution.

Table 4.3 shows the basic information of model residuals including the Shapiro-Wilk test results. Although the mean of residuals is near zero in all four models, which stems from the applied fitness function of GEP, the residuals of rutting and thermal cracking models do not follow a normal distribution as indicated by the Shapiro-Wilk test statistics and corresponding accepted ranges. By plotting the model residuals as Figure 4.7, it can be observed that outliers exist in these two models, especially in the "Validation and Test" dataset. Therefore, the revised group of model residuals which excludes the outliers was analyzed. As indicated in Table 4.3, the revised groups passed the Shapiro-Wilk test. Similar findings are illustrated in the Q-Q plot as Figure 4.8, in which quantiles of model residuals without outliers lie linearly with those of a standard normal distribution. In comparison, typical heavy-tailed Q-Q plot describes the original model residuals of rutting and thermal cracking models (Croarkin and Tobias 2012).

Considering the definition of outliers which indicates significant differences between observations and predictions (Grubbs 1969), the non-normal distribution of model residuals in this study was ultimately caused by the limited model accuracy on the new data. It reminds us that model accuracy on the validation and test datasets should be examined separately and focused as well as the training dataset in the model form selection.

Table 4.3 Information of Model Residuals

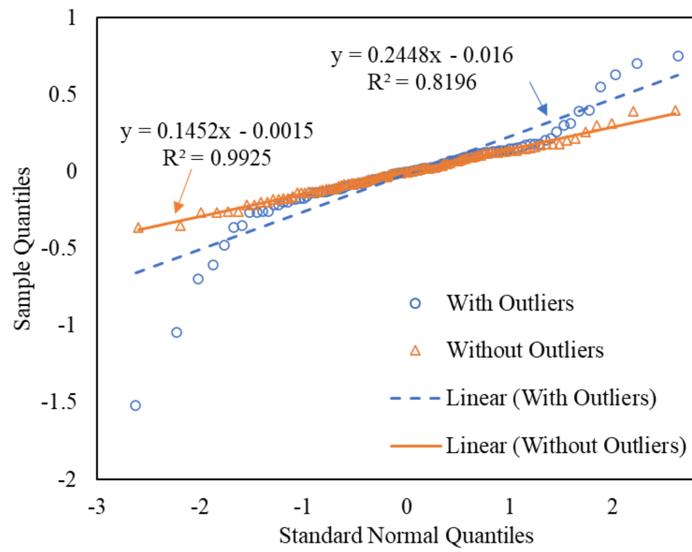
Distress	Rutting	Rutting	Thermal Cracking	Thermal Cracking	Longitudinal Cracking	IRI
Model	Original	Revised	Original	Revised	Original	Original
Mean	-0.0160	-0.0015	-0.0110	-0.0089	-0.02882	0.0073
Standard deviation	0.2701	0.1455	0.1997	0.0867	0.1636	0.1267
Test statistic	0.828	0.992	0.664	0.982	0.956	0.9770
Accepted range at 95% confidence level	[0.978, 1]	[0.976, 1]	[0.958, 1]	[0.956, 1]	[0.946, 1]	[0.970, 1]
Result	Reject	Accept	Reject	Accept	Accept	Accept



(a)

(b)

Figure 4.7 Residual plot of (a) rutting model; and (b) thermal cracking model



(a)

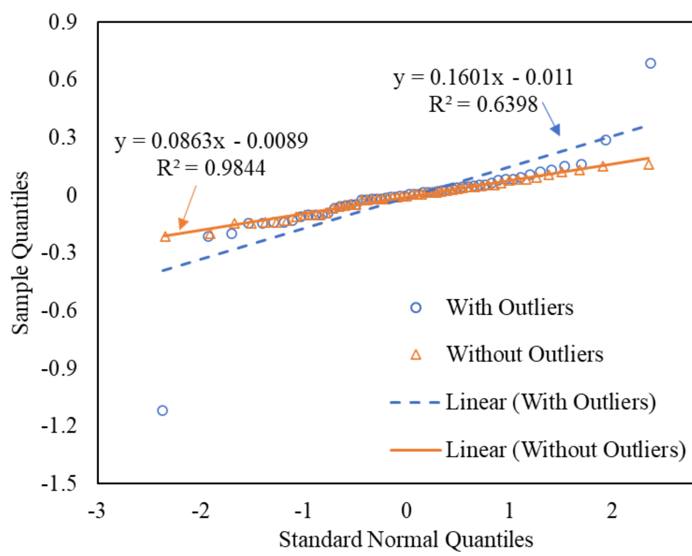


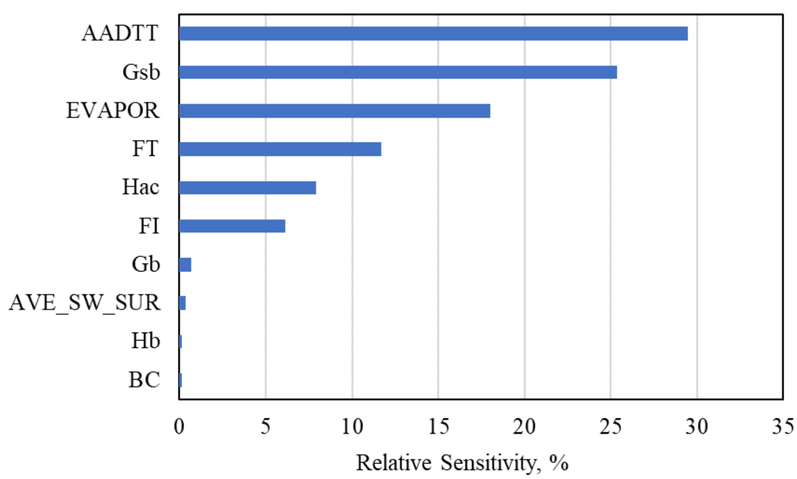
Figure 4.8 Q-Q plot of residuals of (a) rutting model; and (b) thermal cracking model

4.4.3 Model Sensitivity

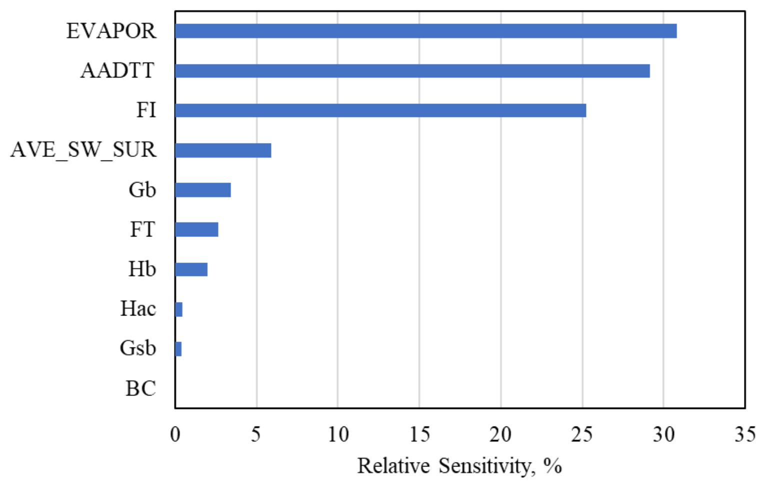
Regarding the superior accuracy of models constructed by GEP mentioned in Section 4.4.1, it was achieved at the expense of less simplicity than LR models and less rationality than ME models. The model quality was greatly affected by the provided data for the model construction. For predictive models of pavement distresses that aim to explore the coupled effects of factors from various sources, model evaluation in terms of rationality is especially necessary.

Figure 4.9 shows the relative sensitivity of input variables. Several general conclusions can be drawn from the figure. First of all, the value and rank of relative sensitivity vary with models. Variables of pavement condition have different contributions to different pavement distresses. Second, compared with neuron networks calibrated with artificial intelligence algorithms, the variation of relative sensitivity is much more obvious in models constructed by GEP (Deng and Shi 2022). The most extreme case is that BC was eliminated from the final form of longitudinal cracking model as shown in Figure 4.9(b). It indicates that the effect of the proposed variable selection method introduced in Section 3.2 on the final model form is limited. An independent variable selection was conducted by GEP according to the relations between model inputs and outputs, which were expressed by the data for the model construction. This characteristic was utilized in a previous study to calculate the variable appearance frequency, which showed the importance and contribution of each input variable to the model output (Yao, Leng et al. 2021).

Additionally, Figure 4.9 reveals that to a very significant extent the model rationality can be expressed by the relative sensitivity of input variables. For instance, it precisely captures the major contributions of traffic (AADTT) to longitudinal cracking and environment (FT and AVE_SW_SUR) to thermal cracking as shown in Figure 4.9(b) and Figure 4.9(c). Referring to mechanics, these two distresses are mainly induced by repetitive traffic loads and temperature cycles (ARA-ERES 2004). However, the relative sensitivity underestimates the effects of some variables on specific distresses, such as the effect of binder content on the rutting development since the asphalt layer of the pavement suffers a lot from permanent deformation under traffic loads at intermediate and high temperatures (Deng, Shi et al. 2021). Overall, the relative sensitivity is worthwhile to be applied in future models as a key indicator for the model rationality and model form selection.

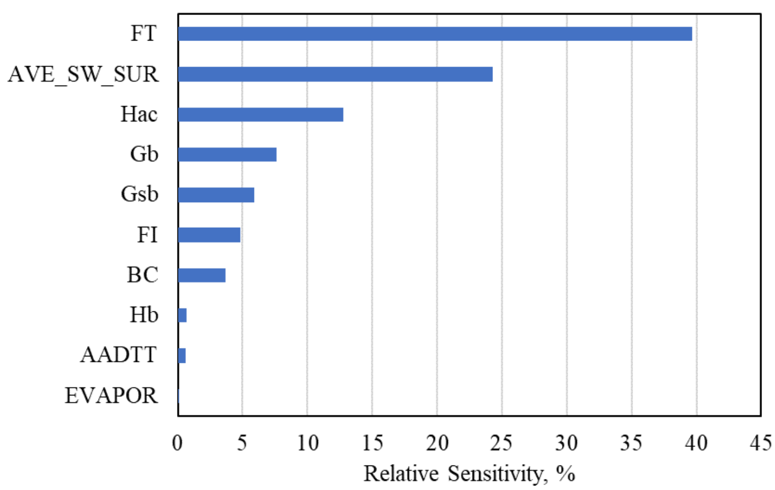


(a)



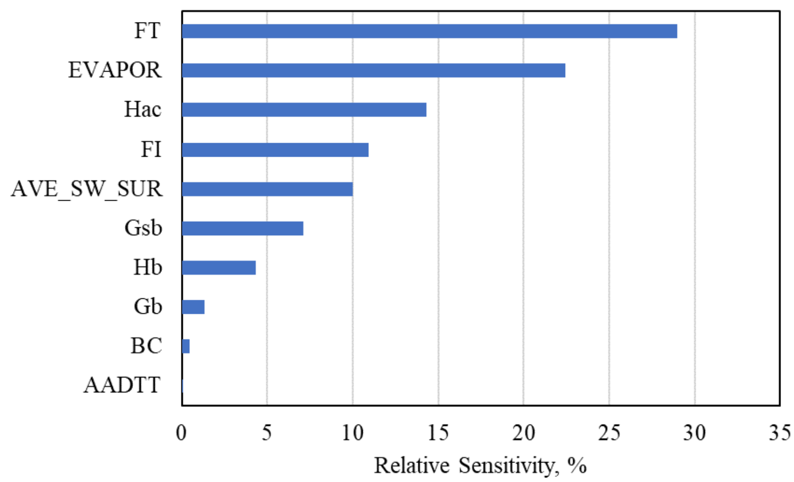
EVAPOR
AADTT
FI
AVE_SW_SUR
Gb
FT
Hb
Hac
Gsb
BC
0
5
10
15
20
25
30
35
Relative Sensitivity, %

(b)



FT
AVE_SW_SUR
Hac
Gb
Gsb
FI
BC
Hb
AADTT
EVAPOR
0
5
10
15
20
25
30
35
40
45
Relative Sensitivity, %

(c)



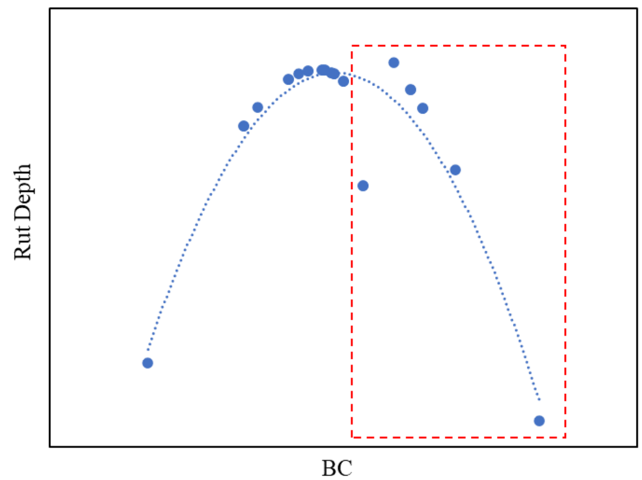
(d)

Figure 4.9 Relative sensitivity of input variables of (a) rutting model; (b) longitudinal cracking model; (c) thermal cracking model; and (d) IRI model

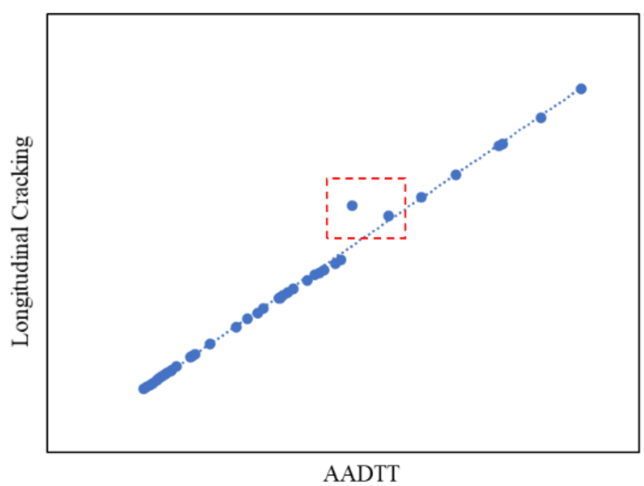
Figure 4.10 depicts the changing pattern of the four distresses with their major contributors. Similar to the relative sensitivity, the pattern was calculated with one input variable over its domain and others fixed at their mean values. However, the changing pattern can show the complex and continuous effect of the variable on the model output which is not indicated by the relative sensitivity. In this study, most of these effects were captured well by the selected models, such as the increases of longitudinal cracking with traffic loads, thermal cracking with radiation and IRI with surface evaporation.

Figure 4.10 also illustrates several defects of the model. The most obvious one is that it provides abnormal predictions as marked in Figure 4.10(a) and Figure 4.10(c). The rut depth and thermal cracking amount were supposed to monotonically increase with binder content (Zhou, Hu et al. 2006) and average radiation (Ling, Chen et al. 2019). Another one is that it overcomplicates the effects of certain variables, which is typically called overfitting (Anderson and Burnham 2002) as marked in Figure 4.10(b). The discontinuity in the longitudinal cracking development with traffic loads cannot be mechanistically interpreted. The former problem primarily stemmed from the outliers in the data for the model construction which reduced the model accuracy, while the latter one resulted from the over pursuit of model accuracy at the expenses of other model properties in the model construction (Anderson and Burnham 2002).

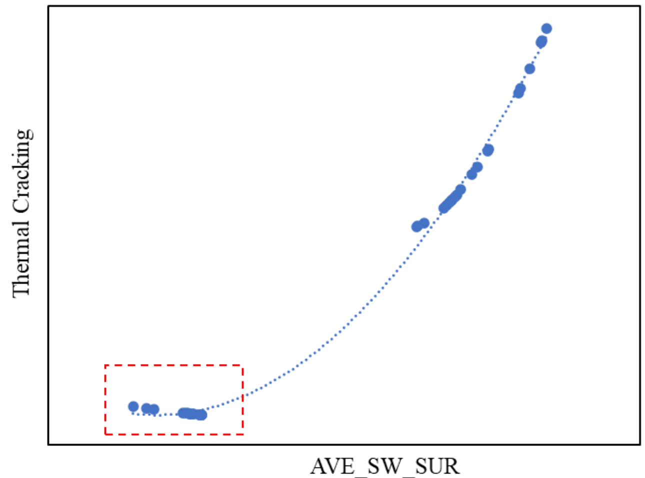
Results in this section present the tradeoff between accuracy and rationality of models constructed by GEP. As indicated by Section 4.1, the balance between the accuracy, complexity and applicability should be focused on for the future predictive models.



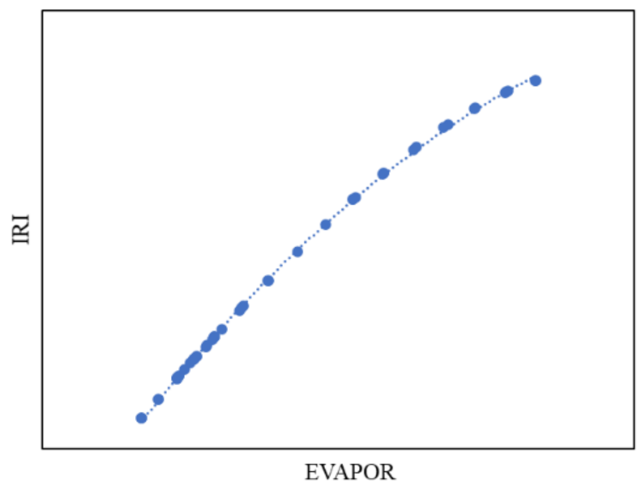
(a)



(b)



(c)



(d)

Figure 4.10 Changing pattern of output with major input in (a) rutting model; (b) longitudinal cracking model; (c) thermal cracking model; and (d) IRI model

4.5 Conclusions

In Chapter 4, predictive models for four typical distresses of asphalt pavements which were generated by GEP were introduced for PMED data. The complete process of model construction and evaluation are described and detailed results and analyses are presented. The major findings from this work are summarized as follows.

- Compared with LR models (R^2 = 0.33 ~ 0.76) and ME models (R^2 = 0.02 ~ 0.60), models generated by GEP have higher and more stable accuracy (R^2 = 0.68 ~ 0.97) for all four distresses;

- Model accuracy on the training dataset cannot guarantee a similar accuracy on the new data. Besides, accuracy reduction and variation on the new data are the main reasons for the non-normal distribution of model residuals;
- Relative sensitivity indicates the major factors causing rutting, longitudinal cracking, thermal cracking and roughness are AADTT (29.5%), surface evaporation (30.8%) and freeze-thaw days (39.7% and 29.0%), respectively;
- Continuous effects of input variables on model outputs can be expressed by the changing pattern of model outputs. Meanwhile, the changing pattern can be utilized to evaluate the model rationality and overfitting.

5. Development of Deep Learning Models for Asphalt Pavement Performance in Idaho

5.1 Introduction

There are two major differences between PMED data and ITD PMS data. The first one is the collection of pavement condition data including material properties, structure configuration, traffic and environmental conditions. They are required as model inputs in ME models as mentioned in Section 2.3. The second one is the existence of maintenance. A maintenance strategy causes an instant change in pavement behaviors and affects subsequent development of distresses. Laboratory experiments with stable and adjustable conditions contribute to comprehensive modeling of pavement materials and structures from construction to failure (Deng, Zhang et al. 2022). Meanwhile, fast developing data mining and processing techniques contribute to accurate prediction of individual distresses and overall condition of pavement from its properties and surrounding environment (Deng and Shi 2022, Deng and Shi 2022). However, models built upon natural deterioration can hardly capture the effects of maintenance actions, of which the time and type are decided by pavement managers and technicians. As shown in Figure 1.2, natural deterioration described by the “Past Deterioration” and “Predicted Deterioration” curves can be modeled by most current models, such as mechanistic-empirical (ME) model (ARA-ERES 2004) and machine learning (ML) model (Deng and Shi 2022). Yet, it is difficult to quantitatively model the effects of maintenance time (e.g., recovered values after “Rehab. A” and “Rehab. D”) and type (e.g., recovered values after “Rehab. A” and “Rehab. B”) on the performance curve. Moreover, material properties and structure configuration of the pavement have changed after maintenance, which makes it difficult to conduct subsequent predictions from historical records.

In general, the occurrence of maintenance action partitions the life-cycle performance of the pavement and results in multiple performance curves with limited relevancy. Accordingly, this study treats them as multiple short-term time series and aims to provide accurate and stable predictions.

The simplest models for time series are polynomials, including univariate ones with specific pavement performance indicators and multivariate ones with additional predictors (Archilla and Madanat 2000). For data of small amounts, desirable fitting accuracy can be easily achieved by various alternative polynomials with basic regression analysis. However, a random selection of polynomials may lead to inaccurate predictions for the follow-up development, because it is difficult to identify which part of the performance curve the modeled data belongs to. To address this issue, the deterioration rate (curve slope) and sensitivity (derivative with respect to predictor) were calculated from previous, especially adjacent datapoints (Abaza 2004, Rogoza 2019). These methods take the advantage of the temporal correlation of data in time series to improve the prediction accuracy and reliability (Rogoza 2019). A similar idea is applied in the Markov chain in which the transition matrix contains deterioration probabilities calculated from adjacent states (Abaza 2016, Wang, Lee et al. 2022).

Another two well-known models with as simple forms as polynomials are exponential smoothing (ES) model and autoregressive integrated moving average (ARIMA) model. Their major difference with the models mentioned above is that the complete historical records can be used in the model construction (Kotu and Deshpande 2018). Exponential smoothing model utilizes the exponentially decaying factors with more weights on the recent data (Brown and Meyer 1961). In the development of ES model, more characteristics of data such as the trend and seasonality can be captured as the number of smoothing parameters increases (Kotu and Deshpande 2018). Autoregressive integrated moving average model combines the autoregressive (AR) model and moving average (MA) model (Box and Jenkins 1970). In addition to the weighted historical data, weighted errors of historical predictions and white noise are introduced in ARIMA model (Hyndman and Athanasopoulos 2018). It can predict time series with trend or seasonality as well by adding additional seasonal terms. It is worth noting that in a previous research (Makridakis, Spiliotis et al. 2018), ES model and ARIMA model outperformed sophisticated machine learning (ML) models in predicting series from the business and economic domains.

With the rapidly-developing artificial intelligence (AI), ML models in predicting pavement performance have gained dramatically increasing applications (Justo-Silva, Ferreira et al. 2021, Marcelino, de Lurdes Antunes et al. 2021). The model output (i.e., pavement performance) can generally achieve desirable prediction accuracy through complex model structures and sophisticated learning algorithms (Murphy 2012). The development of pavement performance was not necessarily treated as time series in current ML models. Using neural networks (NNs) as an example, time effects were reflected on the time as an independent variable (Gong, Sun et al. 2018) or the accumulated/averaged traffic/environmental condition factors (Marcelino, de Lurdes Antunes et al. 2021). Connections were directly built between pavement performance and pavement condition, which were simulated as artificial neurons. As mentioned earlier, these NNs are not suitable for pavements with maintenance. First, the occurrence of maintenance weakens the continuity of time. The variable "time" before and after maintenance have different reference origins and effects on the distress development. Second, pavement after maintenance have different material properties and/or structural configuration with the original one; and the follow-up distresses are developed on the basis of residual damages. It is improper to contain pavement performance before and after maintenance in the same database to train the model.

As with the ES model and ARIMA model, ML models for time series are supposed to learn and capture the characteristics of the series from historical records and make corresponding predictions. This ability is reflected on the architectural ideas of corresponding ML models (LeCun and Bengio 1995). For example, convolutional neural networks (CNNs) can extract image features for recognition and classification problems (Géron 2019). They can also deal with time series to identify, extract and distillate series features for prediction (Brownlee 2018). Recurrent neural networks (RNNs) contain recurrent neurons receiving outputs from the previous time step (Géron 2019). The temporal order and dependencies of the series can be inherently addressed (Hewamalage, Bergmeir et al. 2021). Accordingly, recent cases using these two models and their descendances to predict the development of pavement performance are increasing (Dong, Shao et al. 2019, Bukharin, Yang et al. 2021, Xin, Akiyama et al. 2022).

For a typical distress of asphalt pavement in this study, these two ML models or deep learning (DL) models – CNN and LSTM – were selected to investigate their feasibility in predicting the distress development. The spatial correlation analysis was conducted to transfer raw data to model inputs. The model performance was compared with two statistical models – ES and ARIMA – and improved with three strategies. The rest of this paper is organized as follows: Section 5.2 describes the studied distress and the processing of raw data; Section 5.3 introduces all applied models including their detailed schematics and mathematical equations; Section 5.4 presents the results and analyses from the model construction and comparison; Section 5.5 summarizes the conclusions from this study.

5.2 Data Processing

This section uses rutting as an example. It is directly measured as the permanent deformation from the pavement surface. Previous studies have shown it reflects the material densification, shear flow and failure, and cracking initiation and propagation of layers under traffic loads and at intermediate and high temperatures (Tutumluer and Pan 2008, Deng, Zhang et al. 2022, Zhang, Chen et al. 2022). The field rutting data was collected from the state highway 41 (SH-41) in the state of Idaho and retrieved from the AgileAssets Pavement Analyst software adopted by the Idaho Transportation Department (ITD). State Highway-41, as shown in Figure 5.1, is a state highway running from Interstate 90 in Post Falls to U.S. Route 2 on the Idaho-Washington state line with the total length of 39.06 miles (62.86 km).

Figure 5.1 Illustration of SH-41 (adapted from Google Map and ITD OpenData)

Figure 5.2 presents three series of rutting development measured within one mile (1.61 kilometers) in the longitudinal (driving) direction as examples. As in Figure 1.2, Figure 5.2 illustrates the effects of natural deterioration and maintenance on the rutting development of the pavement. The sudden drops marked by circles in the rut depth indicate the rehabilitation the pavement experiences between current and previous dates of measurement. Otherwise, the rut depth is supposed to continuously increase with pavement service time as marked by rectangles. It can be seen from Figure 5.2 that from 2007 to 2020, this pavement experienced at least three occurrences of rehabilitation. Accordingly, the complete history of rut depth is partitioned into several stages, of which the longest deterioration stage contains 6 datapoints. Therefore, the length of time series in this study is smaller than the ones in previous studies introduced in Section 5.1. It is necessary to comprehensively evaluate and compare the traditional statistical and the DL models in terms of their prediction accuracy and stability.

Figure 5.2 also illustrates the variation of pavement performance in the longitudinal direction as indicated by the differences between these three series starting from 2012. It may result from the variations of construction quality and measuring location, of which the latter reason can also lead to the slightly inconsistent rutting development in the deterioration stage. The combination of series within a certain length in the longitudinal direction can reduce such variations and computational time consumption in this study. Accordingly, the spatial correlation of pavement rutting performance was first investigated to determine the characteristic longitudinal length to process the raw data.

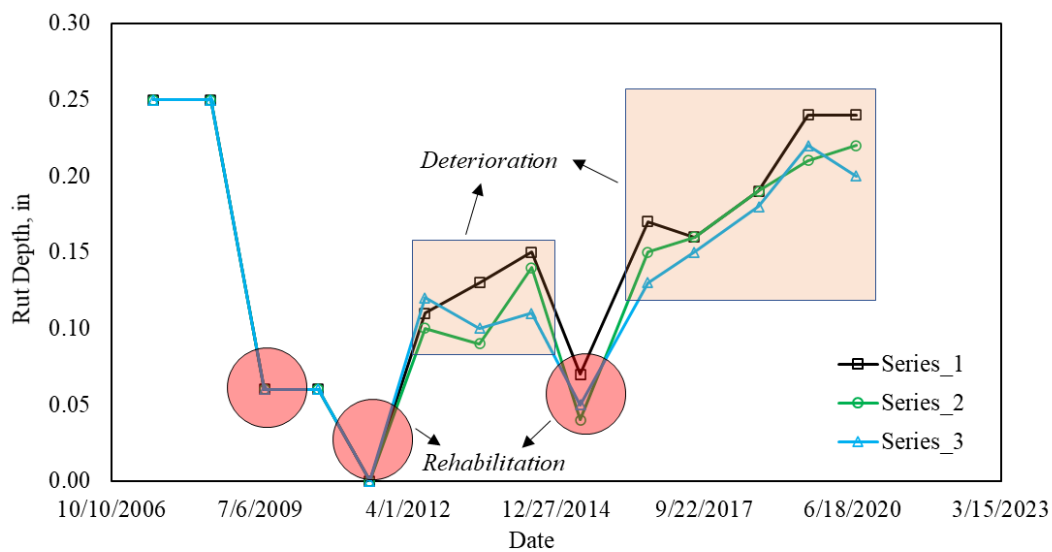


Figure 5.2 Example rutting measurements at different service times

The cosine similarity of vectors defined as the following equation was applied in this section,

$$S_{ij} = \frac{\mathbf{A}_i \bullet \mathbf{A}_j}{\|\mathbf{A}_i\| \|\mathbf{A}_j\|}$$

Where:

$\|A_i\|$, $\|A_j\|$ = the magnitudes of two vectors A_i and A_j

Specifically, the average similarity of the length k was calculated using the following equation,

$$\bar{S}_k = \frac{1}{2C_N^2} \sum_{i,j}^{N} S_{ij}$$

Where:

i, j = values from 1 to N and satisfies $i \neq j$

C^2_N = the number of two-element combinations of N objects

N = the total number of vectors in the length k

Figure 5.3 illustrates two cases of the average similarity with the distance to the start point in one lane. Three distinct stages can be observed from the figure in which the average similarity oscillates in the first stage but remains at a high level. As the data collected far from the start point are involved, the average similarity experiences a continuous drop, which shows the increase of the data variation. In the final stage, the average similarity bounces back and approaches to a steady state, which represents the average similarity of the data of the entire pavement. Therefore, one mile (1.61 kilometers) was selected as the characteristic length in which the measurements at the same time were averaged for the following model construction. As for the transverse direction, the series from ascending and descending lanes were utilized separately considering the variation of the traffic condition (traffic volume, vehicle speed, etc.) in two driving directions.

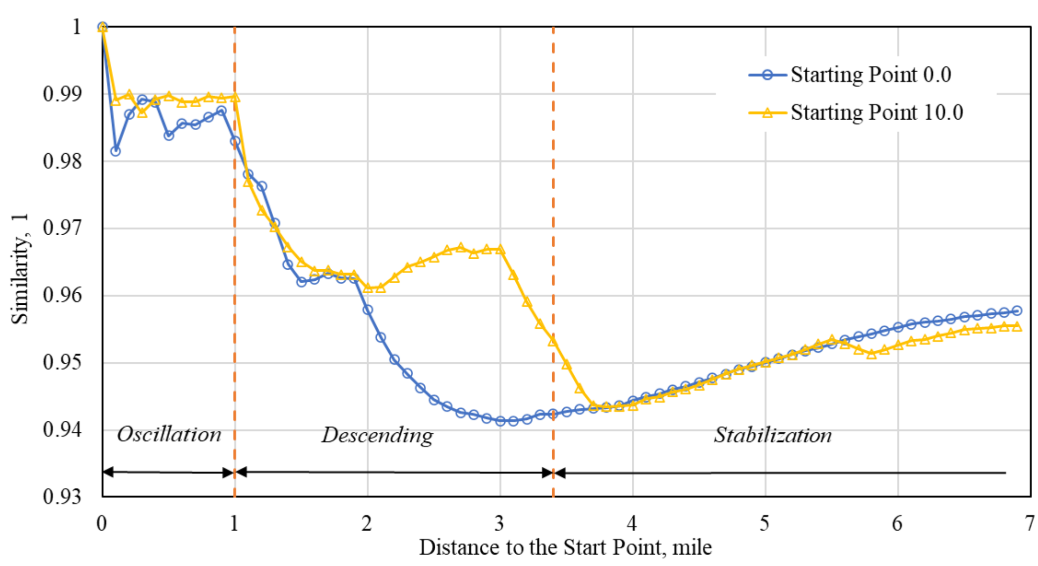


Figure 5.3 Spatial correlation of pavement rutting performance in the longitudinal direction

Table 5.1 presents the summary of organized data from the first 20 miles of two lanes. The shortest series contains four datapoints that can merely construct a model with one two-step training series (the first two steps as the training input, the third one as the training output and the fourth one as the model validation). The longest series contains six datapoints, which is the longest period without major rehabilitation for this pavement. It can be transformed into three types of training series with corresponding number of time steps in one training series and number of training series. Figure 5.4 illustrates the case of $\mathbf{N_{6,3}}$. The rutting data collected in a constant interval (one year) as shown in Figure 5.2 were directly applied in the normalization and then model construction. Otherwise, the missing data were filled via interpolation (Xin, Akiyama et al. 2022). For a better comparison between different input types, five series were selected for each datatype ($\mathbf{N_4}$, $\mathbf{N_5}$ and $\mathbf{N_6}$).

Table 5.1 Information of Collected Data

Series Length	Number of Sections	Number of Time Steps in One Training Series	Number of Training Series	Notation
4	17	2	1	$\mathbf{N_{4,2}}$
5	11	2	2	$\mathbf{N_{5,2}}$
5	11	3	1	$\mathbf{N_{5,3}}$
6	14	2	3	$\mathbf{N_{6,2}}$
6	14	3	2	$\mathbf{N_{6,3}}$
6	14	4	1	$\mathbf{N_{6,4}}$

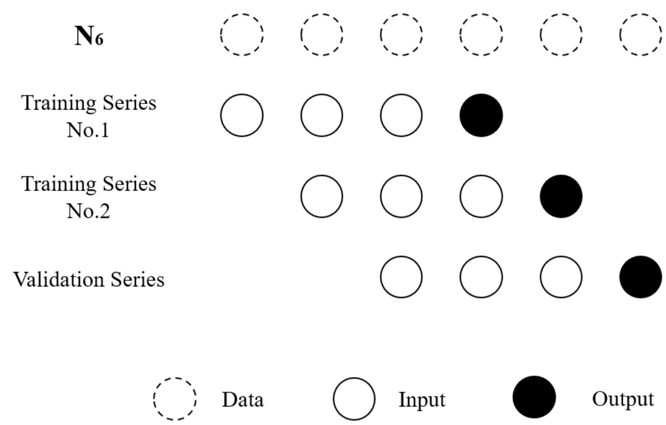


Figure 5.4 Conceptual Diagram of $\mathbf{N_{6,3}}$

5.3 Model Description

5.3.1 ETS and ARIMA

The simplest ES model for time series with no trend or seasonality has the mathematical expression as the following equation (Hyndman and Athanasopoulos 2018),

$$y_{t+1|t} = \sum_{i=1}^{t} \alpha\left(1-\alpha\right)^{t-i} y_i$$

Where:

$y_{t+1|t}$ = prediction for the time *t+1* based on the observation at time *t*

y_i = the observation at time *i*

α = the smoothing parameter

For time series with additional trend, Holt's linear trend model (Holt 2004) can be used as the following equations, in which the prediction for *h* steps ahead of the time *t* is the combination of level estimate (L_t) and trend estimate (T_t) at time *t*,

$$y_{t+h|t} = L_t + hT_t$$

$$L_t = \alpha y_t + \left(1-\alpha\right)\left(L_{t-1} + T_{t-1}\right)$$

$$T_t = \beta\left(L_t - L_{t-1}\right) + \left(1-\beta\right)T_{t-1}$$

Where:

β = the smoothing parameter

A similar approach is applied for the additional seasonality in the Holt-Winters method (Winters 1960).

The ARIMA model can be mathematically expressed as the following equation (Hyndman and Athanasopoulos 2018),

$$y_t^{(d)} = c + \sum_{i=1}^{p} \phi_i y_{t-i}^{(d)} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Where:

$y_{t-i}^{(d)}$ = the *(t-i)*-th observation y_{t-i} after *d* times of differencing

ε_{t-j} = the prediction error of y_{t-j}

ε_t = the white noise for the prediction of y_t

c = constant

φ_i, ϑ_j = model coefficients

The equation above contains the basic formula of AR model and MA model and differencing of observations, in which differencing is typically applied to process time series with trend or seasonality. Non-stationary time series should be transferred to the stationary ones for the construction of ARIMA model (Hyndman and Athanasopoulos 2018). It can be denoted with backshift notation as the following equation.

$$\begin{cases} By_t = y_{t-1} \\ y_t^{(1)} = y_t - y_{t-1} \end{cases} \Rightarrow y_t^{(d)} = (1-B)^d\ y_t$$

Accordingly, the ARIMA model can be rewritten as the following equation, in which the number of AR parts *p*, the degree of differencing *d* and the number of MA parts *q* are three fundamental elements of ARIMA(*p*, *d*, *q*) model.

$$\left[1-\sum_{i=1}^{p}\phi_i B^i\right](1-B)^d\ y_t = c + \left[1+\sum_{j=1}^{q}\theta_j B^j\right]\varepsilon_t$$

In the model descriptions above, trend refers to the long-term developing direction of the data and seasonality refers to the seasonal pattern of the data (Hyndman and Athanasopoulos 2018). Time series with either characteristic have time-dependent statistical properties such as mean and variance and are called non-stationary. The complete rutting development in field pavement from construction to failure does follow a typical pattern (Deng, Zhang et al. 2022). Although the occurrence of maintenance partitions such pattern as shown in Figure 5.2, it is weakly reflected on some residual properties of the applied data such as an increasing trend. Therefore, trend component and differencing were applied in the corresponding ES and ARIMA models in this section.

5.3.2 CNN

The CNN applied in this study is one-dimensional (1D) CNN targeted for 1D signals with lower computational complexity, easier training and implementation than two-dimensional (2D) CNN (Kiranyaz, Avci et al. 2021). Time series can also be seen as images with observations and time distributed on two axes. Characteristics (trend, pattern, etc.) of series can be captured by CNN as features (curve, edge, etc.) of images (Han, Zhao et al. 2019).

Figure 5.5 illustrates the structure of applied 1D CNN with one convolutional layer. As shown in the figure, the raw data as model input are first transformed into an 1D vector as the input layer. It then passes through a convolutional layer in which the input vector is convolved with filters (or kernels) to a set of feature maps. The dot product of one filter containing weights and its overlapped input components is successively calculated to form one feature map in the convolutional layer. Values in the feature map are then processed by an activation function such as Rectified Linear Unit (ReLU) (Fukushima 1969) before passing through the pooling layer, which was designed to reduce the dimension of feature maps. As processed in the convolutional layer, the average (average pooling) or maximum (max pooling) value of the overlapping portion between the filter and feature map is successively calculated. Those processed features are finally flattened to a 1D vector as the input for a fully connected NN. Accordingly, the key structural parameters of the CNN include the size and number of filters and the size of strides (sliding distance of filters) in the convolutional and pooling layers, and as well as the numbers of neurons and layers in the fully connected layer. The weights in the filters and the fully connected NN are learned during the model training.

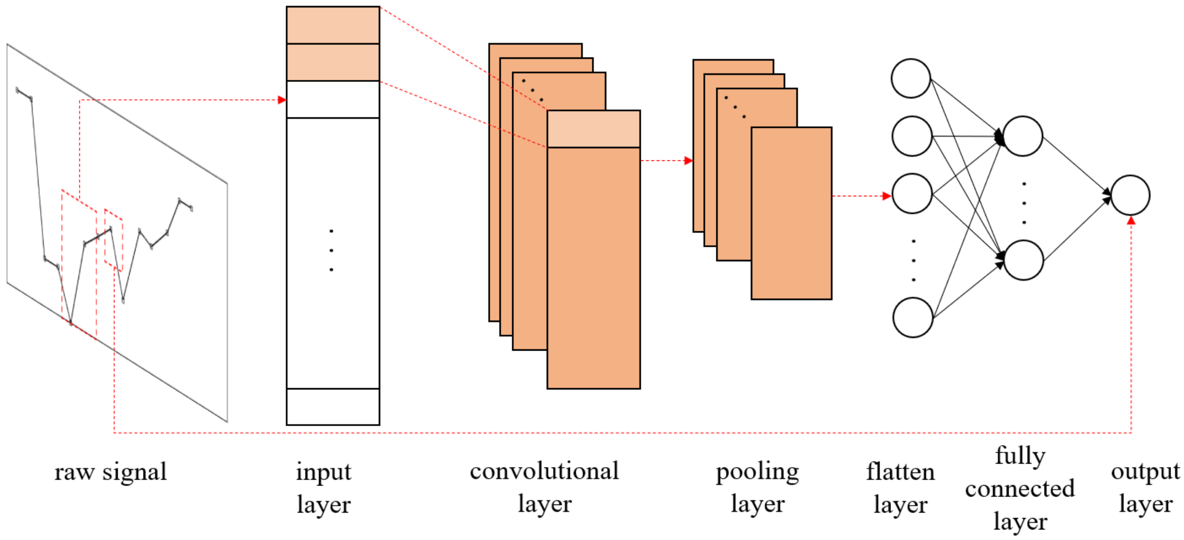


Figure 5.5 Structure of applied 1D CNN (revised from (Chandra, Goyal et al. 2021))

5.3.3 LSTM

Long short-term memory (LSTM) NN is a modified RNN that is capable of storing information in long-term sequences (Hochreiter and Schmidhuber 1997). As shown in Figure 5.6, the current LSTM cell processes the observation *X* as well as the output vector *h* and cell state vector *C* from the previous one. The first step is to calculate the activation vector of the forget gate *f* as the following equation,

$$f_t = \sigma_g \left(W_f X_t + U_f h_{t-1} + b_f \right)$$

Where:

σ_g = the gate activation function (e.g., sigmoid function)
W_f, U_f = the weight matrices
b_f = the bias vector

This step determines the degree of forgetting information of the previous cell state vector. Next, the state cell is updated as the following equation,

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

in which the activation vectors of the update gate i_t and cell input $\tilde{C}_t$ are calculated as the following equations,

$$i_t = \sigma_g \left(W_i X_t + U_i h_{t-1} + b_i \right)$$

$$\tilde{C}_t = \tanh \left(W_C X_t + U_C h_{t-1} + b_C \right)$$

where tanh is the hyperbolic tangent function. This step transforms information from the input to the cell state. Finally, the output of the current step is calculated as the following equation,

$$h_t = o_t \cdot \tanh C_t$$

in which the activation vector of the output gate o_t is calculated as the following equation. This step transforms information from the cell state to the output.

$$o_t = \sigma_g \left(W_o X_t + U_o h_{t-1} + b_o \right)$$

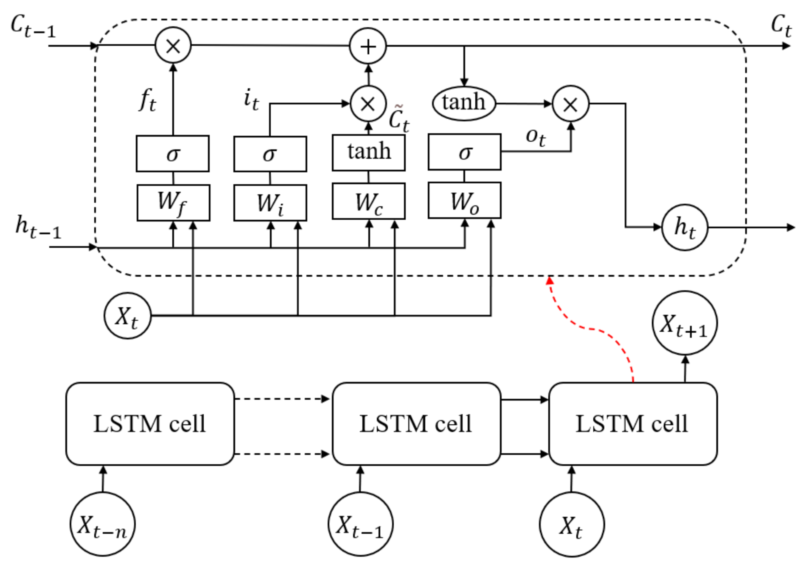


Figure 5.6 LSTM NN Structure with One LSTM Layer (revised from (Pu, Liu et al. 2020))

Compared with traditional RNN in which the information transferred from the previous step to the current one is processed with simple function (e.g., hyperbolic tangent function), the modified structure of the LSTM cell can effectively reduce gradient exploding and vanishing in the model training (Hochreiter and Schmidhuber 1997). Accordingly, the parameters in those weight matrices and bias vectors need to be learned during the model training.

5.3.4 CNN-LSTM and ConvLSTM

In addition to serving as individual models, CNN and LSTM can be combined in hybrid models to predict time series. A typical combination method is that the original series is first processed by a CNN model. Features extracted by convolutional and pooling layers are then flattened to an 1D vector as the input of an LSTM model (Brownlee 2018). It is denoted as the CNN-LSTM model in this study. Considering the information can be lost during the flattening in the CNN-LSTM, convolutional operation is directly applied within the LSTM cell with inputs, cell outputs, cell states, and gates that are set as tensors of higher dimension (Shi, Chen et al. 2015). It was named as ConvLSTM model. Accordingly, the structural and learning parameters of CNN-LSTM and ConvLSTM models are the same as the ones in CNN and LSTM models described in the previous two sections.

It is likely that CNN-LSTM and ConvLSTM models are excessively complicated and unnecessary for problems with as limited data as in this study. Actually, even CNN and LSTM models with the simplest structure have much more coefficients to be calibrated than the data applied in the model training in this study. These models were applied herein to investigate their feasibility in predicting rutting development in field pavement and the effects of increasingly complex models on the model performance. Therefore, model stability and rationality were treated equally as model accuracy to be evaluated in the model evaluation.

5.4 Results and Discussion

Results in this section aim to serve as references for selecting suitable models and hyperparameters for different data types. Furthermore, potential strategies for improving the prediction performance are discussed for future model construction.

5.4.1 Effects of Hyperparameters and Input Type

Hyperparameters of DL models are typically determined first to ensure desirable model performance and provide convenience for a model comparison. Rather than a grid search (Dong, Shao et al. 2019), several typical values were taken for those adjusting hyperparameters to investigate their effects on the model performance. The major reason is that the data type was involved in this study, which has interactive effects with hyperparameters on the model construction and performance. It is explicitly reflected on the fact that the available values of hyperparameters are different in models with different input types. For example, the filter size in the convolutional layer of CNN is controlled by the size of model input (i.e., the number of time steps). The individual effects of the data type should be characterized with fixed values of hyperparameters. Second, it is unlikely that the most desirable hyperparameters determined via a grid search are the same with different model inputs. It is more practical to provide typical values for the corresponding data type in engineering problems.

Type and hyperparameters of the training algorithm and type of the loss function were not adjusted in this study. The adaptive moment estimation (Adam) algorithm (Kingma and Ba 2014) was applied as the training algorithm with default coefficient values in the python DL library – Keras. The mean squared error (MSE) was applied as the loss function.

Table 5.2 shows the hyperparameter values of the cases with the highest model accuracy using the last datapoint of series as the validation data and the root mean square error (RMSE) as the indicator. The epoch number is the number of times that the learning algorithm works through the training dataset (Brownlee 2018). Except for the filter number of CNN, which was valued from 32, 64 and 128, other hyperparameters were valued from 100, 250, and 500. The RMSE of each case with fixed input type and hyperparameters was calculated from 30 predictions (5 series x 6 repetitions). It can be seen from the table that in both CNN and LSTM, there are no certain values of hyperparameters leading to the highest model accuracy.

Table 5.2 Hyperparameters with Best Model Accuracy in (a) CNN and (b) LSTM

(a)

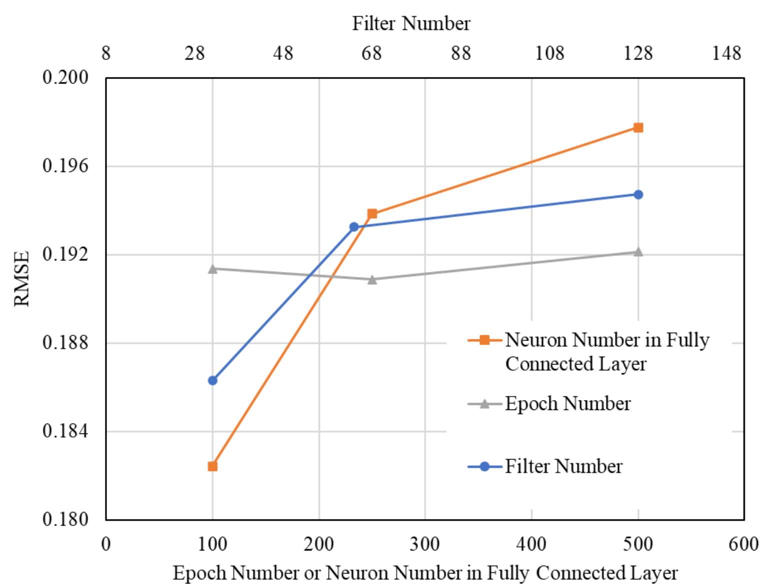
Input Type	Filter Size	Pool Size	Filter Number	Neuron Number in Fully Connected Layer	Epoch Number	RMSE
$N_{4,2}$	2	1	128	250	250	0.179
$N_{5,2}$	2	1	32	500	500	0.194
$N_{5,3}$	2	1	32	100	250	0.360
$N_{5,3}$	2	2	32	100	250	0.245
$N_{5,3}$	3	1	32	100	100	0.223
$N_{6,2}$	2	1	128	250	100	0.053
$N_{6,3}$	2	1	64	100	100	0.066
$N_{6,3}$	2	2	128	100	100	0.063
$N_{6,3}$	3	1	32	100	100	0.065
$N_{6,4}$	2	1	32	100	250	0.108
$N_{6,4}$	2	2	128	500	250	0.113
$N_{6,4}$	3	1	32	100	100	0.085
$N_{6,4}$	3	2	32	100	250	0.089
$N_{6,4}$	4	1	64	100	500	0.085

(b)

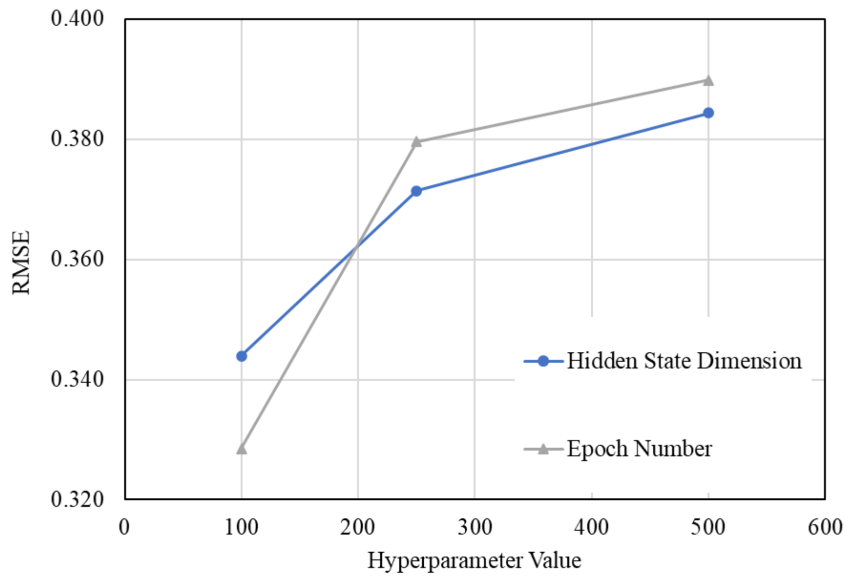
Input Type	Hidden State Dimension	Epoch Number	RMSE
$N_{4,2}$	100	100	0.424
$N_{5,2}$	250	100	0.221
$N_{5,3}$	100	100	0.396
$N_{6,2}$	500	500	0.071
$N_{6,3}$	100	100	0.064
$N_{6,4}$	100	250	0.242

A clearer illustration of the hyperparameter effects on model accuracy is presented in Figure 5.7, in which the cases with different input types were combined. Accordingly, the RMSE of each set of hyperparameters in CNN and LSTM was calculated from 3780 and 540 predictions, respectively. A general increasing trend can be observed from the graph indicating that model accuracy decreases with the increases of filter, neuron, and epoch numbers and LSTM hidden state dimension. However, this trend is negatively affected by the input type and therefore is not clearly reflected on model accuracy in Table 5.2.

As for the individual effect of input type, it can be similarly calculated by combining cases with different hyperparameters, as shown in Figure 5.8. Input type can be represented by its two properties – input length ("Number of Time Steps") and input number ("Number of Inputs") as indicated in Table 5.1. It is also indicated in Table 5.1 that the possible combinations of input length and input number for a certain series are fixed and determined by the length of the series. Model inputs with the input length of 2, 3, and 4 and the input number of 1 were extracted from data type $\mathbf{N}_{4,2}$, $\mathbf{N}_{5,3}$, and $\mathbf{N}_{6,4}$, respectively. Model inputs with the input number of 1, 2, and 3 and the input length of 2 were extracted from data type $\mathbf{N}_{4,2}$, $\mathbf{N}_{5,2}$, and $\mathbf{N}_{6,2}$, respectively. It can be seen from Figure 5.8 that the larger the input number or the input length, the higher the model accuracy. Finally, the balance was achieved in CNN and LSTM that model input with the input length and input number of 2 and 3 (or 3 and 2) resulted in the highest model accuracy as shown in Table 5.2.

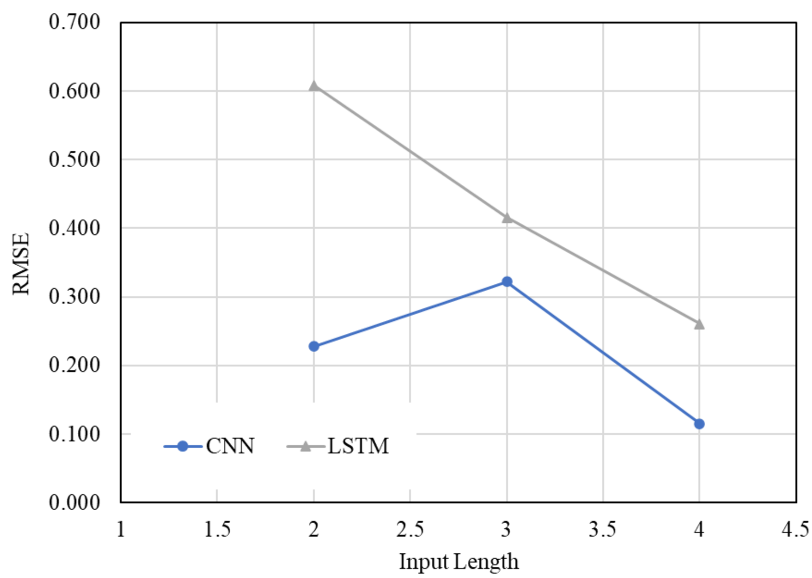


(a)

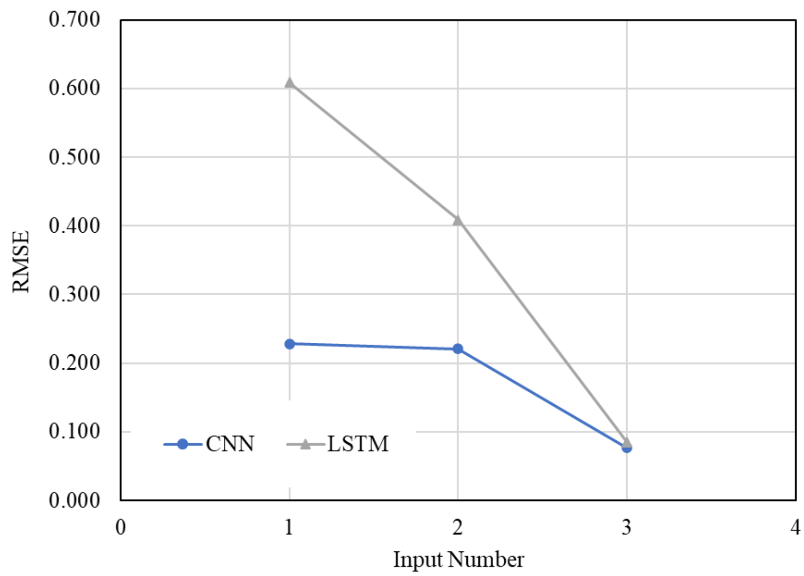


(b)

Figure 5.7 Effect of hyperparameters in (a) CNN and (b) LSTM



(a)

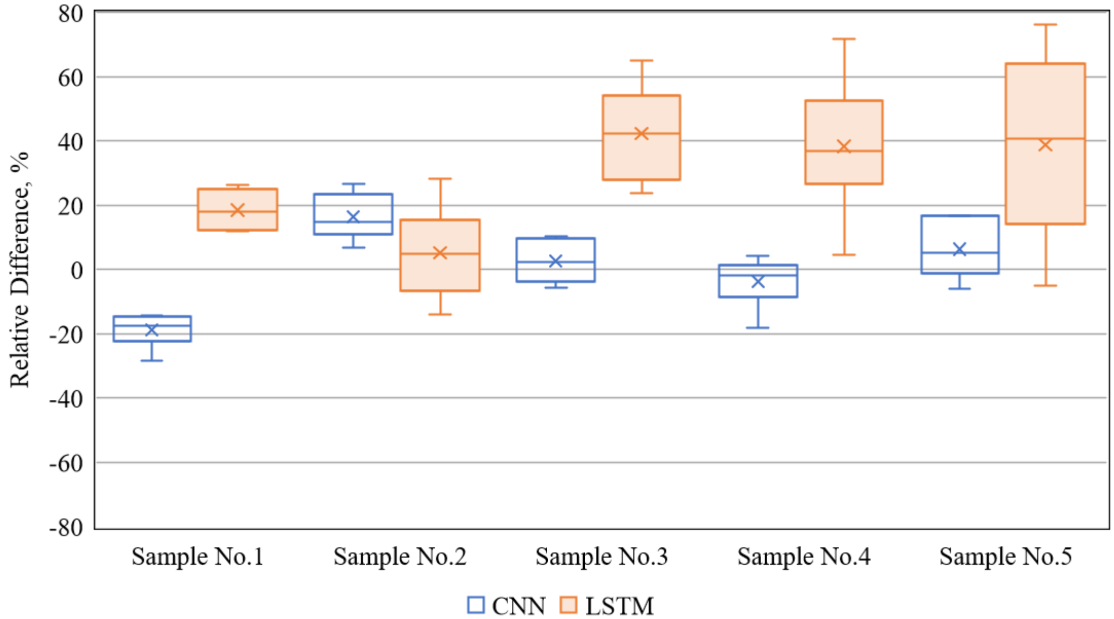


(b)

Figure 5.8 Effect of input type: (a) input length and (b) input number

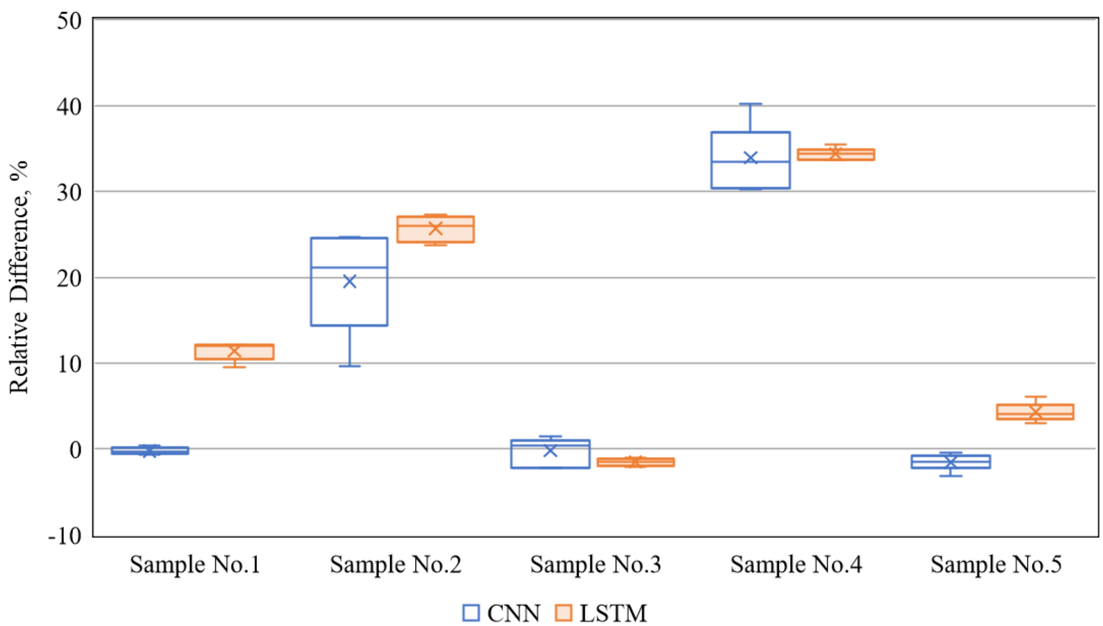
5.4.2 Model Comparison

In this section, the comparison of model accuracy between two applied DL models and two statistical models is presented. Five sample series were randomly selected from series of different lengths to be analyzed. As in the previous section, the validation accuracy was utilized, which was calculated using the last datapoint of each sample series and the model trained with all previous datapoints. To give a general idea of accuracy level, the relative difference between the predicted and measured data in percentage replaced RMSE as the indicator. Results of CNN and LSTM with different input lengths are presented by the boxplot in Figure 5.9. For each input length, the case with the lowest RMSE was selected as marked in bold in Table 5.2. They represent the highest average accuracy CNN and LSTM can achieve with these lengths of inputs. Two key findings can be drawn from Figure 5.9. First, CNN achieved better accuracy than LSTM in most cases, which is indicated by the mean and deviation of the relative difference, and second, model accuracy relied heavily on the shape of input. For example, all predictions from LSTM for Sample No.1, No. 3, and No. 4 of the input type $\mathbf{N_4}$ are higher than the measurements, while for the other two sample series the differences between predictions and measurements distribute across zero line. This indicates the risk of obtaining constantly overestimated or underestimated predictions for certain shapes of rutting development. In general, as input length increases, the relative difference can reach a desirable below 8%.



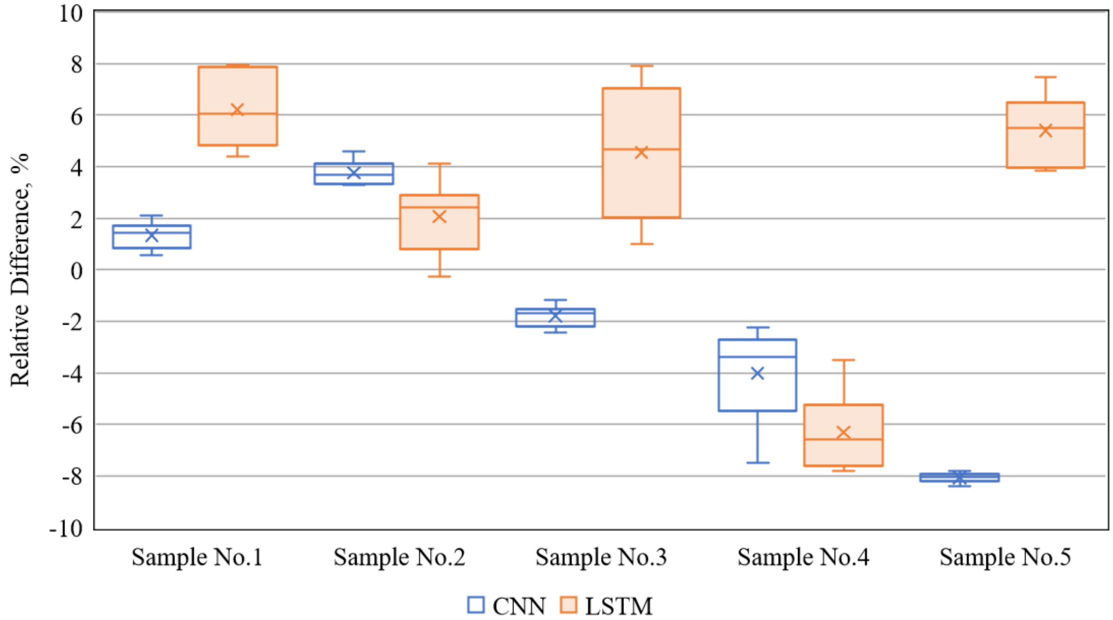
80
60
40
20
0
-20
-40
-60
-80
Relative Difference, %
Sample No.1
Sample No.2
Sample No.3
Sample No.4
Sample No.5
CNN
LSTM

(a)



50
40
30
20
10
0
-10
Relative Difference, %
Sample No.1
Sample No.2
Sample No.3
Sample No.4
Sample No.5
CNN
LSTM

(b)



(c)

Figure 5.9 Comparison between CNN and LSTM with Input Type: (a) N_4, (b) N_5 and (c) N_6

For performance comparison with statistical models, the data in the series $\mathbf{N_6}$ was utilized to calibrate two smoothing parameters of ES model with trend (ETS) and construct ARIMA(2,1,0) model. Since the parameter values in these two models were optimized using all the training data (the first five) in the series, there is no deviation in repetitive predictions by ETS and ARIMA models, as shown in Figure 5.10. It can also be seen from Figure 5.10 that two statistical models have no superiority over CNN and LSTM in predicting the rutting development in this study.

In addition to statistical models, we also conducted a comparison study of CNN models against piece-wise linear regression models currently adopted by ITD for prediction of overall condition index (OCI) of selected ITD asphalt pavement sections, as detailed in Appendix C. CNN models achieved higher prediction accuracy than the piece-wise regression models, for the historical data in the series $\mathbf{N_5}$ and **N6**.

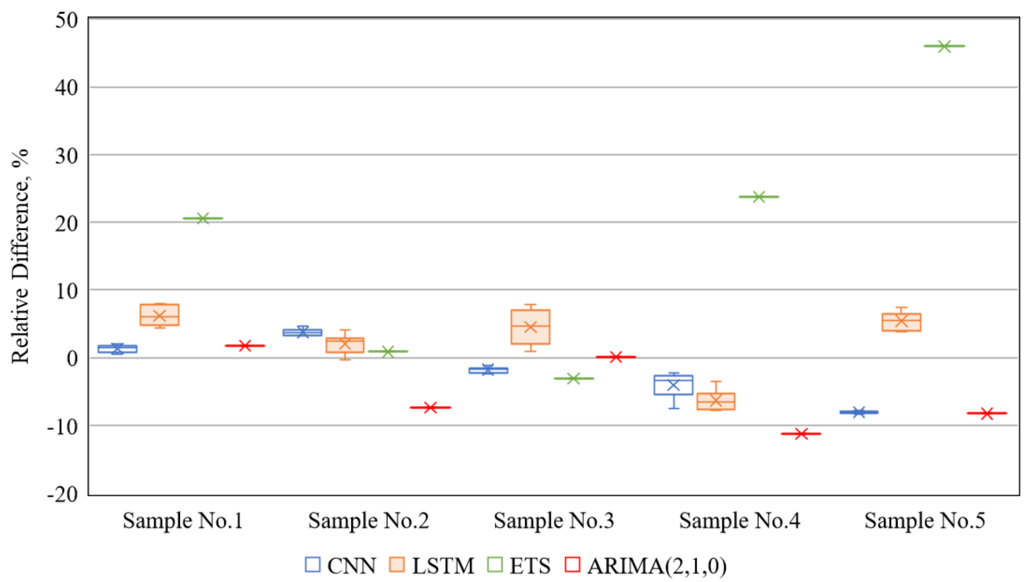
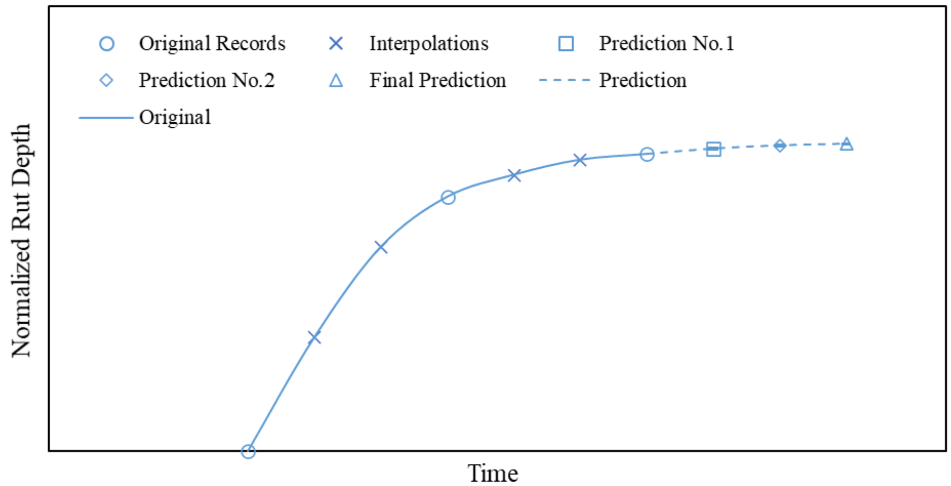


Figure 5.10 Comparison with statistical models

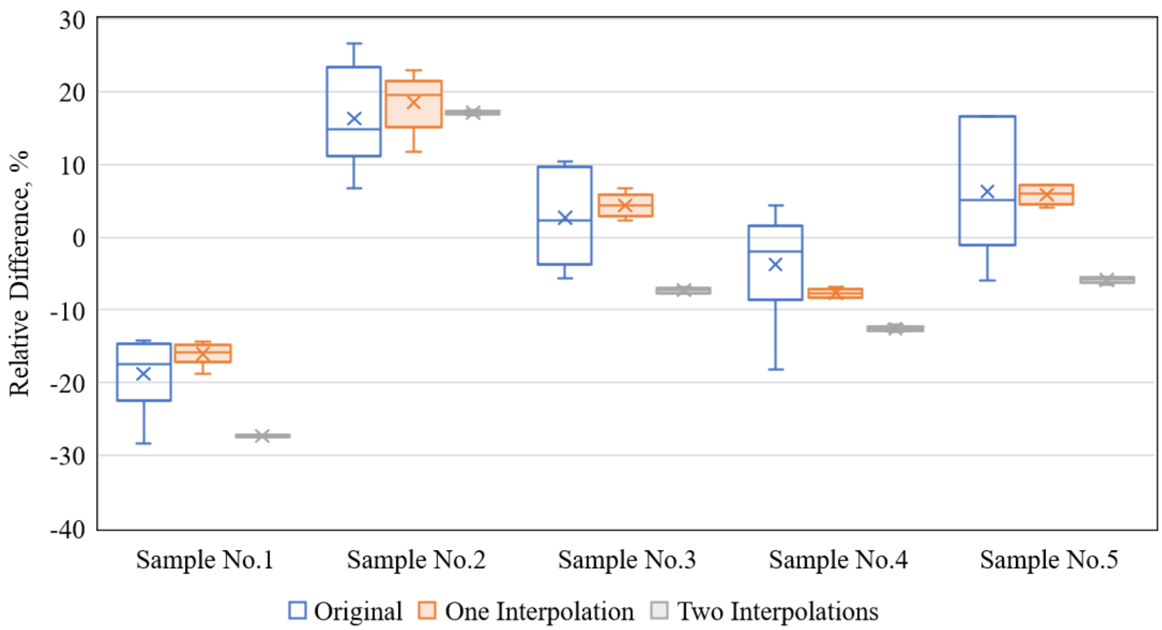
5.4.3 Strategies for Improving the Model Performance

Results in the previous two sections indicate that DL models are worthwhile to be utilized in predicting short-term rutting development. However, data quantity is a major issue limiting their performance. Accordingly, three potential strategies were adopted in this study to investigate their contributions to improving model accuracy and stability.

The first strategy is increasing the input length and number by adding interpolations in the interval. One and two interpolations were separately added in the series $\mathbf{N_4}$ using the cubic spline interpolation method (Hall and Meyer 1976). Accordingly, the input length was increased to five and seven, respectively and the corresponding optimal hyperparameters and input types were set as in Table 5.2. Model performance with modified inputs is presented in Figure 5.11. It can be seen that model stability was improved most as the deviation of repetitive predictions decreased significantly. Model accuracy was improved significantly in some cases such as Series Nos. 3, 4, and 5 with LSTM, but was worsened slightly in some cases such as Series No.1 with LSTM. It can be explained by the fact that the model accuracy relies heavily on the shape of input, which cannot be further described by the added interpolations. Another reason is illustrated in Figure 5.11(a) that with interpolations, the final prediction is not necessarily the only prediction from the model. Perquisite predictions with the equal number of interpolations in the interval are produced first and then serve as the inputs for the final prediction, as “Prediction No. 1” and “Prediction No. 2” in Figure 5.11(a). Errors of these perquisite predictions can accumulate and be reflected on the final prediction.


Original Records
Interpolations
Prediction No.1
Prediction No.2
Final Prediction
Prediction
Original
Normalized Rut Depth
Time

(a)


30
20
10
0
-10
-20
-30
-40
Relative Difference, %
Sample No.1
Sample No.2
Sample No.3
Sample No.4
Sample No.5
Original
One Interpolation
Two Interpolations

(b)

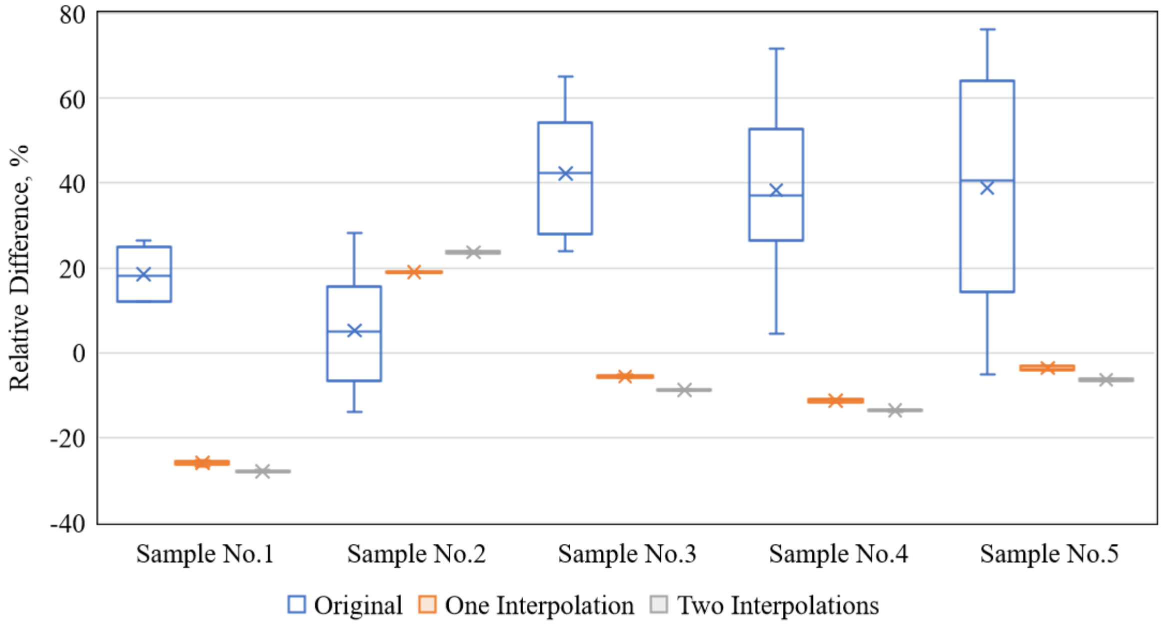


Figure 5.11 Comparison with interpolated inputs: (a) example of original data with two interpolations in the interval, (b) CNN and (c) LSTM

The second strategy is increasing the input dimensionality by adding parallel series of pavement exposure conditions as model input. Two factors – average annual daily traffic (AADT) and freeze-thaw days – were selected as examples to represent traffic and environmental conditions of the pavement for their high impacts on the rutting development (Deng and Shi 2022). They were separately retrieved from ITD Opendata and Long-term Pavement Performance (LTPP) database (Luo, Wang et al. 2022) from the location nearest to the one from which Sample No.1 of $\mathbf{N_4}$ was collected. In this study, multiple input series were processed by individual CNN models and their outputs were then combined for the final prediction. While in LSTM, the multiple input series were directly treated as separate variables (Brownlee 2018). Figure 5.12 shows that with enriched model input, the model's prediction accuracy and stability achieved visible improvement. As fully connected NNs, the contributions of various pavement exposure factors to the rutting development are considered. However, the time effect is dispersedly reflected on the development of individual factors rather than serving as an independent factor. Another advantage of this method is that compared with rutting development, the characterization and prediction of individual pavement exposure conditions (e.g., traffic volume and freeze-thaw days) are more established (George and Santra 2020, Deng, Shi et al. 2021). Accurate and reliable prediction of pavement exposure conditions can contribute to better prediction of rutting development.

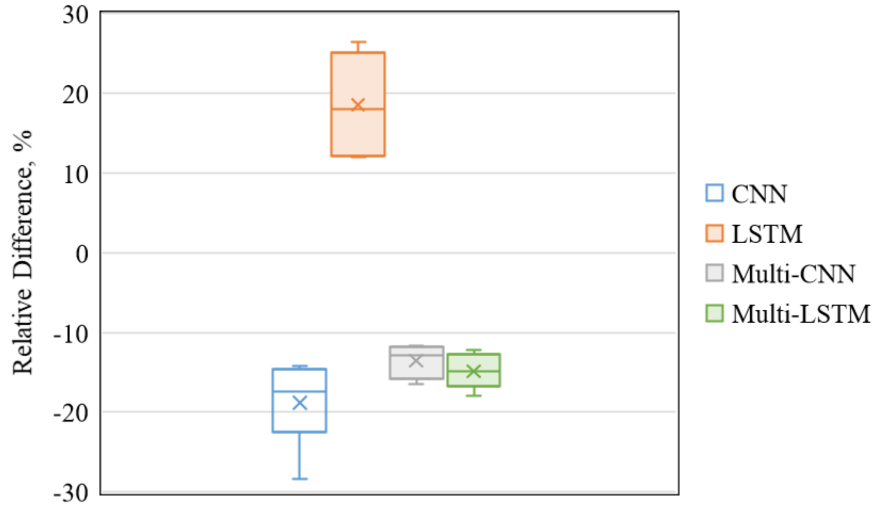


Figure 5.12 Comparison with multiple inputs

The third strategy is increasing the model complexity by applying two hybrid models - CNN-LSTM and ConvLSTM - as introduced in Section 5.3.4. As shown in Figure 5.13, similar to the strategy of adding interpolations, the major improvement was achieved in model stability. The accuracy level of CNN-LSTM is close to that of CNN. Data quantity may be the reason limiting the superiority of hybrid models.

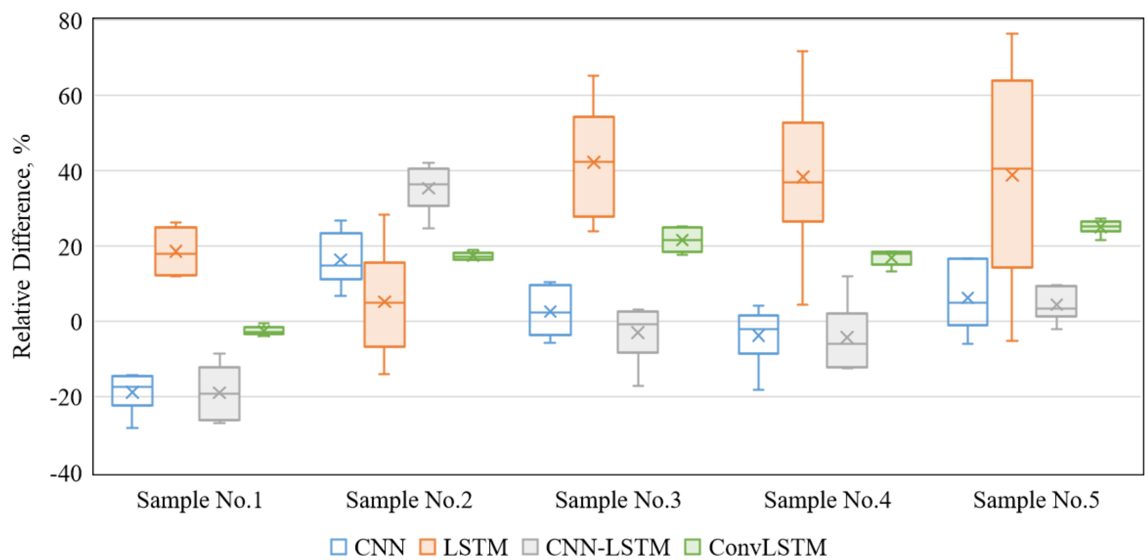


Figure 5.13 Comparison with hybrid models

5.5 Conclusions

In Chapter 5, two DL models: CNN and LSTM in predicting short-term rutting development of a field asphalt pavement were utilized for ITD PMS data. Effects of model hyperparameters and inputs were characterized for an optimal setting for the corresponding data. Model performance was evaluated and compared within DL models and with the statistical models ETS and ARIMA. Three potential strategies were utilized to improve prediction performance. The major findings from this work are summarized as follows.

- The average cosine similarity of rutting development in the longitudinal direction of the pavement has three distinct stages. It indicates the spatial correlation of pavement performance and can capture the characteristic length of pavement to process collected data. One mile (1.61 meters) was selected as the characteristic length of averaging the collected data for the highway pavement in this study. Integration of the raw data effectively mitigated the data inconsistency caused by measuring errors and simplified the model construction;
- Three types of series were collected from the pavement, which were further converted to one, two, and three types of model inputs with the corresponding input lengths and numbers ;
- In CNN and LSTM, the model accuracy decreases with the increases of filter, neuron, and epoch numbers and LSTM hidden state dimension. This trend is negatively affected by the effects of model input, in which model accuracy increases with the increases of input length and input number;
- CNN outperformed LSTM in most cases in which the average relative difference between predicted and measured rut depths was within the range (-18.8% to +16.3%), (-1.5% to +33.9%) and (-8.1% to +3.8%) for the series with four, five, and six datapoints, respectively. In LSTM, these three ranges were (+5.2% to +42.3%), (-1.5% to +34.4%) and (-6.3% to +6.2%), respectively. The statistical models ETS and ARIMA showed no superiority over these two DL models, and their range was (-11.2% to +46.0%) for the series with six datapoints;
- Increasing the data quantity by adding interpolations and increasing the model complexity (by using hybrid models) mainly improved the model's stability. As for model accuracy, the only case of improvement is that the average relative difference range shrank to (-2.5% to +25.1%) for the series with four datapoints by ConvLSTM. In comparison, both model accuracy and stability were improved by the addition of parallel series of pavement condition as model inputs. One single series was tried and the average relative difference range reduced from (-18.8% to +18.6%) to (-14.9% to -13.5%) by adding the AADT and the number of freeze-thaw days.

6. Summary, Conclusions and Recommendations

6.1 Summary

The project is to develop reliable and realistic and enhanced performance curves for ITD asphalt pavements by mining the historical data. To this end, the project reviewed currently applied predictive models in terms of their forms, applications, advantages and limitations. According to characteristics of historical data collected by ITD, models of different types were utilized and compared with traditional models. The flow of major work in this project can be expressed as Figure 6.1. Model introduction, construction, evaluation and comparison were recorded with texts, figures and tables in individual chapters.

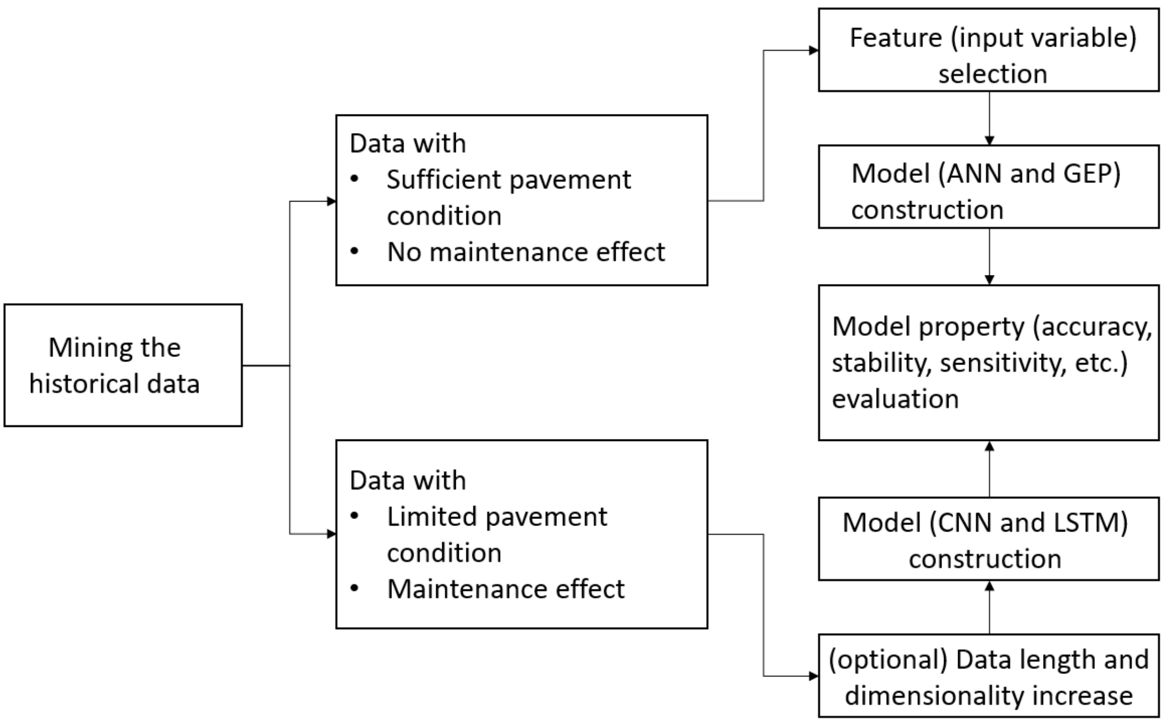


Figure 6.1 Flow of work in this project

6.2 Conclusions

Details of findings and conclusions from each task are presented in the last section of each chapter. In this section, a summary of conclusions from this study is presented.

In Chapter 2, a literature review on current predictive models of asphalt pavement performance and a practitioner survey on the insights and experiences of users on the existing models indicate the basic qualities of a predictive model should have to be applied in practice. The reason why ME model is widely applied is that it takes advantages of mechanical model and empirical model with basic accuracy, rationality and simplicity. Meanwhile, numerical model plays an important part in providing pavement responses. Machine learning model, as a product of cutting-edge technology, takes advantages of the development of artificial intelligence and computing power. It has both promising applications and potential problems to be considered and solved.

In Chapter 3, NNs calibrated with particle swarm optimization was utilized to predict the asphalt pavement rutting in Idaho using PMED data. Performance of models with different calibrated parameters, accuracy and calibration time was evaluated and compared. Meanwhile, a three-step variable selection method with PCA and correlation analysis was utilized to reduce the input dimensionality. It serves as an example of using machine learning model and algorithm to construct predictive model with sufficient data. It indicates that during the design of NN models, one should consider the number of hidden neurons to balance the model accuracy and reproducibility, which are both important for model application.

In Chapter 4, GEP was utilized to construct predictive models for four typical distresses of asphalt pavements in Idaho using PMED data. Compared with neural networks, less coefficients and more operations were applied in the GEP models. As another example of artificial intelligence algorithm in selecting model form and calibrating model coefficient, GEP models gained higher accuracy than linear regression and ME models. This study also indicates the importance of checking the overall model accuracy with both the training and validation datasets.

In Chapter 5, deep learning models - CNN and LSTM were utilized in predicting short-term rutting development of a field asphalt pavement with ITD PMS data. Potential strategies such as increasing data quantity and dimensionality and model complexity were utilized to improve prediction performance as well. This study points out the necessity of considering spatial correlation of pavement performance to mitigate measuring error. When an agency has data with limited quantity, lacking pavement condition and maintenance effect, deep learning models are worth applying in light of their higher prediction accuracy and stability than statistical models.

6.3 Recommendations for Implementation

- Characteristics of data should be investigated before applying predictive models. As for pavement performance, the focuses can be put on the availability of material, structure, traffic and environment conditions as well as maintenance effect. Different model types have different applicability and performance for different data.
- Artificial intelligence model and algorithm are promising in predict pavement performance for the accuracy, efficiency and automation. AI can be applied in model form selection, model calibration, etc. in different ways and to different degrees.

- To avoid overfitting and ensure basic rationality of predictive models, statistical methods are necessary to check the stability, robustness, sensitivity, etc. of constructed models before application.
- Models introduced in this project can extend the application in terms of the distress type, pavement type and areas of interest, as the research team will provide the associated codes and instructions as deliverables in addition to this project final report.
- While outside the scope of this project, a follow-up project could develop an interactive toolkit with user-friendly interface, to facilitate the implementation of the models developed in this project.

References

Abaza, K. (2002). "Optimum flexible pavement life-cycle analysis model." Journal of Transportation Engineering **128**(6): 542-549.
Abaza, K., I. Khatib and S. Ashur (2004). "Integrated pavement management system with a Markovian prediction model." Journal of Transportation Engineering **130**(1): 24-33.
Abaza, K. A. (2004). "Deterministic performance prediction model for rehabilitation and management of flexible pavement." International Journal of Pavement Engineering **5**(2): 111-121.
Abaza, K. A. (2016). "Simplified staged-homogenous Markov model for flexible pavement performance prediction." Road Materials and Pavement Design **17**(2): 365-381.
Abu Al-Rub, R. K., M. K. Darabi, C.-W. Huang, E. A. Masad and D. N. Little (2012). "Comparing finite element and constitutive modelling techniques for predicting rutting of asphalt pavements." International Journal of Pavement Engineering **13**(4): 322-338.
Al-Rub, R. K. A. and M. K. Darabi (2012). "A thermodynamic framework for constitutive modeling of time-and rate-dependent materials. Part I: Theory." International Journal of Plasticity **34**: 61-92.
Alharbi, F. (2018). Predicting pavement performance utilizing artificial neural network (ANN) models. Ph.D., Iowa State University.
Ali, B., M. Sadek and I. Shahrour (2009). "Finite-element model for urban pavement rutting: Analysis of pavement rehabilitation methods." Journal of transportation engineering **135**(4): 235-239.
Alimoradi, S., A. Golroo and S. M. Asgharzadeh (2020). "Development of pavement roughness master curves using Markov Chain." International Journal of Pavement Engineering: 1-11.
Amin, S. and M. Ajakaiye (2020). Machine Learning Algorithms for Rutting Modelling of Bituminous Pavements in West Midlands. 19th Annual International Conference on Highways and Airport Pavement Engineering, Asphalt Technology and Infrastructure, Liverpool, UK.
Anderson, D. and K. Burnham (2002). Model Selection and Multimodel Inference. New York, NY, Springer-Verlag New York, Inc.
ARA-ERES (2004). Guide for mechanistic–empirical design of new and rehabilitated pavement structures. Washington, D.C., National Copperative Highway Research Program, Transportation Research Board, National Research Council.
Archilla, A. R. (2006). "Repeated measurement data analysis in pavement deterioration modeling." Journal of infrastructure systems **12**(3): 163-173.
Archilla, A. R. and S. Madanat (2000). "Development of a pavement rutting model from experimental data." Journal of transportation engineering **126**(4): 291-299.
Aslam, F., F. Farooq, M. N. Amin, K. Khan, A. Waheed, A. Akbar, M. F. Javed, R. Alyousef and H. Alabdulijabbar (2020). "Applications of gene expression programming for estimating compressive strength of high-strength concrete." Advances in Civil Engineering **2020**.
Bayomy, F., A. Muftah, E. Kassem, F. Tousef and H. Alkuime (2018). Calibration of the AASHTOWare Pavement ME Design Performance Models for Flexible Pavements in Idaho. Moscow, ID, University of Idaho.
Behnke, R., I. Wollny, F. Hartung and M. Kaliske (2019). "Thermo-mechanical finite element prediction of the structural long-term response of asphalt pavements subjected to periodic traffic load: Tire-pavement interaction and rutting." Computers and Structures **218**: 9-31.
Bigl, S. R. and R. L. Berg (1996). Modeling of Mn/ROAD Test Sections with the CRREL Mechanistic Pavement Design Procedure. Hanover, NH, Cold Regions Research & Engineering Laboratory.
Box, G. E. and G. M. Jenkins (1970). Time Series Analysis: Forecasting and Control. San Francisco, CA, Holden-Day, Inc.

Brown, R. G. and R. F. Meyer (1961). "The fundamental theorem of exponential smoothing." Operations Research **9**(5): 673-685.
Brownlee, J. (2018). Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python, Machine Learning Mastery.
Bukharin, A. W., Z. Yang and Y. Tsai (2021). "Five-year project-level statewide pavement performance forecasting using a two-stage machine learning approach based on long short-term memory." Transportation Research Record **2675**(11): 280-290.
Cao, H., X. Qian, Z. Chen and H. Zhu (2017). "Layout and size optimization of suspension bridges based on coupled modelling approach and enhanced particle swarm optimization." Engineering Structures **146**: 170-183.
Ceylan, H., M. B. Bayrak and K. Gopalakrishnan (2014). "Neural Networks Applications in Pavement Engineering: A Recent Survey." International Journal of Pavement Research and Technology **7**(6): 434-444.
Chandra, R., S. Goyal and R. Gupta (2021). "Evaluation of deep learning models for multi-step ahead time series prediction." IEEE Access **9**: 83105-83123.
Chen, S., Y. Ren, D. Friedrich, Z. Yu and J. Yu (2020). "Sensitivity analysis to reduce duplicated features in ANN training for district heat demand prediction." Energy and AI **2**: 100028.
Choi, S. and M. Do (2020). "Development of the road pavement deterioration model based on the deep learning method." Electronics **9**(1): 3.
Cleveland, W. S. and S. J. Devlin (1988). "Locally weighted regression: an approach to regression analysis by local fitting." Journal of the American Statistical Association **83**(403): 596-610.
Collop, A. and D. Cebon (1995). "A model of whole-life flexible pavement performance." Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science **209**(6): 389-407.
Cox, D. R. and E. J. Snell (1968). "A general definition of residuals." Journal of the Royal Statistical Society: Series B **30**(2): 248-265.
Croarkin, C. and P. Tobias (2012). The NIST/SEMATECH e-Handbook of Statistical Methods, National Institute of Standards and Technology.
Darabi, M. K., R. K. A. Al - Rub, E. A. Masad and D. N. Little (2012). "Thermodynamic - based model for coupling temperature - dependent viscoelastic, viscoplastic, and viscodamage constitutive behavior of asphalt mixtures." International Journal for Numerical Analytical Methods in Geomechanics **36**(7): 817-854.
Deacon, J. A., J. T. Harvey, I. Guada, L. Popescu and C. L. Monismith (2002). "Analytically based approach to rutting prediction." Transportation research record **1806**(1): 9-18.
Deng, Y., X. Luo, F. Gu, Y. Zhang and R. L. Lytton (2019). "3D simulation of deflection basin of pavements under high-speed moving loads." Construction and Building Materials **226**: 868-878.
Deng, Y., X. Luo, Y. Zhang, S. Cai, K. Huang, X. Shi and R. L. Lytton (2021). "Determination of flexible pavement deterioration conditions using Long - Term Pavement Performance database and artificial intelligence - based finite element model updating." Structural Control and Health Monitoring **28**(2): e2671.
Deng, Y., X. Luo, Y. Zhang and R. L. Lytton (2020). "Determination of complex modulus gradients of flexible pavements using falling weight deflectometer and artificial intelligence." Materials and Structures **53**(4): 1-17.
Deng, Y., X. Luo, Y. Zhang and R. L. Lytton (2021). "Evaluation of flexible pavement deterioration conditions using deflection profiles under moving loads." Transportation Geotechnics **26**: 100434.

Deng, Y. and X. Shi (2022). "An accurate, reproducible and robust model to predict the rutting of asphalt pavement: neural networks coupled with particle swarm optimization." IEEE Intelligent Transportation Systems Transactions **23**(11): 22063-22072.
Deng, Y. and X. Shi (2022). "Development of predictive models of asphalt pavement distresses in Idaho through gene expression programming." Neural Computing and Applications(34): 14913–14927.
Deng, Y., X. Shi and Y. Zhang (2021). "Determination of composite modulus of concrete containing reclaimed asphalt pavement using micromechanical modelling and soft computing techniques." Journal of Cleaner Production: 129486.
Deng, Y., X. Shi, Y. Zhang and J. Chen (2021). "Numerical modelling of rutting performance of asphalt concrete pavement containing phase change material." Engineering with Computers: 1-16.
Deng, Y., Y. Zhang, X. Luo and R. L. Lytton (2022). "Development of equivalent stationary dynamic loads for moving vehicular loads using artificial intelligence-based finite element model updating." Engineering with Computers **38**(4): 2955-2974.
Deng, Y., Y. Zhang, X. Shi, S. Hou and R. L. Lytton (2022). "Stress–strain dependent rutting prediction models for multi-layer structures of asphalt mixtures." International Journal of Pavement Engineering **23**(8): 2728-2745.
Do, L. N., N. Taherifar and H. L. Vu (2019). "Survey of neural network‐based models for short‐term traffic state prediction." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **9**(1): e1285.
Dong, Y., Y. Shao, X. Li, S. Li, L. Quan, W. Zhang and J. Du (2019). Forecasting pavement performance with a feature fusion LSTM-BPNN model. Proceedings of the 28th ACM international conference on information and knowledge management, New York, NY, Association for Computing Machinery.
Dongré, R., J. D'Angelo and A. Copeland (2009). "Refinement of flow number as determined by asphalt mixture performance tester: Use in routine quality control–quality assurance practice." Transportation Research Record **2127**(1): 127-136.
Doré, G., J.-M. Konrad and M. Roy (1999). "Deterioration model for pavements in frost conditions." Transportation Research Record **1655**(1): 110-117.
El-Badawy, S., A. Awed and F. Bayomy (2011). Evaluation of the MEPDG dynamic modulus prediction models for asphalt concrete mixtures. Transportation and Development Institute Congress 2011: Integrated Transportation and Development for a Better Tomorrow, Chicago IL, American Society of Civil Engineers.
El-Badawy, S., F. Bayomy and A. Awed (2012). "Performance of MEPDG dynamic modulus predictive models for asphalt concrete mixtures: local calibration for Idaho." Journal of Materials in Civil Engineering **24**(11): 1412-1421.
Epps, J. A. (2002). Recommended performance-related specification for hot-mix asphalt construction: Results of the WesTrack project. Washington, DC, Transportation Research Board.
Fang, H., J. E. Haddock, T. D. White and A. J. Hand (2004). "On the characterization of flexible pavement rutting using creep model-based finite element analysis." Finite Elements in Analysis and Design **41**(1): 49-73.
Ferreira, C. (2001). "Gene expression programming: a new adaptive algorithm for solving problems." Complex Systems **13**(2): 87-129.
Ferreira, C. (2006). Gene expression programming: mathematical modeling by an artificial intelligence. Berlin, Germany, Springer-Verlag.
Fukushima, K. (1969). "Visual feature extraction by a multilayered network of analog threshold elements." IEEE Transactions on Systems Science and Cybernetics **5**(4): 322-333.
Fwa, T., S. Tan and L. Zhu (2004). "Rutting prediction of asphalt pavement layer using C-φ model." Journal of Transportation Engineering **130**(5): 675-683.

Gandomi, A. H., G. J. Yun and A. H. Alavi (2013). "An evolutionary approach for modeling of shear strength of RC deep beams." Materials and Structures **46**(12): 2109-2119.
George, S. and A. K. Santra (2020). "Traffic prediction using multifaceted techniques: a survey." Wireless Personal Communications **115**(2): 1047-1106.
GEPSOFT (2014). GeneXproTools.
Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. Sebastopol, CA, O'Reilly Media, Inc.
Ghahramani, Z. (2015). "Probabilistic machine learning and artificial intelligence." Nature **521**(7553): 452-459.
Gnanadesikan, R. and M. B. Wilk (1968). "Probability plotting methods for the analysis of data." Biometrika **55**(1): 1-17.
Gong, H., Y. Sun, W. Hu and B. Huang (2021). "Neural networks for fatigue cracking prediction using outputs from pavement mechanistic-empirical design." International Journal of Pavement Engineering **22**(2): 162-172.
Gong, H., Y. Sun, Z. Mei and B. Huang (2018). "Improving accuracy of rutting prediction for mechanistic-empirical pavement design guide with deep neural networks." Construction and Building Materials **190**: 710-718.
Grubbs, F. E. (1969). "Procedures for detecting outlying observations in samples." Technometrics **11**(1): 1-21.
Gulen, S., K. Zhu, J. Weaver, J. Shan and W. Flora (2001). Development of improved pavement performance prediction models for the Indiana pavement management system. West Lafayette, IN, Joint Transportation Research Program, Indiana Department of Transportation**:** 69.
Gupta, A., P. Kumar and R. J. K. J. o. C. E. Rastogi (2014). "Critical review of flexible pavement performance models." **18**(1): 142-148.
Haas, R. (1998). Pavement Management: A Great Past But What About the Future? Fourth International Conference on Managing Pavements, Durban, South Africa, CSIR Transportek.
Hall, C. A. and W. W. Meyer (1976). "Optimal error bounds for cubic spline interpolation." Journal of Approximation Theory **16**(2): 105-122.
Han, Z., J. Zhao, H. Leung, K. F. Ma and W. Wang (2019). "A review of deep learning models for time series prediction." IEEE Sensors Journal **21**(6): 7833-7848.
Heris, M. K. (2015). "Quadratic Assignment Problem (QAP) using GA, PSO and FA." from https://yarpiz.com/359/ypap104-quadratic-assignment-problem.
Hewamalage, H., C. Bergmeir and K. Bandara (2021). "Recurrent neural networks for time series forecasting: Current status and future directions." International Journal of Forecasting **37**(1): 388-427.
Hill, L., A. Cheetham and R. Hass (1991). "Development and implementation of a pavement management system for Minnesota." Transportation Research Record **1311**: 230-241.
Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory." Neural Computation **9**(8): 1735-1780.
Holt, C. C. (2004). "Forecasting seasonals and trends by exponentially weighted moving averages." International journal of forecasting **20**(1): 5-10.
Hong, F., E. Perrone, M. Mikhail and A. Eltahan (2017). Planning pavement maintenance and rehabilitation projects in the New Pavement Management System in Texas. Transportation Research Board 96th Annual Meeting. Washington, D.C., Transportation Research Board.
Hossain, M., L. Gopisetti and M. Miah (2017). Prediction of international roughness index of flexible pavements from climate and traffic data using artificial neural network modeling. Airfield and highway pavements 2017**:** 256-267.

Hou, S., X. Shi, Y. Deng and F. Gu (2018). "Evaluation of rutting and friction resistance of hot mix asphalt concrete using an innovative vertically loaded wheel tester." Construction and Building Materials **176**: 710-719.
Huang, B., L. N. Mohammad and M. Rasoulian (2001). "Three-dimensional numerical simulation of asphalt pavement at Louisiana accelerated loading facility." Transportation research record **1764**(1): 44-58.
Huang, Y. H. (2004). Pavement analysis and design. Upper Saddle River, NJ, Pearson Education, Inc.
Hyndman, R. J. and G. Athanasopoulos (2018). Forecasting: Principles and Practice. Melbourne, Australia, OTexts.
Iooss, B. and A. Saltelli (2017). Introduction to sensitivity analysis. Handbook of uncertainty quantification, Springer**:** 1103-1122.
ITD (2019). Transportation Asset Management Plan. Boise, ID, Idaho Transportation Department.
Ji, X., N. Zheng, Y. Hou and S. Niu (2013). "Application of asphalt mixture shear strength to evaluate pavement rutting with accelerated loading facility (ALF)." Construction and Building Materials **41**: 1-8.
Juang, C. and S. Amirkhanian (1992). "Unified pavement distress index for managing flexible pavements." Journal of Transportation Engineering **118**(5): 686-699.
Justo-Silva, R., A. Ferreira and G. Flintsch (2021). "Review on machine learning techniques for developing pavement performance prediction models." Sustainability **13**(9): 5248.
Kandhal, P. S. and L. A. Cooley (2003). Accelerated laboratory rutting tests: Evaluation of the asphalt pavement analyzer. Washington, DC, Transportation Research Board. **508**.
KEI (2015). AgileAssets Pavement Management System Engineering Configuration: Final Configuration Document. Boise, ID, Kercher Engineering, Inc.
Kenis, W. (1978). An Interim Design Method for Flexible Pavements Using the VESYS Structural Subsystem. Washington, DC.
Kennedy, J. and R. Eberhart (1995). Particle swarm optimization. International Conference on Neural Networks (ICNN'95), Perth, WA, Australia, IEEE.
Kerkhoven, R. E. and G. M. Dormon (1954). Some considerations on the California bearing ratio method for the design of flexible pavements. London, Shell Research Centre.
Khan, A. S. and S. Huang (1995). Continuum theory of plasticity. New York, NY, John Wiley & Sons.
Khedr, S. A. and M. Mikhail (1996). "Design of flexible pavements and overlay using an expert system." Transportation research record **1543**(1): 20-28.
Khraibani, H., T. Lorino, P. Lepert and J.-M. Marion (2012). "Nonlinear mixed-effects model for the evaluation and prediction of pavement deterioration." Journal of Transportation Engineering **138**(2): 149-156.
Kim, S.-H. and N. Kim (2006). "Development of performance prediction models in flexible pavement using regression analysis method." KSCE Journal of Civil Engineering **10**(2): 91-96.
Kim, W. J., H. J. Lee and H. T. Phan (2017). "Calibration and validation of a rutting model based on shear stress to strength ratio for asphalt pavements." Construction and Building Materials **149**: 327-337.
Kingma, D. P. and J. Ba (2014). "Adam: A method for stochastic optimization." arXiv preprint.
Kiranyaz, S., O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj and D. J. Inman (2021). "1D convolutional neural networks and applications: A survey." Mechanical Systems and Signal Processing **151**: 107398.
Korkiala-Tanttu, L. and A. Dawson (2007). "Relating full-scale pavement rutting to laboratory permanent deformation testing." International Journal of Pavement Engineering **8**(1): 19-28.
Kotu, V. and B. Deshpande (2018). Data Science: Concepts and Practice. Burlington, MA, Morgan Kaufmann.
Lea, J. and J. T. Harvey (2002). Data Mining of the Caltrans Pavement Management System (PMS) Database. Richmond, CA, Pavement Research Center, University of California, Berkeley.

LeCun, Y. and Y. Bengio (1995). Convolutional networks for images, speech, and time series. The Handbook of Brain Theory and Neural Networks. Cambridge, MA, MIT Press.
Lee, J. Y., F. Dernoncourt and P. Szolovits (2017). "Transfer learning for named-entity recognition with neural networks." arXiv preprint arXiv:1705.06273.
Li, J., L. M. Pierce and J. Uhlmeyer (2009). "Calibration of flexible pavement in mechanistic–empirical pavement design guide for Washington state." Transportation Research Record **2095**(1): 73-83.
Li, L., X. Huang, D. Han, M. Dong and D. Zhu (2015). "Investigation of rutting behavior of asphalt pavement in long and steep section of mountainous highway with overloading." Construction and Building Materials **93**: 635-643.
Li, Y., A. Cheetham, S. Zaghloul, K. Helali and W. Bekheet (2006). "Enhancement of Arizona pavement management system for construction and maintenance activities." Transportation Research Record **1974**(1): 26-36.
Ling, M., Y. Chen, S. Hu, X. Luo and R. L. Lytton (2019). "Enhanced model for thermally induced transverse cracking of asphalt pavements." Construction and Building Materials **206**: 130-139.
Lu, Q. and J. T. Harvey (2006). "Evaluation of Hamburg wheel-tracking device test with laboratory and field performance data." Transportation research record **1970**(1): 25-44.
Luo, X., F. Gu, Y. Zhang, R. L. Lytton and B. Birgisson (2018). "Kinetics-based aging evaluation of in-service recycled asphalt pavement." Journal of Cleaner Production **200**: 934-944.
Luo, X., H. Li, Y. Deng and Y. Zhang (2020). "Energy-Based Kinetics Approach for Coupled Viscoplasticity and Viscofracture of Asphalt Mixtures." Journal of Engineering Mechanics **146**(9): 04020100.
Luo, X., R. Luo and R. L. Lytton (2013). "Characterization of fatigue damage in asphalt mixtures using pseudostrain energy." Journal of Materials in Civil Engineering **25**(2): 208-218.
Luo, X., H. Wang, Y. Deng and H. Li (2022). "Automatic inverse analysis of asphalt pavement field aging based on system identification." IEEE Transactions on Intelligent Transportation Systems **23**(11): 22106-22115.
Lytton, R. L. (2000). "Characterizing asphalt pavements for performance." Transportation Research Record **1723**(1): 5-16.
Lytton, R. L., X. Luo, M. Ling, Y. Chen, S. Hu and F. Gu (2018). A Mechanistic–Empirical Model for Top–Down Cracking of Asphalt Pavements Layers. Washington, DC, Texas A&M Transportation Institute
Lytton, R. L., X. Luo, S. Saha, Y. Chen, Y. Deng, F. Gu and M. Ling (2019). Proposed Enhancements to Pavement ME Design: Improved Consideration of the Influence of Subgrade and Unbound Layers on Pavement Performance. College Station, TX, Texas A&M Transportation Institute.
Ma, T., D. Zhang, Y. Zhang and J. Hong (2016). "Micromechanical response of aggregate skeleton within asphalt mixture based on virtual simulation of wheel tracking test." Construction and Building Materials **111**: 153-163.
Ma, T., D. Zhang, Y. Zhang, Y. Zhao and X. Huang (2016). "Effect of air voids on the high-temperature creep behavior of asphalt mixture based on three-dimensional discrete element modeling." Materials & Design **89**: 304-313.
Makridakis, S., E. Spiliotis and V. Assimakopoulos (2018). "Statistical and machine learning forecasting methods: concerns and ways forward." PLOS ONE **13**(3): e0194889.
Mandapaka, V., I. Basheer, K. Sahasi, P. Ullidtz, J. T. Harvey and N. Sivaneswaran (2012). "Mechanistic-empirical and life-cycle cost analysis for optimizing flexible pavement maintenance and rehabilitation." Journal of Transportation Engineering **138**(5): 625-633.
Marcelino, P., M. de Lurdes Antunes, E. Fortunato and M. C. Gomes (2021). "Machine learning approach for pavement performance prediction." International Journal of Pavement Engineering **22**(3): 341-354.

Mazari, M. and D. D. Rodriguez (2016). "Prediction of pavement roughness using a hybrid gene expression programming-neural network technique." Journal of Traffic and Transportation Engineering **3**(5): 448-455.
Mitchell, T. M., J. G. Carbonell and R. S. Michalski (2012). Machine Learning: A Guide to Current Research, Springer US.
Morris, M. D. (1991). "Factorial sampling plans for preliminary computational experiments." Technometrics **33**(2): 161-174.
Murphy, K. P. (2012). Machine learning: a probabilistic perspective. Cambridge, MA, MIT press.
Ningyuan, L., T. Kazmierowski, S. Tighe and R. Haas (2001). Integrating dynamic performance prediction models into pavement management maintenance and rehabilitation programs. 5th International Conference on Managing Pavements, Seattle, WA, Engineering Professional Programs, University of Washington.
Nisbet, R., J. Elder and G. Miner (2009). Handbook of statistical analysis and data mining applications. Cambridge, MA, Academic press.
Okuda, T., K. Suzuki and N. Kohtake (2018). Non-parametric Prediction Interval Estimate for Uncertainty Quantification of the Prediction of Road Pavement Deterioration. 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE.
Park, H. M. and Y. R. Kim (2003). "Prediction of remaining life of asphalt pavement with falling-weight deflectometer multiload-level deflections." Transportation Research Record **1860**(1): 48-56.
Peng, B. (2014). Discrete element method (DEM) contact models applied to pavement simulation. PhD, Virginia Tech.
Prozzi, J. A. (2001). Modeling pavement performance by combining field and experimental data. Ph.D., University of California, Berkeley.
Prozzi, J. A. and S. Madanat (2002). A non-linear model for predicting pavement serviceability. Berkeley, CA, UC Berkeley: University of California Transportation Center.
Prozzi, J. A. and S. M. Madanat (2003). "Incremental nonlinear model for predicting pavement serviceability." Journal of transportation Engineering **129**(6): 635-641.
Pu, Z., C. Liu, X. Shi, Z. Cui and Y. Wang (2020). "Road surface friction prediction using long short-term memory neural network based on historical data." Journal of Intelligent Transportation Systems: 1-12.
Queiroz, C. (1983). "A mechanistic analysis of asphalt pavement performance in Brazil." Journal of Association of Asphalt Paving Technology **52**: 474-488.
Ricardo Archilla, A. and S. Madanat (2001). "Statistical model of pavement rutting in asphalt concrete mixes." Transportation Research Record **1764**(1): 70-77.
Ringwood, J. V., A. Mérigaud, N. Faedo and F. Fusco (2019). "An analytical and numerical sensitivity and robustness analysis of wave energy control systems." IEEE Transactions on Control Systems Technology **28**(4): 1337-1348.
Rogoza, W. (2019). "Method for the prediction of time series using small sets of experimental samples." Applied Mathematics and Computation **355**: 108-122.
Romero, P. and K. Stuart (1998). "Evaluating accelerated rut testers." Public Roads **62**(1).
Saha, S., F. Gu, X. Luo and R. L. Lytton (2018). "Use of an artificial neural network approach for the prediction of resilient modulus for unbound granular material." Transportation Research Record **2672**(52): 23-33.
Saidi, R., W. Bouaguel and N. Essoussi (2019). Hybrid feature selection method based on the genetic algorithm and pearson correlation coefficient. Machine Learning Paradigms: Theory and Application. A. E. Hassanien. Cham Switzerland, Springer. **801:** 3-24.
Schapery, R. A. (1997). "Nonlinear viscoelastic and viscoplastic constitutive equations based on thermodynamics." Mechanics of Time-Dependent Materials **1**(2): 209-240.

Schmidhuber, J. (2015). "Deep learning in neural networks: An overview." Neural Networks **61**: 85-117.
Seitllari, A., M. A. Lanotte and M. E. Kutay (2019). Calibration of the MEPDG rutting model: issues and consequences on rutting prediction. Transportation Research Board 98th Annual Meeting. Washington, D.C., Transportation Research Board.
Shabbir, F. and P. Omenzetter (2015). "Particle swarm optimization with sequential niche technique for dynamic finite element model updating." Computer‐Aided Civil and Infrastructure Engineering **30**(5): 359-375.
Shami, H. I., J. S. Lai, J. A. D'angelo and T. P. Harman (1997). "Development of temperature-effect model for predicting rutting of asphalt mixtures using Georgia loaded wheel tester." Transportation Research Record **1590**(1): 17-22.
Shapiro, S. S. and M. B. Wilk (1965). "An analysis of variance test for normality (complete samples)." Biometrika **52**(3/4): 591-611.
Shell International Petroleum Company, L. (1978). Shell pavement design manual: Asphalt pavements and overlays for road traffic. London, England, Shell International Petroleum Company.
Shenoy, A. and P. Romero (2001). "Superpave shear tester as a simple standardized measure to evaluate aggregate-asphalt mixture performance." Journal of testing and evaluation **29**(5): 472-484.
Shi, X., Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong and W.-c. Woo (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. NIPS2015. C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett. Montreal, Canada, Curran Associates, Inc. **28**.
Shi, X., P. Schillings and D. Boyd (2004). "Applying artificial neural networks and virtual experimental design to quality improvement of two industrial processes." International Journal of Production Research **42**(1): 101-118.
Shojaeefard, M. H., M. Akbari, M. Tahani and F. Farhani (2013). "Sensitivity analysis of the artificial neural network outputs in friction stir lap joining of aluminum to brass." Advances in Materials Science and Engineering **2013**.
Shoukry, S. N. and G. W. William (1999). "Performance evaluation of backcalculation algorithms through three-dimensional finite-element modeling of pavement structures." Transportation Research Record **1655**(1): 152-160.
Stathakis, D. (2009). "How many hidden layers and nodes?" International Journal of Remote Sensing **30**(8): 2133-2147.
Stephens, J. (1990). The effects of temperature cycles on the aging of bituminous concrete. Storrs, CT, University of Connecticut, Storrs**:** 49.
Su, D., H. Zhang, H. Chen, J. Yi, P.-Y. Chen and Y. Gao (2018). Is Robustness the Cost of Accuracy?--A Comprehensive Study on the Robustness of 18 Deep Image Classification Models. Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany.
Sullivan, B. W. (2002). Development of flow number and flow time candidate simple performance test for asphalt mixtures. Master of Science, Arizona State University.
Tan, S.-A., B.-H. Low and T. Fwa (1994). "Behavior of asphalt concrete mixtures in triaxial compression." Journal of Testing and Evaluation **22**(3): 195-203.
Timm, D., B. Birgisson and D. Newcomb (1998). "Development of mechanistic-empirical pavement design in Minnesota." Transportation Research Record **1629**(1): 181-188.
Tseng, K.-H. and R. L. Lytton (1989). Prediction of permanent deformation in flexible pavement materials. Implication of aggregates in the design, construction, and performance of flexible pavements. H. Schreuders and C. Marek. West Conshohocken, PA, ASTM International**:** 154-172.
Turtschy, J. C. and G. Sweere (1999). Long Term Pavement Performance Modeling Based on Accelerated Pavement Testing. The 1st International Conference on Accelerated Pavement Testing, Reno, NV.

Tutumluer, E. and T. Pan (2008). "Aggregate morphology affecting strength and permanent deformation behavior of unbound aggregate materials." Journal of Materials in Civil Engineering **20**(9): 617-627.
Ullidtz, P. (1999). "Deterioration models for managing flexible pavements." Transportation Research Record **1655**(1): 31-34.
Wang, F., C.-C. B. Lee and N. G. Gharaibeh (2022). "Network-level bridge deterioration prediction models that consider the effect of maintenance and rehabilitation." Journal of Infrastructure Systems **28**(1): 05021009.
Wang, S.-C. (2003). Artificial neural network. Interdisciplinary Computing in Java Programming. Boston, MA, Springer. **743:** 81-100.
Welch, B. L. (1947). "The generalization of 'STUDENT'S' problem when several different population varlances are involved." Biometrika **34**(1-2): 28-35.
Winters, P. R. (1960). "Forecasting sales by exponentially weighted moving averages." Management Science **6**(3): 324-342.
Witczak, M. W. (2007). Specification Criteria for Simple Performance Tests for Rutting, Volume I: Dynamic Modulus (E*) and Volume II: Flow Number and Flow Time. Washington, DC, Arizona State University**:** 94.
Wold, S., K. Esbensen and P. Geladi (1987). "Principal component analysis." Chemometrics and Intelligent Laboratory Systems **2**(1-3): 37-52.
Wu, Z., X. Chen, X. Yang and Z. Zhang (2011). "Finite element model for rutting prediction of flexible pavement with cementitiously stabilized base–subbase." Transportation research record **2226**(1): 104-110.
Xin, J., M. Akiyama, D. M. Frangopol and M. Zhang (2022). "Multi-objective optimisation of in-service asphalt pavement maintenance schedule considering system reliability estimated via LSTM neural networks." Structure and Infrastructure Engineering: 1-18.
Xue, Q. and L. Liu (2013). "Hydraulic-stress coupling effects on dynamic behavior of asphalt pavement structure material." Construction and Building Materials **43**: 31-36.
Xue, Q., L. Liu, Y. Zhao, Y.-J. Chen and J.-S. Li (2013). "Dynamic behavior of asphalt pavement structure under temperature-stress coupled loading." Applied Thermal Engineering **53**(1): 1-7.
Yang, J., J. J. Lu, M. Gunaratne and B. Dietrich (2006). "Modeling crack deterioration of flexible pavements: Comparison of recurrent Markov chains and artificial neural networks." Transportation Research Record **1974**(1): 18-25.
Yao, L., Z. Leng, J. Jiang, F. Ni and Z. Zhao (2021). "Nondestructive prediction of rutting resistance of in-service middle asphalt layer based on gene expression programing." Construction and Building Materials **293**: 123481.
Ye, Z., X. Shi, C. K. Strong and T. H. Greenfield (2009). "Evaluation of effects of weather information on winter maintenance costs." Transportation Research Record **2107**(1): 104-110.
Yin, F., E. Arambula, R. Lytton, A. E. Martin and L. G. Cucalon (2014). "Novel method for moisture susceptibility and rutting evaluation using Hamburg wheel tracking test." Transportation Research Record **2446**(1): 1-7.
Yu, H. and B. M. Wilamowski (2018). Levenberg–marquardt training. Intelligent Systems. B. M. Wilamowski and J. D. Irwin. Boca Raton, FL, CRC Press**:** 12-11-12-16.
Yu, L., K. K. Lai, S. Wang and K. He (2007). Oil price forecasting with an EMD-based multiscale neural network learning paradigm. International Conference on Computational Science, Beijing China, Springer.
Yuan, J. and M. A. Mooney (2003). "Development of adaptive performance models for Oklahoma Airfield pavement management system." Transportation Research Record **1853**(1): 44-54.

Zhang, J., J. Pei and Z. Zhang (2012). "Development and validation of viscoelastic-damage model for three-phase permanent deformation of dense asphalt mixture." Journal of materials in civil engineering **24**(7): 842-850.
Zhang, Y. (2012). Anisotropic characterization of asphalt mixtures in compression. Ph.D., Texas A&M University.
Zhang, Y., H. Chen, P. Xiao, Y. Deng and A.-H. Kang (2022). "Investigation of average asphalt film thickness of dense graded asphalt mixtures with compaction effects." Construction and Building Materials **326**: 126696.
Zhang, Y., F. Gu, B. Birgisson and R. L. Lytton (2017). "Viscoelasticplastic–Fracture Modeling of Asphalt Mixtures under Monotonic and Repeated Loads." Transportation Research Record **2631**(1): 20-29.
Zhang, Y., R. Luo and R. L. Lytton (2013). "Mechanistic modeling of fracture in asphalt mixtures under compressive loading." Journal of Materials in Civil Engineering **25**(9): 1189-1197.
Zhang, Y., T. Ma, X. Ding, T. Chen, X. Huang and G. Xu (2018). "Impacts of air-void structures on the rutting tests of asphalt concrete based on discretized emulation." Construction and Building Materials **166**: 334-344.
Zhou, F., S. Hu and T. Scullion (2006). Integrated asphalt (overlay) mixture design, balancing rutting and cracking requirements. College Station, TX, Texas Transportation Institute, Texas A&M University System.
Zhou, F. and T. Scullion (2002). VESYS5 Rutting model calibrations with local accelerated pavement test data and associated implementation. College Station, TX, Texas Transportation Institute, Texas A&M University System.
Zhou, F., T. Scullion and L. Sun (2004). "Verification and modeling of three-stage permanent deformation behavior of asphalt mixes." Journal of Transportation Engineering **130**(4): 486-494.
Zienkiewicz, O. C., C. Humpheson and R. Lewis (1975). "Associated and non-associated visco-plasticity and plasticity in soil mechanics." Geotechnique **25**(4): 671-689.

Appendix A. Complete Version of the Practitioner Survey

This survey, as part of a research project entitled "Developing Enhanced Performance Curves of ITD Asphalt Pavements by Mining the Historical Data", is designed to gather information about pavement performance management from managers and engineers in roadway agencies. The project is funded by the Idaho Transportation Department (ITD) and the National Center for Transportation Infrastructure Durability & Life-Extension (TriDurLE). The survey aims to capture your insights and experiences (as users) on the existing performance deterioration models for asphalt pavements, in terms of model performance, input and output parameters, consideration of maintenance and rehabilitation (M&R) history, implementation considerations, etc. Thank you for supporting research and rest assured that your feedback will be anonymized or aggregated to protect your privacy.

Please provide your Name, Title, Agency and Email.

- Name ____
- Title ____
- Agency ____
- Email ____

1. Which distress modes of asphalt pavements are you primarily interested in your work? (select all that apply)

    a. rutting

    b. transverse cracking

    c. longitudinal cracking

    d. alligator cracking

    e. edge cracking

    f. block cracking

    g. pothole

    h. patch

    i. raveling

    j. roughness

    k. others (please specify) ____

2. Which resources do you have to obtain/develop models of distress modes, and how many years has your agency used it? (select all that apply and provide information about the resources you use as needed)

    a. AASHTOWare Pavement ME Design
    ____

    b. software developed or purchased by your agency (please specify)
    ____

    c. EXCEL sheets with programmed models (please specify the references)
    ____

    d. other (please specify) ____

3. What limitations do you think the models you use have? (select all that apply)

    a. too many/too complicated inputs

    b. poor prediction

    c. unfriendly user interface

    d. other (please specify) ____

4. Which inputs are required in the models you use? (select all that apply)

    a. traffic data

    b. climatic data

    c. material properties data

    d. pavement structure data

    e. other (please specify) ____

5. Which inputs are difficult to obtain when you use the models? (select all that apply)

    a. traffic data

    b. climatic data

    c. material properties data

    d. pavement structure data

    e. other (please specify) ____

6. Please specify the specific factors of Traffic which are difficult to obtain as model input (select all that apply)
    a. traffic volume
    b. load magnitude & distribution
    c. load speed
    d. load spectrum
    e. other (please specify) ____
7. Please specify the specific factors of Climate which are difficult to obtain as model input (select all that apply)
    a. temperature
    b. precipitation
    c. wind speed
    d. relative humidity
    e. other (please specify) ____
8. Please specify the specific factors of Pavement Material which are difficult to obtain as model input (select all that apply)
    a. asphalt mixture properties
    b. asphalt binder properties
    c. base and subgrade properties
    d. other (please specify) ____
9. Please specify the specific factors of Pavement Structure which are difficult to obtain as model input (select all that apply)
    a. layer thickness
    b. overlays information
    c. construction information
    d. other (please specify) ____
10. For the inputs that are not available, what approaches do you usually take in order to use the models? (select all that apply)

a. use default or recommended values in the software/EXCEL sheet

b. use values recorded in the database (LTPP, NSRDB, etc.) for your pavement section (please specify the names of databases)
____

c. use typical values in the references (papers, reports, slides, etc.)

d. other (please specify) ____

11. What are your purposes of using the models? (select all that apply)

    a. to obtain distress indices for pavement management (PMS, PMIS, etc.)

    b. to obtain distress indices for innovative materials/structural designs

    c. to validate the models you use

    d. to calibrate the models you use

    e. other (please specify) ____

12. Do you use historical data to validate or refine your asphalt pavement performance models?

    a. No

    b. Yes

13. After repairs, the structural and material conditions of your pavement may be different. How do you consider such inconsistency when you use models? In other words, do your pavement performance models consider the M&R (maintenance and rehabilitation) history of the asphalt pavement? Please comment:

    ____

14. How do you rate the models you use if you have compared model predictions with field measurements?

    a. 7-10: most of them are accurate

    b. 4-6: some of them are accurate

    c. 0-3: few of them are accurate

    d. I didn’t do such comparisons

15. Have you ever faced any of the following problems when using your asphalt pavement performance models? (select all that apply) Please specify the model name and comment.

a. you can get very different predictions in multiple runs of your model ____

b. you can get very different predictions in pavements with similar conditions ____

c. your model is too sensitive or insensitive to certain model inputs ____

d. other (please specify) ____

16. For a pavement performance model, what values of R-squared (R^2) do you think are tolerable, if fitting the predicted indices vs. actual performance data?

    a. above 0.60

    b. above 0.80

    c. above 0.90

    d. above 0.95

17. Which distresses do the models you use gave better prediction? (up to three selections)

    a. rutting

    b. transverse cracking

    c. longitudinal cracking

    d. alligator cracking

    e. edge cracking

    f. block cracking

    g. pothole

    h. patch

    i. raveling

    j. roughness

    k. other (please specify) ____

18. Which distresses do the models you use gave poorer prediction? (up to three selections)

    a. rutting

b. transverse cracking

c. longitudinal cracking

d. alligator cracking

e. edge cracking

f. block cracking

g. pothole

h. patch

i. raveling

j. roughness

k. other (please specify) ____

19. In your opinion, which factors caused the poor prediction? (select all that apply)

    a. oversimplification of model inputs (please specify which inputs, e.g., traffic speed) ____

    b. inappropriate model forms (please specify which distress modes and model types, e.g., linear) ____

    c. insufficient or inappropriate data for model calibration

    d. other (please specify) ____

20. How often does your agency check the prediction accuracy and update the database of the pavement performance models?

    a. 0-2 years

    b. 3-5 years

    c. over 5 years

    d. never

21. What strategies have your agency adopted to improve the accuracy of the pavement performance models? (select all that apply)

    a. do the local calibration using the data in your state

    b. add or reduce the number of model parameters based on the conditions of your state

c. other (please specify) ____

22. Please rank the following factors according to your priorities (first is the most important) when considering pavement performance models. (you can drag these factors)

    ______ accurate prediction

    ______ ease of use

    ______ reliability and ruggedness (e.g., get consistent results in multiple runs)

    ______ clear form

    ______ solid foundation on mechanistic theory (vs. empirical model)

    ______ other (please specify the name)

23. Please provide your opinion on HOW current performance deterioration models (for asphalt pavements) can be improved.

    ____

24. How much are artificial intelligence (AI) models (e.g., neural networks) involved in your work?

    a. I use such models in my work (please specify the model names if possible) ____

    b. I know and am willing to use such models but haven't found a chance yet

    c. I know such models but don't think they are appropriate to be used in my work

    d. I don't know such models

25. If now you have a chance to try the artificial intelligence models as the performance deterioration models for asphalt pavements.

    a. I would use the artificial intelligence models to replace the ones I currently use

    b. I would try both the artificial intelligence models and the models I currently use, and pick the better one

    c. I would rather use and improve traditional empirical/mechanistic-empirical models

    d. other (please specify) ____

Other relevant information, comments or suggestions you would like to provide:

____

Appendix B. Supplement Information of GEP Models

The contents included in this appendix are listed as follows.

- Variable normalization function
- Maximum and minimum values of model inputs and outputs for variable normalization
- Equations of final models

$$X = \left(x - x_{\min}\right)\frac{2}{x_{\max} - x_{\min}} - 1$$

Where:

x, X = original and normalized values
x_{max}, x_{min} = the maximum and minimum values of the variable

Table B.1 Maximum and Minimum Values of Model Inputs and Outputs

Distress	Rutting	Rutting	Longitudinal Cracking	Longitudinal Cracking	Thermal Cracking	Thermal Cracking	IRI	IRI
-	Max	Min	Max	Min	Max	Min	Max	Min
AADTT, 1	10426	308	10426	707	10426	493	10426	493
Hac, mm	304.00	60.96	228.60	60.96	228.60	60.96	304.00	60.96
Hb, mm	777.24	152.40	777.24	170.69	765.05	170.69	777.24	152.40
BC, %	6.04	4.40	6.04	4.86	6.04	4.40	6.04	4.40
Gb, 1	1.035	1.020	1.035	1.023	1.035	1.023	1.035	1.020
Gsb, 1	2.941	2.554	2.941	2.561	2.941	2.554	2.941	2.554
EVAPOR, mm	8499.6	530.0	8508.3	811.1	8508.3	811.1	8508.3	811.1
FI, °C	4805	215	5277	433	5277	433	4805	433
FT, day	1734	112	1734	231	1734	251	1734	230
AVE_SW_SUR, W/s2	151404	119798	151014	119798	151014	119798	150968	119798
Output*	9.418	1.016	69.419	0.104	164.557	0.896	3074.016	891.728

* Units of outputs of rutting, longitudinal cracking, thermal cracking and IRI models are mm, m/km, m/km and mm/km, respectively

Table B.2 Equation of Final Models of Four Distresses

Distress	Equation	Coefficients
Rutting	y = (d(2)+(((((G1C5*G1C4)+(d(9)/d(1)))+exp(exp(d(8))))+(d(5)+(G1C4-d(6))))*((d(9)+(d(2)*G1C3))/((G1C8^2)-d(6)))))+(d(7)+(1.0/(((((d(5)+d(2))+((G2C6*G2C2)-G2C1))-((G2C9-G2C2)*(d(1)*d(7))))-(((d(10)/d(6))-(d(1)*G2C6))-(d(4)^2))))))+(d(10)*((d(7)*(exp(d(3))-(d(5)/((((G3C2*G3C0)*d(4))/exp(d(6)))-(G3C5+(d(3)-d(8)))))))*d(8)))+((d(2)-((d(6)*(((G4C6/G4C0)-d(3))+((d(10)+d(4))-d(7))))/(((realsqrt(G4C8)-(G4C7*d(6)))+(G4C1*G4C2))^2)))*d(6))	G1C4 = -5.72; G1C3 = 0.32; G1C8 = -3.69; G1C5 = 1.13; G2C1 = 8.18; G2C9 = -4.14; G2C2 = 0.69; G2C6 = -7.93; G3C5 = -1.83; G3C2 = 4.49; G3C0 = -2.45; G4C1 = 6.56; G4C2 = -1.27; G4C6 = -8.04; G4C0 = -6.65; G4C8 = 5.09; G4C7 = -6.89
Longitudinal Cracking	y = ((((d(8)+d(10))-(d(10)*(d(7)+d(3))))-((d(5)*d(10))+d(1)))*(d(9)/(((d(9)+G1C5)+d(6))+(d(5)^2))))+((d(1)+d(7))+((d(6)-(d(6)+realsqrt((((d(10)*exp(d(2)))*(d(10)+d(10)))*(G2C4*G2C4)))))-d(8)))+((d(10)-((1.0/((((((d(3)-G3C8)-d(6))/(1.0/(d(3))))+((d(1)-G3C9)-(d(1)*d(1))))-(1.0/(((G3C3*d(1))-d(1)))))))^2))/(1.0/(d(7))))+(((exp(d(1))+d(10))*((d(5)-((exp(d(10))/(G4C8*d(5)))*((G4C2*d(5))^2)))/(((d(8)*d(8))+(d(9)+G4C2))+G4C3)))^2)	G1C5 = -2.28; G2C4 = -0.29; G3C9 = 4.03; G3C3 = -4.27; G3C8 = 2.62;

		G4C3 = 2.66; G4C2 = 3.98; G4C8 = -5.97
Thermal Cracking	y = (((d(4)*(d(3)*(d(5)*((((d(10)+d(1))/(d(6)+d(6)))+d(3))-(((G1C1+d(8))+(d(6)/G1C7))^2)))))+G1C4)+d(2))+exp((d(10)-((((1.0/(G2C1))+((d(6)-(d(5)-d(1)))+((1.0/(d(5)))*d(9))))*exp(((G2C9-G2C2)-(d(10)+G2C4))))/d(10))))+((d(4)+(((G3C4-(((d(5)/d(2))*(d(5)-G3C1))*(G3C8*d(8))))/(((d(3)/d(1))^2)-((G3C9*G3C5)-(d(9)+d(5)))))+d(8)))*d(2))+(1.0/((1.0/(((d(10)+(((((d(1)+d(1))*(d(7)*d(10)))*((d(7)*d(7))*d(6)))/(((G4C3+G4C9)+(G4C5/d(2)))+(G4C0-d(10))))^2))*d(1))))))	G1C4 = -1.99; G1C1 = 0.95; G1C7 = 0.95; G2C1 = 0.69; G2C9 = 3.62; G2C2 = 2.87; G2C4 = 4.15; G3C4 = -0.45; G3C8 = 0.21; G3C9 = 0.54; G3C5 = -3.93; G3C1 = 3.64; G4C0 = -0.98; G4C3 = 8.13; G4C9 = -4.91; G4C5 = 0.11
IRI	y = (1.0/((((((d(6)/((G1C3*d(5))+d(3)))*((G1C6+d(5))-d(6)))*(exp(G1C9)^2))-(G1C0*exp(d(4))))^2)))+(exp((((((d(2)^2)-d(6))-(d(6)-d(1)))+G2C0)-(((G2C8-d(10))-(d(5)-G2C8))+((d(2)*d(1))/d(3)))))-(d(9)-d(10)))+((((d(10)*(d(7)*G3C4))+(1.0/(G3C8)))*d(10))-exp(((((d(8)+d(2))+d(7))+G3C5)+(G3C0-(G3C7+d(9))))))+(d(7)-((((((d(5)-d(10))*(d(7)/d(1)))-((G4C0*d(10))-(d(8)+G4C8)))-(((d(3)-d(8))-d(9))-d(5)))*((d(3)/(d(2)-G4C3))^2))^2))	G1C0 = 9.08; G1C9 = -0.58; G1C6 = -8.30; G1C3 = -0.81;

		G2C0 = 6.74; G2C8 = 5.54; G3C8 = -1.20; G3C5 = 1.70; G3C0 = -4.29; G3C4 = 0.92; G3C7 = -1.40; G4C3 = -3.72; G4C0 = 5.75; G4C8 = -2.44

*d(1)~d(10) represent ten input variables in Table B.1

Appendix C. Comparison of Different Predictive Models for OCI of Asphalt Pavements

The contents included the comparison of two modelling methods – piecewise linear regression curves adopted by ITD currently and DL models applied in this project.

The first method is using piecewise linear functions to predict the development of Overall Condition Index (OCI) in asphalt pavements with different treatments. OCI is an index reflecting the general health condition of the pavement section which was defined from the combination of individual distresses such as cracking and raveling. Figure C.1 shows the four development curves of OCI with the duration (in year) to the treatment. The first step of using this method is to select the curve according to the most recent treatment to the pavement section. Next, shift the curve to the location of the latest record of OCI as shown in Figure C.2. Finally, predict the OCI value at the future time of interest using the shifted development curve.

The four models (curves) of OCI development in Figure C.1 were constructed by ITD using historical records of OCI in asphalt pavement sections with different treatment types. They can be categorized as empirical models introduced in Chapter 2. This method assumes that the variation in the OCI development in different pavement sections entirely results from the treatment type. Effects of the variations in the material properties, structure configuration, traffic and environmental conditions as described in Chapter 5 are ignored. Although the causing factors to the variation in OCI development are simplified, the advantages of this method are obvious, including simplicity, ease of implementation and application, etc.

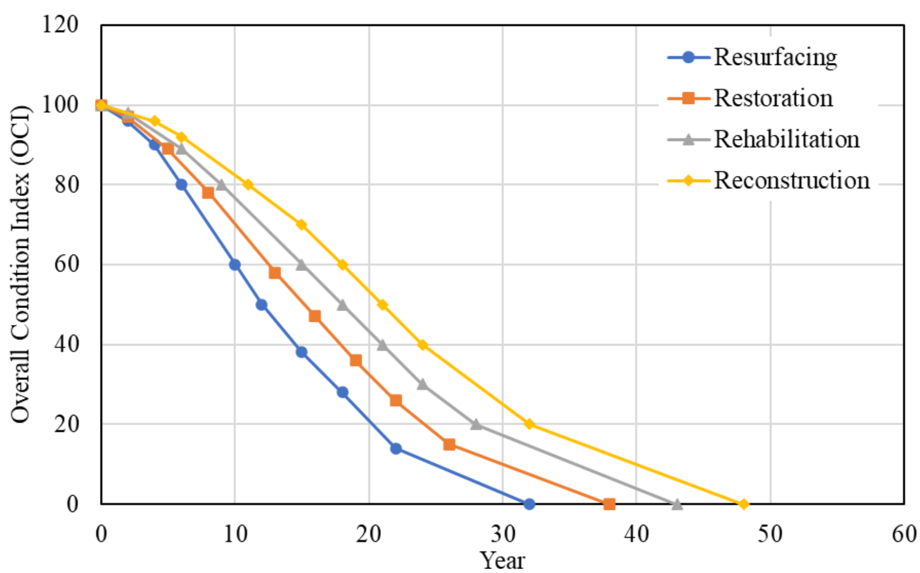


Figure C.1 Four Piecewise Linear Regression Curves of OCI Development in Asphalt Pavements

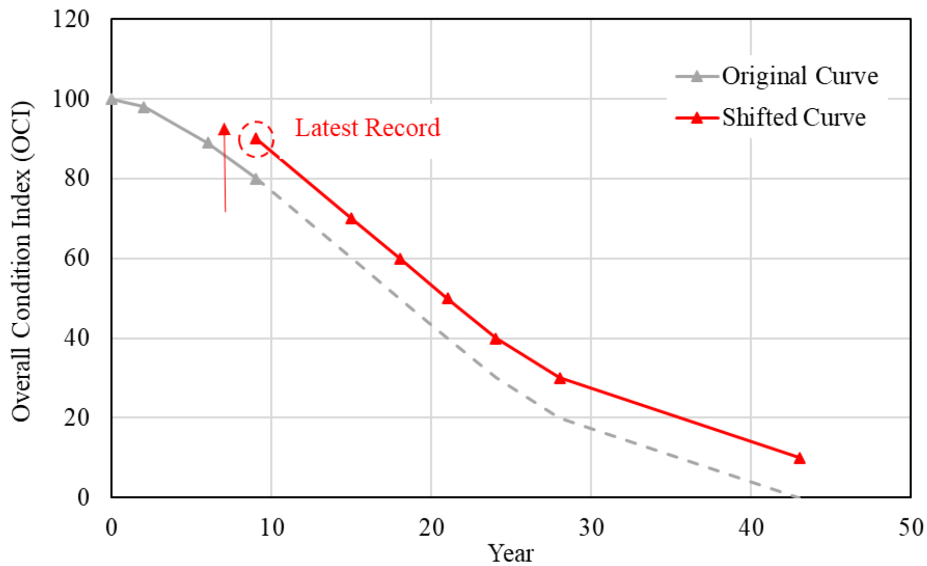


Figure C.2 Curve Shift for Predicting OCI in Different Asphalt Pavements

The second method is using DL models introduced in Chapter 5 of this final report, which are able to predict pavement performance in terms of individual distresses such as rutting as well as comprehensive condition indices such as OCI in this appendix. Compared with the first method, no specific functions should be pre-defined. The pattern of OCI development can be captured from historical records by complicated model structures as CNN and LSTM. However, DL models typically require more historical records (i.e., series with more than three datapoints) than the first method for the model training. Besides, it has been proved in Chapter 5 that the more historical data the model includes, the more precise the prediction will be.

In this appendix, OCI was utilized in the comparison between piecewise linear regression model and CNN model to show the potential application of DL models in predicting pavement performance. Historical records of OCI between 2013-2020 in four routes (Route ID: IN 015 A M, IN 084 A M, US 012 A M and US 020 A M) in the state of Idaho were collected and organized. Finally, series with three different lengths as defined in Table 5.1 (i.e., $\mathbf{N_4}$, $\mathbf{N_5}$ and $\mathbf{N_6}$) were obtained to be utilized in the model comparison. The (sample) numbers of three series are 568, 143 and 21, respectively. For CNN models, hyperparameter values were directly taken as in Table 5.2(a) which led to the highest accuracy for the rutting development. Therefore, results in this appendix can be also seen as the validation of constructed CNN models in Chapter 5 to new data.

Two indicators – absolute relative difference and RMSE were utilized to show model accuracy as in the following equations. As in Chapter 5, the last datapoint of each series was treated as the validation data, which is *y* in the following equations.

$$\text{Absolute Relative Difference} = \frac{|\hat{y} - y|}{y} \times 100\%$$

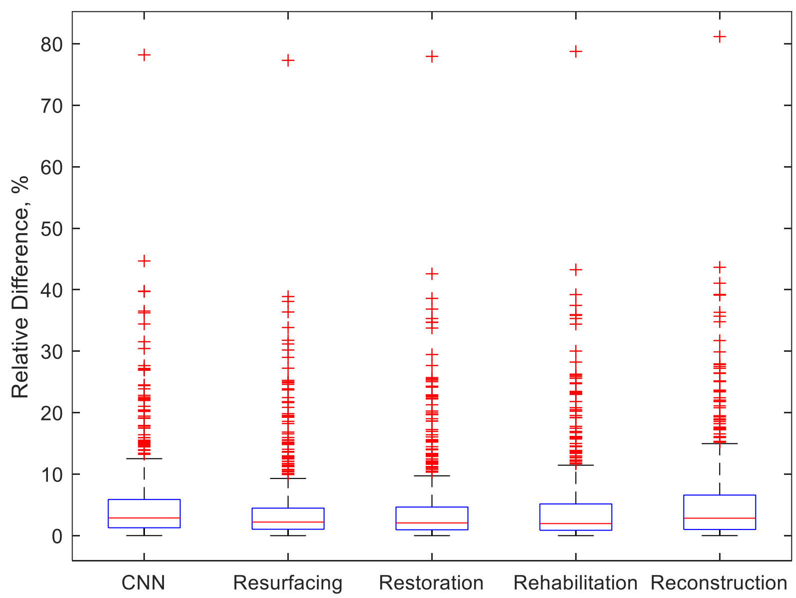
$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(\hat{y} - y\right)^2}{N}}$$

Where:

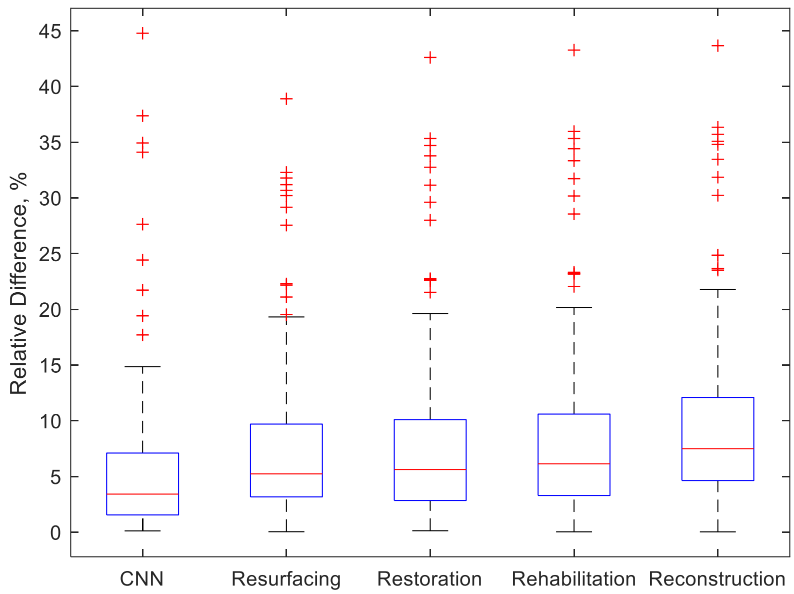
y, $\hat{y}$ = actual and predicted OCI values

N = the total number of predictions

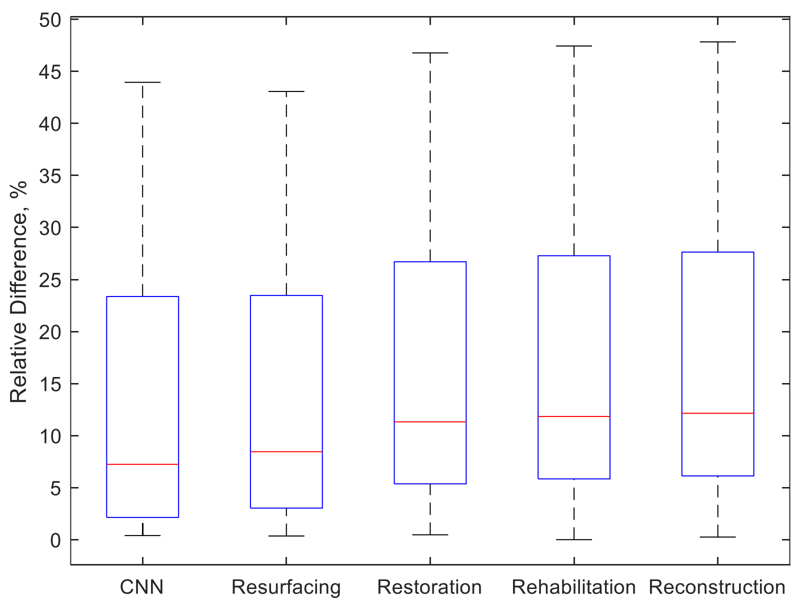
Figure C.3 shows the boxplots of absolute relative difference of predictions in three series types. Descriptions of boxplot can be found in Figure 3.7(b) of this final report. It can be seen from the plots that for the OCI development with very limited historical records (i.e., three years of available data), piecewise regression models have slightly higher prediction accuracy, indicated by the medium value of absolute relative difference in Figure C.3(a). But with the increase of series length, CNN models have higher prediction accuracy than any of four piecewise regression models, as shown in Figures C.3(b) and C.3(c). Besides, CNN models and piecewise linear regression models have similar prediction levels for the outliers, which cannot be captured by neither the trend of OCI development in previous years (captured by CNN models) nor the pattern of OCI development in other pavement sections (captured by piecewise regression models).



(a)



(b)



(c)

Figure C.3 Boxplot of Absolute Relative Difference of Predictions in Series (a) N_4, (b) N_5 and (c) N_6

Table C.1 shows RMSE values of OCI predictions in three series types. Similar to Figure C.3, the finding that CNN models have higher prediction accuracy than any of four piecewise regression models when the length of historical records for the model training is greater than three, indicated by the lowest value of RMSE.

Table C.1 RMSE of OCI Predictions in Different Series Types

Series Type	CNN	Resurfacing	Restoration	Rehabilitation	Reconstruction
N_4	6.42	5.71	5.93	6.13	6.78
N_5	6.75	8.30	8.76	9.06	9.89
N_6	12.96	13.15	14.94	15.27	15.48

As a supplement of results in Chapter 5, example in this appendix shows promising applications of DL models in predicting asphalt pavement performance in light of higher prediction accuracy than the existing method. Besides, implementation of DL models is of a similar level of complexity to piecewise linear regression models considering that data processing, modelling and hyperparameter value determination were finished and provided in this project. Moreover, DL models have more space to improve prediction accuracy and stability by adopting strategies provided in Chapter 5 such as increasing data quantity and dimensionality and model complexity.