

የኢፌዲሪ የቴክኒክና ሙያ ስልጠና ኢንስቲትዩት
FDRE TECHNICAL & VOCATIONAL TRAINING INSTITUTE

FEDERAL TECHNICAL AND VOCATIONAL TRAINING INSTITUTE
FACULTY OF CIVIL TECHNOLOGY
SCHOOL OF GRADUATE STUDIES
DEPARTMENT OF CONSTRUCTION TECHNOLOGY AND MANAGEMENT

A NEURAL NETWORK APPROACH FOR PREDICTING LONG-TERM PAVEMENT PERFORMANCE IN MEKELLE, ETHIOPIA

BY
HAILAY BERIHU ABAY

JUNE, 2026

ADDIS ABABA ETHIOPIA



A NEURAL NETWORK APPROACH FOR PREDICTING LONG-TERM PAVEMENT PERFORMANCE IN MEKELLE, ETHIOPIA

BY

HAILAY BERIHU ABAY

ADVISOR: MESERET GETNET (Ph.D.)

A RESEARCH SUBMITTED TO THE FACULTY OF CIVIL TECHNOLOGY, SCHOOL OF GRADUATE STUDIES, FEDERAL TECHNICAL AND VOCATIONAL TRAINING INSTITUTE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN CONSTRUCTION TECHNOLOGY AND MANAGEMENT

JUNE, 2026

ADDIS ABABA ETHIOPA

DECLARATION

I confirm that research work titled "A Neural Network Approach for Predicting Long-Term Pavement Performance in Mekelle, Ethiopia" is my own work. The work has not been presented elsewhere. Where material has been used from other sources it has been properly acknowledged.

HAILAY BERIHU (ID.NO MTR/589/13) _____ _____

Signature Date

ADVISOR APPROVAL SHEET

I, the principal advisor and supervisor of this thesis, hereby certify that I have closely advised and supervised the student while developing this thesis and read the draft thesis entitled **"**A Neural Network Approach for Predicting Long-Term Pavement Performance in Mekelle, Ethiopia**."** submitted in partial fulfillment of the requirements for the degree of Master Science with specialization in Construction Technology and Management and has been carried out by Hailay Berihu under my supervision/ guidance. Therefore, I recommend that the student has fulfilled the requirements and hence hereby can submit the research to the department.

Name	Signature	Date
Advisor: Meseret Getnet (Ph.D.)		_____

EXAMINERS' APPROVAL SHEET

We, the undersigned, members of the Board of Examiners of the final open defense by Hailay Berihu have read and evaluated his research entitled "A Neural Network Approach for Predicting Long-Term Pavement Performance in Mekelle, Ethiopia" and examined the candidate. This is, therefore, to certify that the research has been accepted in partial fulfillment of the requirements for the Master of Science in Construction Technology and Management.

Approved by:

1.	_____	_____	_____
	Internal Examiner	Signature	Date
2.	_____	_____	_____
	External Examiner	Signature	Date
3.	_____	_____	_____
	Department	Signature	Date
4.	_____	_____	_____
	Dean, Graduate studies	Signature	Date

Final approval and acceptance of the research are contingent upon the submission of the final copy of the research to the office of Postgraduate Studies (OPGS) through the Department/School Graduate Committee (DGC/SGC) of the candidate's department.

SGC: _____

Date: ____/____/_____

ACKNOWLEDGMENTS

Prior to all, I must give many thanks and glory to God for all things being done throughout my life. I wish to express my profound gratitude to my Advisor/ Supervisor: Meseret Getnet (Ph.D.) who did not only team up to provide the guidance and constructive criticism required for the work, but also encouraged me to do good research.

I would like to thank staff of Mekelle City Road Authority and the Tigray Construction Road and Transport Bureau, Mekelle Traffic Police and municipal records, the National Meteorology Agency's Mekelle branch, municipal and construction bureaus and sure construction, involved in responding to project data collection format. Their participations and honesty to show direction for me was wonderful. Those who gave their time and willingness for information, I would like to thank them once again. Further, I would like to thank all those who were participated directly or indirectly for the success of data collection.

Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study. My friends your help and encouragement through the process of researching and writing this thesis was great. This accomplishment would not have been possible without them.

Hailay Berihu

JUNE, 2026

ABSTRACT

Accurate prediction of long-term pavement performance is essential for effective road maintenance planning, reducing rehabilitation costs, and improving sustainable transportation infrastructure management. In developing countries such as Ethiopia, pavement deterioration assessment is often based on conventional engineering approaches and empirical models that may not effectively capture the complex and nonlinear interactions among traffic, structural, and environmental factors affecting pavement behavior over time. This limitation reduces the reliability of maintenance planning and decision-making. Therefore, this study aimed to develop and evaluate an Artificial Neural Network (ANN)-based predictive model for long-term pavement performance of asphalt roads in Mekelle City, Ethiopia, and compare its predictive capability with the traditional Multiple Linear Regression (MLR) model. The study employed a quantitative research approach using historical pavement-related datasets collected from relevant road and infrastructure institutions in Mekelle City. The dataset included key pavement deterioration influencing variables such as pavement age, pavement thickness, traffic volume, traffic load, rainfall, temperature, and road condition-related parameters. Prior to model development, data preprocessing techniques including data cleaning, normalization, and feature selection were conducted to improve model reliability and predictive performance. A Multilayer Perceptron Artificial Neural Network (MLP-ANN) was developed and trained to predict long-term pavement performance, while Multiple Linear Regression (MLR) was used as a baseline model for comparison. Model performance was evaluated using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The results demonstrated that the ANN model significantly outperformed the MLR model in predictive accuracy MLR R^2 value of 0.143, a high MAE of 15.34, and RMSE of 18.78. The ANN achieved high prediction accuracy with an R^2 value of 0.9614, a low MAE of 1.853, and RMSE of 2.239, indicating strong capability in modeling nonlinear pavement deterioration relationships. In contrast, the MLR model showed weaker predictive performance with an R^2 value of 0.143 and a higher MAPE of 32.26%, demonstrating limited effectiveness in handling complex pavement behavior. To improve model interpretability and transparency, Explainable Artificial Intelligence techniques including SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) were applied. The interpretation results consistently identified pavement age, pavement thickness, traffic volume, and traffic load as the most influential factors contributing to pavement deterioration prediction. The study concludes that ANN-based predictive modeling provides a more accurate, reliable, and practical approach for long-term pavement performance prediction compared with conventional regression techniques under local Ethiopian road conditions. Additionally, the integration of explainable machine learning methods contributes to improved understanding of model behavior and supports evidence-based pavement maintenance planning. This research provides a useful foundation for future pavement infrastructure management studies and the application of machine learning techniques in road engineering within similar developing-country contexts.

Keywords: Pavement Performance Prediction, Artificial Neural Network (ANN), Multiple Linear Regressions (MLR), Pavement Deterioration, SHAP, LIME, Explainable.

TABLE OF CONTENTS

DECLARATION i
ADVISOR APPROVAL SHEET ii
EXAMINERS' APPROVAL SHEET iii
ACKNOWLEDGMENTS iv
ABSTRACT v
LIST OF TABLES x
LIST OF FIGURE xi
LIST OF EQUATION xii
LIST OF ACRONYMS xiii
CHAPTER ONE 1
1. INTRODUCTION 1
1.1 BACKGROUND OF THE STUDY 1
1.2 Statement of the Problem 3
1.3 Research Questions 4
1.4. Objectives of the Study 5
1.4.1 General Objectives 5
1.4.2 Specific objectives 5
1.5 Significance of the Study 5
1.6 Scope of the Study 6
1.7 Limitation of the Study 7
1.8 Organization of the Research 8
CHAPTER TWO 9
LITERATURE REVIEW 9
2.1 Introduction 9
2.1.1 Theoretical review 9
2.2 Pavement Deterioration and Performance Modeling 10
2.3 Evolution of Pavement Performance Prediction Models 11
2.3.1 Mechanistic-Empirical (M-E) Models: 12
2.3.2 Strengths of M-E Models 12

2.4 Machine Learning in Pavement Engineering 12
2.5 The Critical Role of Environmental Factors and Local Context 14
2.6 Enhancing Trust and Usability: Model Interpretability 15
2.7 Empirical Review Literature 16
2.7.1 Interpretation of Empirical Literature (Based on Table 2.1) 18
2.8 Summary and Identification of Research Gaps 20
2.8.1 Summary 20
2.8.2 Research Gaps 21
CHAPTER THREE 23
RESEARCH METHODOLOGY 23
3.1 Introduction 23
3.2 Study Area 23
3.3 Research Design and Approach 24
3.4 Population, Sample Size, and Sampling Technique 27
3.4.1 Population of the Study 27
3.4.2 Sample Size Determination 27
3.4.3 Sampling Technique 28
3.5 Data Sources and Collection 28
3.5.1 Data Sources 28
3.5.2 Methods of Data Collection 29
3.5.2.1 Questionnaire Survey 29
3.5.2.2 Interview 30
3.5.2.3 Case Study 30
3.6 Data Preprocessing and Feature Engineering 30
3.6.1 Data Preprocessing 30
3.6.2 Feature Engineering 31
3.7 Model Development and Training 31
3.7.1 Data Analysis Techniques 31
3.7.2 Relative Importance Index (RII) 31
3.7.3 Linear Regression Model 32

3.7.3.1 Multiple Linear Regression (MLR) 32
3.7.4 Model Development 33
3.7.4.1 Artificial Neural Network (ANN) 34
3.7.5 Model Training 35
3.8 Model Evaluation and Interpretability 35
3.8.1 Model Evaluation 35
3.8.2 Model Interpretability 35
3.9 Software 36
3.10 Reliability and Validity of the Research 36
3.10.1. Reliability analysis 37
3.10.2. Validity analysis 38
CHAPTER 4 41
DATA ANALYSIS AND DISCUSSION 41
4.1 Background Information about Respondents 41
4.1.1 Rate of return of questionnaire 41
4.1.2 Working experience of respondent 42
4.1.3 Education level of respondent 43
4.2 Major Significant Factors Affect Performance Prediction Road Projects 44
4.3 Presentation of Data 48
4.4 Data Mining Techniques 53
4.4.1 Data Correlation 53
4.5 Multiple Linear Regressions (MLR) for Pavement Performance Prediction 56
4.5.1 Assumptions of Multiple Linear Regression Analysis 56
4.5.2 Assumptions of multiple linear regressions Analysis 59
4.5.2 Validation of the MLR Model 61
4.6: Artificial Neural Network Structure for Pavement Performance Prediction 62
Statistics for Target Variable: PCI 68
Summary of Variables 68
Neural Network Parameters 69
4.7, Compare ANN and MLR 71

Interpretation of Results 72
4.8 SHapley Additive exPlanations (SHAP) 72
4.9Local Interpretable Model-Agnostic Explanations (LIME) 74
4.10 COMPARE LIME, SHAP and EXPERT-BASED ASSESSMENT RANKING 75
CHAPTER 5 76
SUMMARY OF MAIN FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS 76
5.1 Discussion of Main Findings and Implications 76
5.2 Conclusions 77
5.3 RECOMMENDATIONS 79
5.3.2 Recommendations for Further Research Works 79
REFERENCES 81
Appendix 1 85

LIST OF TABLES

Table 2.1: Empirical Review Literature on a Neural Network Approach for Predicting Long-Term Pavement Performance 16
Table 3.1 Reliability Statistics for Objective (Quantitative) factors 37
Table 3.2 Reliability Statistics for Subjective (Qualitative) factors 37
Table 3.3 Reliability Statistics for other factors 37
Table 4.1 Respondent rate of return of questionnaire 41
Table 4.2 Relative importance index for Objective (Quantitative) factors 44
Table 4.3 Relative importance index for Subjective (Qualitative) factors 45
Table 4.4 Relative importance index for other factors 46
Table 4.5: Description of Variables 48
Table 4.6 Performance of Asphalt Road Project Mekelle City 52
Table 4.7: Correlation Matrix before Variable Transformation 53
Table 4.8: Correlation Matrix after Variable Transformation 55
Table 4.9: Multiple Linear Regression Model Summaries for Pavement Performance Prediction 56
Table 4.10: ANOVA for Pavement Performance Prediction Model 57
Table 4.11: Coefficients for Pavement Performance Prediction Model 58
Table 4.12: Comparison of Actual and Predicted PCI using the MLR Model 61
Table 4.13: General Statistics of the ANN Predictive Model (Using Python Neural Network) .. 68
Table 4.14: Statistics for the Training Data for the ANN Predictive Model 69
Table 4.15: Comparison of Actual and Predicted PCI using the ANN Model 70
Table 4.16: Performance Comparison between MLR and ANN Models for PCI Prediction 71
Table 4.17: Comparison of Influential Pavement Deterioration Factors Identified Using LIME, SHAP, and Expert-Based Assessment 75
Table 5.1: PCI Classification and Recommended Maintenance Actions 80

LIST OF FIGURE

Figure 3.1 Source Google map 24
Figure 3.2: Research process adopted for this Study 25
Figure 3.3: ANN development flow chart. 34
Figure 4.1: Respondents' working experience 42
Figure 4.2: Respondents' level of education. 43
Figure 4.6: Histogram of standardized residual for MLR Model 59
Figure 4.7: Scatter Plot of Standardized Residual versus Predicted Pavement Condition Index (PCI) for the Developed Multiple Linear Regression (MLR) Model. 60
Figure 4.8: Multilayer Perceptron (Python Neural Network) structure for pavement performance prediction. 63
Figure 4.9: Workflow of ANN Pavement Performance Prediction Model 64
Figure 4.10: Actual versus Predicted PCI Values for ANN Model 65
Figure 4.11: Regression plot between Actual versus Predicted PCI Values for ANN Model 66
Figure 4.12: ANN Training and Validation Loss Curve 67
Figure 4.13: SHAP Feature Importance for ANN Pavement Condition Index (PCI) Prediction Model 73
Figure 4.14: LIME Local Feature Explanation for ANN Pavement Condition Index (PCI) Prediction Model. 74

LIST OF EQUATION

Equation 3.1 Normalization formula 30
Equation 3.2 The Relative Importance Index (RII) formula..... 32
Equation 3.3 The general form of the MLR model 33
Equation 3.4 The ANN mathematical expression..... 34
Equation 3.5 The formula for calculating (R^2) 37
Equation 3.6 Mean Absolute Percentage Error (MAPE)..... 37
Equation 3.7 Root Mean Squared Error (RMSE) 37

LIST OF ACRONYMS

AASHTO	American Association of State Highway and Transportation Officials
ASTM	American Society for Testing and Materials
ANOVA =	Analysis of Variance
ANN	Artificial Neural Network
AI	Artificial Intelligence
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-agnostic Explanations
ML	Machine learning
LTPP	Long-Term Pavement Performance
FHWA	Federal Highway Administration
MLR	Multiple Linear Regression
M-E	Mechanistic-Empirical
IRI	International Roughness Index
ADT	Average Daily Traffic
ESALs	Equivalent Single Axle Loads
PCI	Pavement Condition Index
R^2	Coefficient of Determination
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Squared Error

CHAPTER ONE

1. INTRODUCTION

1.1 BACKGROUND OF THE STUDY

The construction industry plays a crucial role in the development of any country, contributing to the national economy, job creation, and improving living standards. In urban areas, the demand for Infrastructure is expected to increase significantly in the coming years, making the construction sector even more important (Eliang, 2023). In the context of Mekelle, Ethiopia, the construction industry has experienced notable growth since 2001, with its contribution to the national GDP reaching approximately 5.6% (Tebeje Zewdu, 2015). This growth is expected to continue, supported by rising foreign direct investment and the resumption of halted development projects.

The development of a reliable transportation infrastructure is essential for economic growth, public safety, and improved quality of life. Road pavements, as critical components of transportation systems, are continuously subjected to traffic loading, environmental conditions, and material deterioration. As a result, predicting pavement performance over time is importance for effective maintenance planning, cost reduction, and sustainable infrastructure management. In recent years, Artificial Neural Networks (ANNs) have become one of the most effective machine learning techniques for predicting pavement deterioration because they can handle complex and nonlinear relationships among different variables. Previous studies have applied ANN models to predict pavement condition indices such as PCI by using variables like pavement age, traffic loading, and environmental factors. The findings showed that ANN models provide accurate predictions of long-term pavement performance. However, most of these studies were conducted on airport pavements and in developed countries, with limited focus on urban road pavements under local conditions such as those found in Mekelle City, Ethiopia (Mitchell, 1995).

Pavement performance prediction plays a key role in road asset management systems, particularly in determining maintenance and rehabilitation strategies. Traditionally, such

predictions have relied on empirical and regression-based models, such as Multiple Linear Regression (MLR)

those introduced in early AASHTO design guides (Abdollahi et al., 2023). While valuable, these models often oversimplify the complex and non-linear interactions between causal factors. To address this, Mechanistic-Empirical (M-E) models were developed, which integrate principles of structural mechanics with empirical observations to provide a more sophisticated analysis of pavement response under various loads (Burmister, 1945), (Adlinge & Gupta, 2009). However, the efficacy of M-E models is highly dependent on extensive calibration with high-quality, region-specific data, limiting their practical application in data-scarce contexts like Ethiopia.

The advent of artificial intelligence has introduced a paradigm shift in civil engineering, with Artificial Neural Networks (ANNs) emerging as powerful tools for modeling complex systems. Inspired by the biological structure of the human brain, ANNs can identify intricate, non-linear patterns within large datasets without pre-defined physical models [(Huang & Moore, 1997), (Abambres & Ferreira, 2019a)]. Early applications demonstrated the superiority of ANNs over traditional statistical methods for predicting specific pavement distresses (Alharbi & Smadi, 2019), (Abambres & Ferreira, 2019b). More recently, Artificial Neural Networks (ANNs) have demonstrated strong capability in modeling complex nonlinear pavement deterioration behavior, have shown exceptional promise in capturing the temporal dependencies inherent in long-term pavement deterioration (Xin et al., 2023), (Hochreiter & Schmidhuber, 1997).

Environmental factors, including temperature, rainfall, and moisture variations, significantly influence pavement performance. Studies have shown that incorporating these variables improves prediction accuracy (Rezapour et al., 2022), (Cai et al., 2023). This is particularly important in regions like Mekelle, Ethiopia, which experiences a semi-arid climate with significant temperature fluctuations. Despite the growing application of ANN models globally, limited research has focused on urban road pavements under local Ethiopian conditions. To ensure these advanced models are practical for engineering decision-making, the challenge of their "black box" nature must be addressed. Interpretability techniques like SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) have been developed to provide transparency, allowing engineers to understand the key factors driving a

model's predictions (Shen et al., 2022), (Adolph, 2016). Therefore, this study aims to apply Artificial Neural Network techniques to predict long-term pavement performance in Mekelle City by incorporating traffic, material, and environmental factors. Additionally, interpretability methods such as SHAP and LIME considered enhancing model transparency and supporting practical decision-making in pavement management.

1.2 Statement of the Problem

The lack of reliable local pavement performance prediction models in Mekelle City limits the ability of road agencies to forecast deterioration trends and allocate maintenance resources effectively. This limitation results in mostly reactive maintenance strategies rather than proactive planning, leading to higher lifecycle costs, reduced pavement service life, and possible risks to user safety. The core of this problem is caused by several interrelated challenges;

Firstly, pavement deterioration is an inherently complex and non-linear process, driven by the dynamic interactions between traffic loads, material properties, and environmental stressors. Traditional empirical and mechanistic-empirical models often fail to fully capture these complexities or require extensive local calibration, which is frequently impractical (Abdollahi et al., 2023), (Mahmood et al., 2016). The local data available in Mekelle, including pavement condition surveys, traffic counts, and climate records, are often inconsistent, incomplete, or stored in heterogeneous formats, thereby complicating the development of reliable pavement performance prediction models. During data collection from the Mekelle City Municipality, pavement information was obtained from multiple sources, including pavement inspection reports, traffic surveys, and climate records. Several datasets contained missing observations, inconsistent variable formats, and non-standardized recording systems. Consequently, substantial data cleaning and preprocessing were required before model development.

Secondly, Mekelle's semi-arid climate imposes unique environmental stresses that are not adequately addressed by generic prediction models. High diurnal temperature swings induce thermal fatigue and cracking, while intense solar radiation accelerates binder aging and seasonal rainfall contributes to moisture-related damage (Dahmani et al., 2022). Existing models, often developed for temperate climates, do not sufficiently account for these specific regional

influences, leading to inaccurate performance forecasts. This is compounded by operational factors such as variability in construction quality and increasing traffic loads from rapid urbanization, which further accelerate deterioration (Abambres & Ferreira, 2019a).

Finally, while advanced machine learning models like ANNs offer a powerful solution for capturing these complex relationships, their lack of transparency presents a significant barrier to practical implementation. For pavement management authorities to trust and adopt these models, it is critical to understand why a prediction is being made. Without transparent and interpretable outputs, it is difficult to formulate targeted and defensible maintenance interventions (Wilson & Partners, 2020).

Therefore, a critical research gap exists for a predictive framework that is specifically tailored to the environmental and operational context of Mekelle. There is a pressing need for a model that not only leverages advanced, time-aware neural network architectures to achieve high predictive accuracy but also incorporates interpretability techniques to provide actionable insights for local road authorities. This study aims to address this gap by developing and validating such a model, thereby providing a foundational tool for enhancing the sustainability and resilience of road infrastructure in Mekelle.

1.3 Research Questions

To achieve the objectives of this study, the following research questions are formulated:

- What are the essential components, characteristics, and preprocessing requirements for constructing a dataset to effectively model long-term pavement performance in Mekelle?

- What extent can the MLP-based Artificial Neural Network model improve the prediction of long-term pavement deterioration in Mekelle?

- How does the predictive accuracy (measured using RMSE, MAE, and R^2) of neural network models compare with traditional statistical regression methods in predicting long-term pavement performance?

• Which historical, environmental, and traffic-related factors most significantly influence long-term pavement deterioration in Mekelle, and what is their relative importance as determined by the developed neural network model and compare the findings with expert-based assessments?

1.4. Objectives of the Study

1.4.1 General Objectives

The primary objective of this study is to develop and validate a reliable neural network-based model for predicting the long-term performance of asphalt pavements in Mekelle by integrating historical condition data, traffic loads, and local environmental factors.

1.4.2 Specific objectives

The Specific objectives of the study are

- To compile and preprocess a comprehensive dataset for pavement performance in Mekelle, incorporating historical condition surveys.
- To train and optimize an MLP-based Artificial Neural Network model for predicting pavement deterioration over a 2–15 year period.
- To benchmark the predictive accuracy of the developed neural network models against traditional statistical regression methods.
- To identify and quantify the most influential factors driving pavement deterioration in Mekelle's semi-arid climate using model interpretability techniques and compare the findings with expert-based assessments.

1.5 Significance of the Study

This study contributes to both academic knowledge and engineering practice, particularly in the field of urban road infrastructure management in developing countries.

The significance of this study can be articulated through several key areas.

First, the study develops a predictive model tailored to Mekelle City, which can support local road authorities in making informed, data-driven decisions. By improving the accuracy of long-term pavement performance prediction, the model enables a shift from reactive maintenance practices to proactive and preventive strategies. This transition is expected to enhance resource allocation, reduce maintenance costs, improve road safety, and extend the service life of pavement infrastructure.

Second, the study provides a methodological contribution through the application and evaluation of artificial neural network models in a data-constrained and semi-arid environment. While such models have been widely applied in other regions, their use in the Ethiopian context remains limited. Therefore, this research offers insight into the applicability, performance, and limitations of these techniques under local conditions, and establishes a framework that can be adapted to other cities with similar characteristics.

Third, the study enhances model transparency by incorporating interpretability techniques such as SHAP and LIME to identify the most influential factors affecting pavement performance. This allows engineers and decision-makers to better understand the relationship between environmental, material, and traffic variables and pavement deterioration. As a result, the findings can support improvements in pavement design, construction practices, and maintenance planning.

Finally, the study compiles and processes a structured dataset of pavement condition, traffic, and environmental variables for Mekelle City. This dataset can serve as a valuable resource for future research and support the advancement of data-driven approaches in infrastructure management within Ethiopia.

1.6 Scope of the Study

The scope of this research is defined to ensure a focused and feasible investigation while addressing the main research objectives. The study focuses on the development and evaluation of predictive models for asphalt road pavements within the urban area of Mekelle, Ethiopia.

The methodological scope emphasizes the application of Artificial Neural Network (ANN) models using Multilayer Perceptron (MLP) architecture. The performance of these models is compared with traditional statistical regression methods to provide a comprehensive benchmarking analysis. The study focuses on long-term pavement performance prediction for periods ranging from 2 to 15 years.

The input data for the models are obtained from selected road sections in Mekelle and include historical pavement condition indicators, material property, traffic characteristics (e.g., Average Daily Traffic (ADT)), and environmental factors, including temperature and rainfall. The study also incorporates model interpretability analysis using techniques such as SHAP or LIME to identify the key factors influencing pavement deterioration and expert-based assessments.

This study is subject to several delimitations. It does not consider concrete or unpaved road types, focusing only on asphalt pavements. The developed models are based on data from Mekelle; therefore, their direct applicability to other regions with different environmental and traffic conditions requires further calibration and validation. Additionally, the study is constrained by the availability and quality of historical data, and road sections with incomplete or inconsistent records may be excluded. Finally, while different configurations of the MLP-based ANN model were evaluated, the study does not cover all possible model configurations but instead focuses on selected, well-established architectures relevant to the research problem.

1.7 Limitation of the Study

The limitation of this study includes: Data quality and availability: The study relies on historical data, which may be incomplete, inconsistent, or limited in coverage. Such data limitations can affect the accuracy of model training and the reliability of predictions. Institutional Constraints: Limited coordination among relevant governmental agencies and resource constraints may restrict data collection, validation, and access to updated information, thereby influencing the comprehensiveness of the dataset. Technical and Knowledge Constraints: The application of advanced neural network models and interpretability techniques may be constrained by limited local expertise and familiarity. This could affect model implementation, interpretation, and

practical adoption by practitioners. Environmental Variability: The semi-arid climate of Mekelle, characterized by high temperature fluctuations and seasonal rainfall, introduces additional complexity in pavement performance behavior. These conditions may limit the generalizability of the developed models to other regions with different environmental characteristics.

1.8 Organization of the Research

This study contains five main chapters. Chapter one introduction, chapter two literature review, chapter three methodology, chapter four result and discussion and chapter five conclusion and recommendation. Each of the above chapters has the following contents;

Chapter one- describes the research overview, background, statement problem, objectives, scope, limitation and significance of the study. It also indicates the organization contents of the research.

Chapter two- covers the basic concepts of the research and literature review part of the thesis. It discusses Pavement Deterioration and Performance Modeling, Evolution of Pavement Performance Prediction Models, Machine Learning in Pavement Engineering, The Critical Role of Environmental Factors and Local Context, Enhancing Trust and Usability: Model Interpretability, and summary of the literature and research gap.

Chapter three- covers the research methodology. It contains introduction, location of the study area, research design, population of the study, sample size determination, sampling techniques, source of data, method of data collection, method of data analysis, validity and reliability and ethical considerations.

Chapter Four- contains the data analysis and research discussion. It also includes introduction, questionnaire survey response rates, demographic characteristics of respondents, data pre-processing.

Chapter five- contains the summery of main findings, conclusion and recommendations are presented.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Literature review involves the systematic identification, evaluation, and synthesis of existing research related to a specific problem. This chapter presents a comprehensive review of literature on long-term pavement performance prediction using neural networks, with particular emphasis on the role of historical data and environmental factors.

The chapter discusses key theories, concepts, and modeling approaches used in pavement engineering, including traditional empirical methods, mechanistic-empirical models, and modern machine learning techniques. It further examines the relationships among traffic loading, environmental conditions, and pavement deterioration. This review provides the foundation for understanding the research problem and highlights the need for improved predictive models tailored to local conditions.

2.1.1 Theoretical review

According to (Adams et al., 2017), a theory is a set of systematically interrelated concepts, definitions and propositions that are advanced to explain and predict phenomena (facts). In pavement engineering, theoretical frameworks are essential for understanding the mechanisms of pavement deterioration and for developing predictive models. Pavement deterioration theory explains how factors such as traffic loading, environmental conditions and material properties interact over time to influence pavement performance. Mechanistic theory focuses on stress–strain relationships within pavement layers, while empirical theory relies on observed performance data.

In the context of machine learning, predictive modeling theory emphasizes the ability of algorithms to learn patterns from historical data and generalize them to future conditions. Artificial Neural Networks (ANNs), in particular, are based on the theory of interconnected processing units that can approximate complex nonlinear relationships.

These theoretical perspectives form the foundation for integrating engineering principles with data-driven approaches in pavement performance prediction.

2.2 Pavement Deterioration and Performance Modeling

Several studies focused on pavement deterioration and performance modeling using different approaches, including deterministic, probabilistic, and artificial neural network (ANN) methods. They commonly studied the effects of traffic loading, pavement age, and environmental factors such as temperature and freeze–thaw cycles on pavement condition. Among the methods, ANN models showed better accuracy in predicting pavement distress and long-term performance than traditional regression and Markov models (Alharbi & Smadi, 2019), (Mahmood et al., 2016), and(Mohammadi et al., 2019).

Pavement deterioration is primarily influenced by Traffic Loading: Increased axle loads and traffic volume lead to accelerated wear, represented by Equivalent Single Axle Loads (ESALs).

Environmental Factors: Temperature fluctuations cause thermal cracking and binder aging, while moisture can lead to stripping and weaken the subgrade. Solar radiation also affects material durability. Material Properties and Aging: The performance of asphalt binders and aggregates, along with structural design, significantly impacts long-term durability Pavement Performance Indicators (PPIs) (Abambres & Ferreira, 2019b).

Key metrics for assessing pavement condition include: Functional Performance: The International Roughness Index (IRI) measures ride quality, indicating surface smoothness. Structural Performance: Metrics like rutting (permanent deformation) and various cracking types (fatigue, thermal, top-down) provide insights into the pavement's structural integrity and health (Dahmani et al., 2022). Mekelle’s semi-arid climate presents unique challenges for pavement durability. High diurnal temperature swings, seasonal rainfall, and intense solar radiation cause specific distress mechanisms like thermal cracking, moisture damage, and accelerated aging (Dahmani et al., 2022),(Paper et al., 2016). Limited local data availability and resource constraints hinder precise modeling and proactive maintenance planning. Incorporating region-specific environmental data and leveraging advanced neural network models are essential steps toward improving pavement management practices in Mekelle.

2.3 Evolution of Pavement Performance Prediction Models

The study focuses on the use of machine learning techniques in pavement management systems, especially for condition assessment, performance prediction, and maintenance decision-making. It shows that ML models improve prediction accuracy compared to traditional empirical models (Tamagusko & Correia, 2024). The study develops an analysis-ready pavement materials dataset by organizing layer properties and material parameters into a structured database to support pavement performance analysis and model development (Afsharikia, 2023).

According Abdollahi et al., (2023),The evolution of pavement performance prediction models reflects the ongoing need for accurate and reliable methods to assess the longevity and structural integrity of pavement systems. Historically, empirical models have dominated this field, notably the AASHTO Design Guide, which remains a cornerstone in pavement design. However, as traffic demands and material technologies advance, there is a pressing need to develop new methodologies that better address these changes.

Empirical models are built on the foundation of observed data from real-world pavement performance. The AASHTO Design Guide, cited by researcher, exemplifies this approach. It utilizes historical data from controlled experiments, such as the AASHO Road Test, to establish relationships between traffic loads, material properties, and pavement performance.

The methodology of empirical models typically involves regression analysis of field data. By collecting and analyzing performance data from various pavement sections, researchers derive equations that predict future performance based on current conditions. This regression-based approach allows for the identification of key variables that influence pavement behavior under specific conditions.

The limitations of the researcher are Lack of Transferability, Oversimplification of Complex Interactions and Inability to Account for New Materials or Loads.

2.3.1 Mechanistic-Empirical (M-E) Models:

Mechanistic-Empirical (M-E) models, such as the AASHTO Pavement ME Design, represent a significant advancement over traditional empirical methods by integrating mechanical principles of stress, strain, and deflection with empirical calibration (Burmister, 1945) (Adlinge & Gupta, 2009). This integration allows for a more accurate representation of pavement behavior under varying conditions.

2.3.2 Strengths of M-E Models

Robustness: M-E models provide a comprehensive framework that can adapt to specific loading and environmental scenarios, improving design accuracy and Predictive Capability: They facilitate the analysis of various factors affecting pavement performance, leading to better decision-making in design and maintenance.

Despite their strengths, M-E models have notable limitations like High Data Requirements: They necessitate extensive data inputs, including material properties, traffic patterns, and environmental conditions, which can be challenging to gather, Complex Calibration: The calibration process can be intricate, requiring expertise and access to sophisticated software tools and Regional-Specific Data Needs: The effectiveness of M-E models often hinges on having regionally specific data, making them particularly difficult to implement in data-scarce regions like Ethiopia.

2.4 Machine Learning in Pavement Engineering

Introduction to Artificial Neural Networks (ANNs):

Artificial Neural Networks (ANNs) are computational models inspired by biological neurons in the human brain. They consist of interconnected neurons organized into layers that process input signals and transmit outputs. ANNs mimic the communication and learning processes of biological neurons, adapting their connections through training. A key advantage of ANNs is their ability to model complex, non-linear relationships without needing explicit physical equations. This flexibility allows them to effectively solve intricate problems across various

fields. Consequently, ANNs have become essential tools in data analysis and predictive modeling(Huang & Moore, 1997), (Abambres & Ferreira, 2019a). The study introduced Artificial Neural Networks (ANNs) as a machine learning technique capable of modeling nonlinear relationships between pavement age, traffic, climate, and condition indices for accurate long-term pavement performance prediction(Mitchell, 1995).Early studies applying basic feed forward networks, specifically Multi-Layer Perceptron's (MLPs), demonstrated their effectiveness in predicting pavement distresses. These studies often highlighted MLPs' superior ability to model complex, non-linear relationships compared to traditional statistical methods. Findings indicated that MLPs yielded higher accuracy and better predictive performance, particularly in capturing the intricacies of pavement deterioration. This established a foundation for further advancements in ANN applications within pavement engineering (Alharbi & Smadi, 2019) (Abambres & Ferreira, 2019b).

Advanced Neural Network Architectures for Time-Series Forecasting:

Recent advances in deep learning have introduced architectures such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer models for time-series forecasting. LSTM and GRU improve upon traditional recurrent neural networks by addressing the vanishing gradient problem and effectively capturing long-term temporal dependencies(Hochreiter & Schmidhuber, 1997). Transformer models employ attention mechanisms to model long-range relationships within sequential data and have demonstrated strong performance in a variety of forecasting applications (Xin et al., 2023), (Mitchell, 1995). Although these architectures have shown promising results, they were not implemented in this study because the research focused on developing and evaluating a Multilayer Perceptron (MLP)-based Artificial Neural Network for pavement performance prediction using the available dataset. Given the relatively limited sample size and the objectives of comparing MLP with Multiple Linear Regression (MLR), the MLP was considered an appropriate and computationally efficient modeling approach. Nevertheless, LSTM, GRU, and Transformer models represent potential directions for future research using larger longitudinal pavement datasets.

2.5 The Critical Role of Environmental Factors and Local Context

Previous studies have shown that environmental factors and local context play a critical role in pavement performance and deterioration. The Montreal case study found that climate-related conditions, particularly freeze–thaw cycles, significantly accelerated pavement deterioration and in some cases had a greater impact than traffic loading. Similarly, the machine learning–based PCI prediction study emphasized that pavement condition is strongly influenced by local climate, distress characteristics, traffic, and maintenance history, showing that prediction accuracy improves when local field conditions are considered. In addition, the LTPP materials dataset study highlighted that pavement performance depends on environment, subgrade soil, material properties, loading, and maintenance practices, confirming the importance of region-specific factors in performance modeling. Together, these studies indicate that reliable pavement performance models must incorporate environmental and local contextual variables rather than relying on generalized models developed for other regions (Mohammadi et al., 2019), (Huang & Moore, 1997) and(Afsharikia, 2023).

Integrating environmental data, such as temperature and precipitation, has been shown to significantly enhance the accuracy of pavement performance models. Studies indicate that including these climatic variables as inputs leads to more reliable predictions, addressing the multifaceted nature of pavement deterioration. This connection underscores the importance of contextualizing advanced modeling techniques within specific environmental conditions to effectively tackle the research problem at hand (Mohammadi et al., 2019) , (Hemed et al., 2020). Studies conducted in semi-arid regions like Relizane, Algeria, have effectively applied neural networks to predict environmental conditions such as solar radiation under high temperature variability. These regions experience harsh climatic effects similar to Mekelle, including extreme heat and diurnal temperature swings. Such conditions are known to accelerate thermal cracking, moisture-induced damage, and surface fatigue in pavements. This validates the need for localized, AI-driven pavement performance models sensitive to semi-arid climate stressors (Dahmani et al., 2022).

2.6 Enhancing Trust and Usability: Model Interpretability

The complexity of artificial neural networks (ANNs) often renders them “black boxes,” limiting their adoption in engineering fields that demand transparent decision-making. In pavement performance prediction, where investment and safety decisions are critical, trust in model outputs is essential. SHAP and LIME offer post-hoc interpretability, allowing engineers to understand feature influence on predictions. These tools bridge the gap between AI accuracy and practical usability, especially in resource-constrained contexts like Mekelle (Wilson & Partners, 2020).

Recent applications of machine learning in road construction have demonstrated high predictive accuracy but limited transparency, raising the challenge of the "black box" problem in ANN-based models. This makes their practical adoption difficult in engineering contexts, where interpretability and trust are essential. SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) have emerged as vital tools in Explainable AI (XAI), offering a way to quantify the impact of each input feature on the model’s predictions. These techniques are particularly helpful in complex regression tasks like pavement performance forecasting. Studies in civil engineering have used SHAP/LIME to reveal how traffic load, temperature, and moisture levels influence pavement deterioration, providing actionable insights for decision-makers. By integrating interpretability, machine learning becomes not only accurate but also transparent and usable. Their application addresses the usability gap between AI models and real-world infrastructure decision-making, especially in data-scarce regions like Ethiopia(Shen et al., 2022), (Adolph, 2016).

2.7 Empirical Review Literature

Table 2.1: Empirical Review Literature on a Neural Network Approach for Predicting Long-Term Pavement Performance.

No	Author(year)	Country	Title	Finding
1	(Weng On Tam, 2016)	United States	Use of Long-Term Pavement Performance Data to Develop Traffic Defaults in Support of Mechanistic-Empirical Pavement Design Procedures	Established traffic defaults using LTPP data to enhance mechanistic-empirical pavement design. .
2	(Marcelino et al., 2019)	Portugal	Machine learning approach for pavement performance prediction	Proposed a machine learning framework for predicting pavement performance using various algorithms, emphasizing generalization capabilities.
3	(**Qiang Wang,** 2022)	China	A Comparative Study of AI-Based International Roughness Index (IRI) Prediction Models for Jointed Plain Concrete Pavement (JPCP)	Compared machine learning models (SVM, DT, RF) for predicting IRI of JPCP, finding RF had the highest accuracy.

4	(Meegoda et al., 2015)	United States	Roughness Progression Model for Asphalt Pavements Using Long-Term Pavement Performance Data	Developed an IRI prediction model using LTPP data, highlighting the influence of traffic load, climate, and pavement structure on roughness progression.
5	(Mohandes et al., 2022)	Hong Kong	Towards enhancement in reliability and safety of construction projects: developing a hybrid multi-dimensional fuzzy-based approach	Developed a 5D-SRAM model integrating fuzzy AHP and FTOPSIS for comprehensive safety risk assessment in construction.
6	(Junaid et al., 2024)	China	Evaluation of asphalt mixtures modified with low-density polyethylene and high-density polyethylene using experimental results and machine learning models	LDPE-modified asphalt mixtures showed significantly better resistance to permanent deformation and higher dynamic modulus compared to HDPE and control mixtures.
7	(Learning et al., 2021)	Singapore	Deep Learning Based Pavement Inspection Using Self-Reconfigurable Robot	Developed a deep learning framework for pavement inspection using a self-reconfigurable robot, achieving high detection accuracy for cracks and garbage.

8	(Lundberg & Lee, 2017)	USA	A Unified Approach to Interpreting Model Predictions	Introduced SHAP values as a unified framework for interpreting predictions of complex models, enhancing interpretability while maintaining accuracy.
9	(Ribeiro & Guestrin, 2016)	USA	"Why Should I Trust You?" Explaining the Predictions of Any Classifier	Proposed LIME, a method for interpreting model predictions by locally approximating complex models with interpretable ones.
10	(Gulcehre et al., 2014)	Canada	Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation	Proposed an RNN Encoder-Decoder model that improves translation performance by capturing linguistic regularities in phrase pairs.

2.7.1 Interpretation of Empirical Literature (Based on Table 2.1)

The empirical studies presented in Table 2.1 demonstrate the progressive development of pavement performance analysis and prediction techniques, with a clear transition from traditional data-driven approaches toward advanced machine learning and artificial intelligence methods.

Early studies, such as Weng on Tam, (2016) and Meegoda et al. (2015), rely heavily on Long-Term Pavement Performance (LTPP) data to support mechanistic-empirical pavement design and roughness prediction. These studies emphasize the critical role of traffic loading, climatic conditions, and pavement structure in influencing pavement performance. Their findings

highlight the importance of high-quality historical datasets in developing reliable predictive models. However, they also indicate a dependence on well-established data systems.

Subsequent research, including Marcelino et al. (2019) and Mohandes et al (2022), reflects a shift toward machine learning techniques for pavement performance prediction. These studies demonstrate that algorithms such as Random Forest, Support Vector Machines, and Decision Trees can effectively model pavement condition indicators like the International Roughness Index (IRI). In particular, the superior performance of Random Forest reported by Mohandes et al., (2022) suggests that ensemble learning methods offer improved predictive accuracy compared to traditional approaches.

In addition to performance prediction, some studies extend machine learning applications to related areas within pavement engineering. For example, Balakrishnan Ramalingam, (2021) applied deep learning for automated pavement inspection, achieving high accuracy in detecting surface distresses. Similarly, Junaid et al. (2024) integrated machine learning with experimental analysis to evaluate asphalt mixture performance, demonstrating the potential of AI in material characterization and optimization.

Furthermore, the inclusion of studies by Lundberg and Lee (2017) and Ribeiro and Guestrin (2016) highlights the growing importance of model interpretability in machine learning applications. These studies introduced SHAP and LIME, respectively, as methods for explaining complex model predictions. Their contributions address a key limitation of advanced models, namely the lack of transparency, which is particularly critical in engineering decision-making contexts.

The study by Gulcehre et al. (2014), although originating from the field of machine translation, introduces the RNN Encoder-Decoder architecture, which forms the foundation for advanced sequence modeling techniques. This development is relevant to pavement performance prediction, where time-dependent data plays a significant role.

Despite these advancements, several common limitations emerge from the reviewed studies. First, the methodologies vary considerably, ranging from traditional statistical approaches to machine learning and deep learning techniques, making direct comparisons difficult. While conventional models such as Multiple Linear Regression (MLR) provide interpretable relationships between variables, machine learning methods, including Random Forest and Artificial Neural Networks (ANNs), generally demonstrate superior capability in modeling the nonlinear behavior of pavement deterioration. Second, most studies were conducted using large, well-structured datasets from developed countries, limiting the applicability of their findings to data-constrained environments such as Ethiopia. Third, although prediction accuracy has been widely investigated, relatively few studies have incorporated explainable artificial intelligence techniques, such as SHAP and LIME, to improve the transparency and interpretability of prediction models.

Based on these identified gaps, the present study contributes to the existing body of knowledge by comparing the predictive performance of Multiple Linear Regression (MLR) and a Multilayer Perceptron (MLP) neural network using pavement condition, traffic, and environmental data collected from Mekelle City, Ethiopia. Furthermore, the study integrates SHAP and LIME to explain the predictions of the developed MLP model, thereby improving model transparency and supporting practical decision-making in pavement management under data-limited conditions.

2.8 Summary and Identification of Research Gaps

2.8.1 Summary

This chapter reviewed existing literature on pavement deterioration mechanisms and performance prediction models, with a focus on both traditional and advanced approaches. The review showed that pavement performance is influenced by multiple interacting factors, including traffic loading, environmental conditions, and material properties.

Traditional empirical and mechanistic-empirical (M-E) models have been widely used for pavement design and performance prediction. While these models provide useful insights, they

often struggle to capture complex, nonlinear relationships among influencing variables and require extensive calibration using region-specific data.

Recent advancements in machine learning, particularly Artificial Neural Networks (ANNs), have significantly improved the accuracy of pavement performance prediction. Models such as Multi-Layer Perceptrons (MLP), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) have demonstrated strong capability in modeling nonlinear and time-dependent relationships. In addition, the inclusion of environmental variables such as temperature, rainfall, and solar radiation has been shown to enhance prediction reliability, especially in regions with harsh climatic conditions.

Despite these advancements, the review also highlighted key challenges, including limited availability of localized data, lack of model transparency, and insufficient application of advanced neural network techniques in developing countries. To address the issue of model interpretability, techniques such as SHAP and LIME have been introduced to improve transparency and support practical decision-making.

Overall, the literature indicates that while significant progress has been made in pavement performance prediction, there remains a need for localized, accurate, and interpretable models tailored to specific environmental and data conditions.

2.8.2 Research Gaps

Although numerous studies have been conducted on pavement performance prediction using traditional and machine learning approaches, several important gaps remain, particularly in the context of developing countries such as Ethiopia.

Geographical Gap: Although pavement performance prediction studies have been conducted in both developed and developing countries, there is very limited published research on the application of machine learning techniques for pavement performance prediction using Ethiopian urban road network data, particularly under the climatic, traffic, and pavement conditions of Mekelle City. Many existing prediction models were developed using datasets collected under

different environmental and operational conditions, which may limit their direct applicability to the Ethiopian context. Consequently, there remains a need to develop and evaluate prediction models using locally representative pavement, traffic, and environmental data.

Methodological Gap: While Artificial Neural Networks (ANNs) have been widely applied, although advanced time-series models such as LSTM and GRU have shown strong performance in previous studies, their application in pavement engineering remains limited in developing countries. Furthermore, very few studies integrate these advanced models with interpretability techniques such as SHAP and LIME to enhance model transparency.

Data and Contextual Gap: Many previous studies rely heavily on generalized datasets and do not adequately incorporate localized environmental factors such as temperature variation, rainfall patterns, and solar radiation. In semi-arid regions like Mekelle, these factors play a critical role in pavement deterioration, yet they are often underrepresented in predictive models.

Practical Gap: Despite the high predictive accuracy of machine learning models, their practical application in engineering decision-making is limited due to the “black box” nature of these models. There is a lack of studies that focus on making these models interpretable and usable for local engineers and policymakers.

Content Gap: Previous research often focuses on material performance, short-term condition assessment, or specific indicators such as IRI, rather than providing a comprehensive long-term pavement performance prediction framework. Additionally, comparative analysis between traditional models and advanced machine learning techniques is still limited in the context of developing regions.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents the research methodology employed to achieve the objectives of the study. It outlines the procedures used to develop a predictive model for long-term pavement performance in Mekelle using Artificial Neural Networks (ANN). The chapter describes the research design, study area, population and sampling techniques, data sources and collection methods, data preprocessing and feature engineering, model development and training, evaluation techniques, and ethical considerations. As defined by(Groenewald, 2004), research methodology is the systematic way to solve a research problem; This methodology provides a systematic framework to ensure the validity, reliability, and reproducibility of the research findings.

3.2 Study Area

The geographical focus of this research is Mekelle, the capital city of the Tigray Region in northern Ethiopia. Located at approximately 13°29′N latitude and 39°28′E longitude, with an elevation of around 2084 meters, the city's climate is classified as tropical savanna (Aw) with semi-arid (Bsk) characteristics. This environment presents unique challenges for pavement durability, including high diurnal temperature variations, intense solar radiation, and a distinct wet season concentrated in July and August [(Wilson & Partners, 2020)]. These specific climatic conditions, coupled with rapid urbanization and increasing traffic loads, make Mekelle an ideal and critical case study for developing a regionally-calibrated pavement performance prediction model.

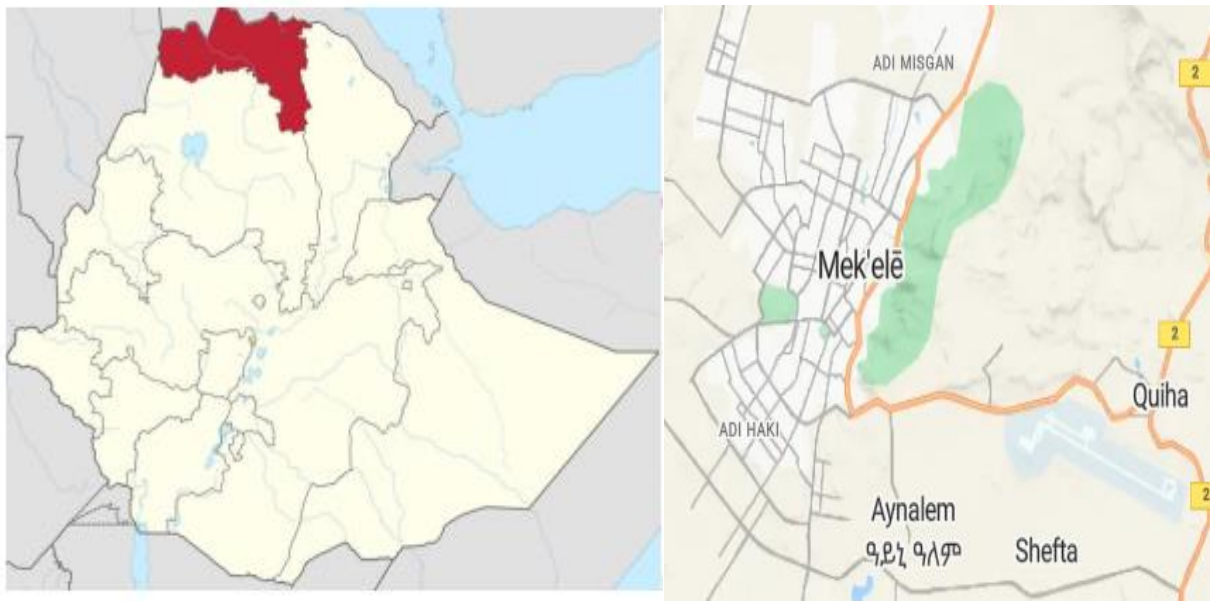


Figure 3.1 Use Source Google map

3.3 Research Design and Approach

The most important problem after defining the research problem or statement of problem is preparing the design of the research project, which is properly known as the research design. A research design is the setup of parameters for data collection and analysis with the goal of balancing procedural economy with relevance to the research question. The research design, which serves as the guide for data collection, measurement, and analysis, is actually the conceptual framework within which research is carried out. Consequently, the design comprises a summary of the researcher's steps, from formulating the hypothesis and considering its operational implications to doing the last data analysis (KOTHARI, n.d.).

The research process consists of sequential stages required to conduct the study systematically and efficiently(KOTHARI, n.d.). Information processing and data collection are thus important to the research process. Individuals can use the knowledge they produce to expand their horizons and possibly even acquire wisdom.

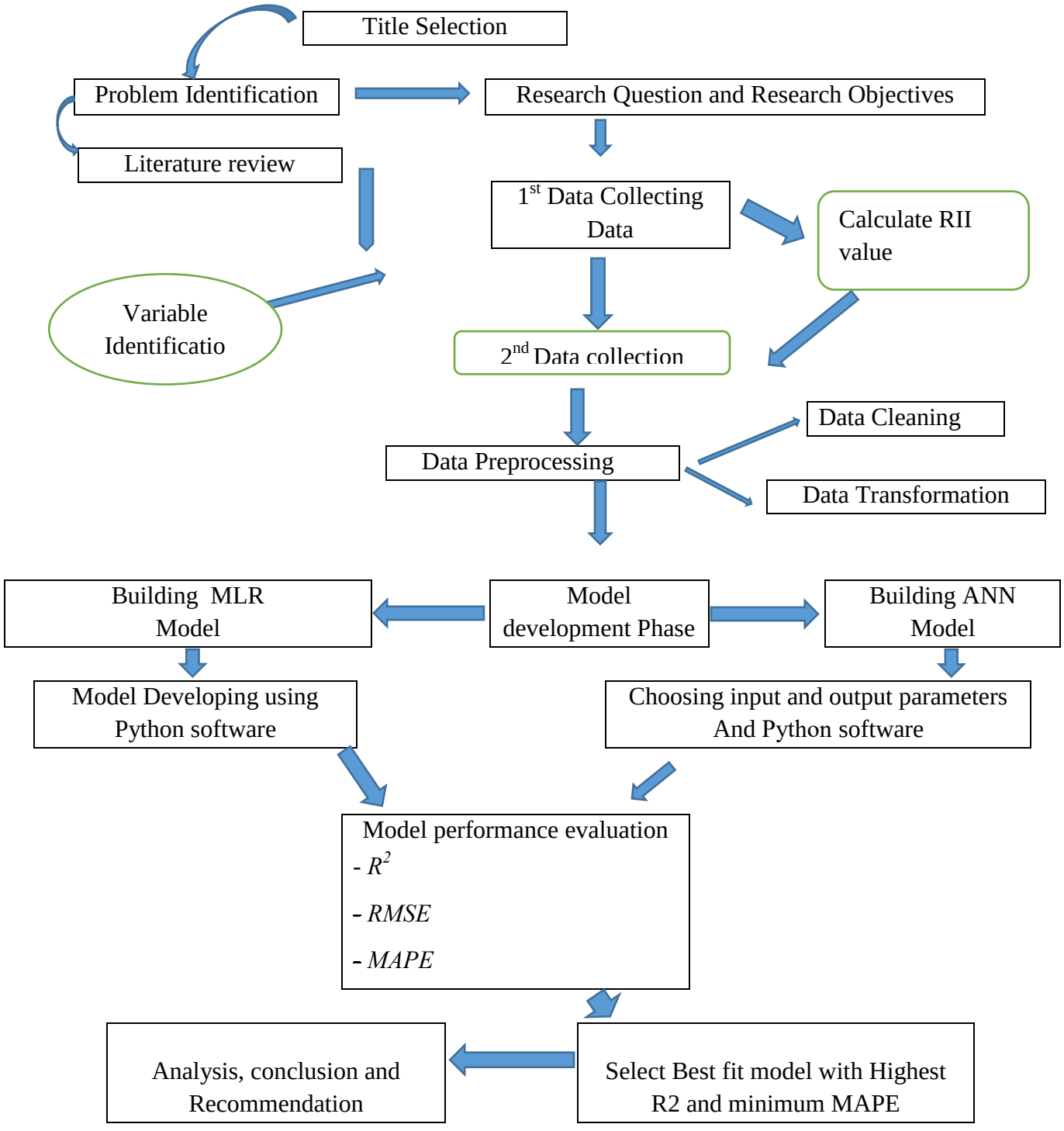


Figure 3.2: Research process adopted for this Study

As shown in the Figure 3.2 the research design used in this study is as follows: Potential input variables were identified from literature review, historical pavement datasets, and engineering standards, while the RII analysis was used only as a supplementary expert-based assessment to compare practitioners' perceptions of the relative importance of pavement deterioration factors with the data-driven findings obtained from the machine learning models. The RII analysis was not used to validate the predictive accuracy of the ANN model, which was evaluated independently using the testing dataset and performance metrics. Before model development outliers were eliminated from data sets to minimize effect of outliers on prediction ability of developed models. Afterwards, both statistical and machine learning tool were used to develop the models and their process were discussed under the method of data analysis section.

This study adopted a quantitative, predictive, and computational research design. This approach is fundamentally rooted in the positivist research philosophy, which posits that phenomena can be observed, measured, and explained through empirical evidence and logical analysis. The choice of a quantitative design is justified by the study's objective to develop a data-driven model that can forecast numerical pavement performance indicators. The predictive nature of the design focuses on leveraging historical data to make informed forecasts about future conditions, which is central to the research problem.

The core of the methodology involves a computational approach centered on the development of Artificial Neural Network (ANN) models. This design integrates several key phases: systematic data collection, careful data preprocessing and feature engineering, iterative model development and optimization, and reliable validation through benchmarking and interpretability analysis. In addition to the computational modeling framework, a supplementary expert-based assessment was conducted using questionnaire surveys and Relative Importance Index (RII) analysis. This component was intended to capture the perceptions and practical experiences of pavement professionals regarding the major factors influencing pavement deterioration in Mekelle City. The findings from the RII analysis were compared with the feature importance results obtained from ANN interpretability techniques such as SHAP and LIME to assess the consistency between expert judgment and data-driven model outcomes.

To compare the predictive performance of the specific objective four based on the data from 30 respondents and for the rest of other objective hypotheses were tested on 10 road pavement projects completed in the past years (2010–2025).

3.4 Population, Sample Size, and Sampling Technique

3.4.1 Population of the Study

The population of a study is defined as a group of subjects from whom the researcher expects to draw conclusions about the study(Groenewald, 2004). The population considered for this study comprises pavement engineers, transport engineers, consultants, road authority experts, and project managers involved in road infrastructure management in Ethiopian, Mekelle City transport and Construction Bureau, sur construction and The population of the study covered all listed pavement road segments within Mekelle that have: Recorded historical performance data, documented environmental conditions (e.g., temperature, rainfall and material property).

3.4.2 Sample Size Determination

Determining the sample size involves selecting the number of observers or replicates to include in the statistical sample, and it is a crucial stage in research methodology (Community & Nursing, 2017).The sample size of this study respondent from the Mekelle City Road Authority and the Tigray Construction, Road and Transport Bureau, Mekelle Traffic Police and municipal records, National Meteorology Agency's Mekelle branch, Sur construction plc. Consultants, Office Engineers, project Managers, site engineers.

There are many formulas to determine sample size to draw data from a given population. However, in this study, a total of 30 respondents for questionnaire were used because on account of the population being considerably small and on the merit of the increase in the accuracy of the developed model based on purposive sampling. Recently completed projects would be suitable for modelling due to their similarity in road project practice and uniformity in terms of technological utilization. In the case of road projects this study used all completed projects between the periods of 2010 to 2025 due to limited number of projects completed within the period and scarcity of data. And also the target sample size consisted of approximately 10 distinct road segments, with data spanning a historical period of 2 to 15 years. This yielded a

dataset of approximately 91 rows and 3,185 longitudinal records, which is considered a sufficient volume for effectively training, validating, and testing complex neural network models while mitigating the risk of over fitting.

3.4.3 Sampling Technique

Sampling is the process of selecting a subset from a given sampling frame or population as a whole and it can be used to draw conclusions about a community or to draw broad conclusions about current theories (Taherdoost, 2022). A purposive sampling technique was employed for this purpose. This non-probability sampling method is justified because the primary goal is not to generalize findings to a larger population in a statistical sense, but rather to select road segments that provide the highest quality and most relevant data for training a predictive model. The criteria for inclusion in the sample was included the availability of complete and consistent historical pavement condition data, documented environmental and traffic data, and a sufficient variation in performance levels, ranging from good to poor.

3.5 Data Sources and Collection

3.5.1 Data Sources

The sources of data for this study were primary and secondary data. The source of data for this study were Pavement condition data, including key performance, historical condition surveys, and inspection reports from the Mekelle City Road Authority and the Tigray Construction, Road and Transport Bureau. Traffic data, including Average Daily Traffic (ADT) and vehicle classification, was obtained from the Mekelle Traffic Police and municipal records.

Crucially, environmental data was sourced from the National Meteorology Agency's Mekelle branch and potentially augmented with satellite data providers. This included daily time-series records of temperature rainfall, solar radiation, and humidity over the past 2 to 15 years. Additional data related to construction and materials, such as pavement type and structural layer properties, were collected from relevant municipal and construction bureaus and Sur construction plc. The output Pavement Condition Index (PCI) was collected from standard PCI scale (ASTM D6433/ Ethiopia concept) To provide a benchmark for model development and feature selection,

data from the Long-Term Pavement Performance (LTPP) database was also be consulted, focusing on pavement types and climatic conditions that bear similarities to those in Mekelle.

The sources of secondary data for this thesis paper were record documents and journal publications.

3.5.2 Methods of Data Collection

To achieve the research objectives different data collection procedures were adopted. Data for the research were collected using both primary and secondary sources (Mazhar et al., 2021). Primary data sources was used in order to get firsthand information from the Road projects and it helped in providing information for specific purpose of addressing the problem. This study collected data from several primary and secondary sources through access historical database, field surveys, interview and questionnaires.

3.5.2.1 Questionnaire Survey

In this research study data was gathered through Questionnaire survey and by direct observation. Questionnaires were developed and distributed to pavement engineers, project managers, consultants, and transport authority professionals with experience in road construction and maintenance. The responses were analyzed using the Relative Importance Index (RII) method to rank the perceived significance of deterioration factors. These rankings were later compared with the feature importance results generated from ANN interpretability techniques such as SHAP and LIME. Close-ended questions were designed using a five-point Likert scale. Scale of agreement followed for Likert scale questions in this study were:

1 = Strongly Disagree

2 = Disagree

3 = Moderately Agree

4 = Agree

5 = Strongly Agree

3.5.2.2 Interview

A limited number of expert interviews were conducted to support understanding of local pavement deterioration practices and data availability, and critical observation of the planning activity was also done in order to obtain reliable information about the factors affecting pavement deterioration and prediction accuracy.

3.5.2.3 Case Study

Secondary data was gathered through reviewing, examination of documents, reports, and records of published sources and other pieces of literature related to factors or parameters that can help to predict performance of road project. The data collected from stated source were checked whether it matches with consultant for some selected projects for reliability of the data before data analysis. A thorough literature review was carried out to assemble the theoretical framework for the study by reviewing similar works done and establish the parameters to be used as inputs in the development of the model. Furthermore, archival analysis of project records was carried out at relevant government offices to gather data on the parameters to be used in the development of the model.

3.6 Data Preprocessing and Feature Engineering

3.6.1 Data Preprocessing

Raw data collected from various sources are often inconsistent and not suitable for direct input into a neural network. Therefore, a rigorous data preprocessing and feature engineering phase was undertaken. The preprocessing phase began with data cleaning to address missing values through appropriate imputation methods, such as mean, median, or model-based techniques, and to correct inconsistencies or remove outliers identified through statistical analysis. Following this, data normalization was performed using Min-Max scaling.

And a common Normalization formula is:

$$x' = \frac{x - xmin}{Xmax - xmin} \qquad \text{----------Eq 3.1}$$

Where:

x = original value

x' = normalized value

This step is essential as it ensures all input features are on a comparable scale, which prevents variables with larger magnitudes from disproportionately influencing the model and facilitates more efficient training.

3.6.2 Feature Engineering

Feature engineering was conducted to improve the quality and predictive capability of the dataset used in model development. This process involved selecting the most relevant variables through correlation analysis and transforming raw data into more meaningful inputs.

In addition, data transformation techniques were applied to reduce inconsistencies and improve model performance. New derived variables, such as climatic indices (e.g., seasonal average temperature and cumulative rainfall over specific time periods), were created to better represent environmental effects on pavement deterioration.

3.7 Model Development and Training

3.7.1 Data Analysis Techniques

Appropriate data analysis techniques were employed to process and interpret the collected data effectively. Data analysis involves inspecting, cleaning, transforming, and modeling data to obtain meaningful information, draw conclusions, and support decision-making. In this study, Microsoft Excel was used to calculate the Relative Importance Index (RII) and analyze questionnaire data. Python was used for data preprocessing, feature engineering, statistical analysis, and the development of predictive models using Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN).

3.7.2 Relative Importance Index (RII)

Relative importance index (RII) methods are used for quantitative data analysis once the data collected from various sources has been compiled. The most important factors can be identified as those having the largest values of final index (Ortega et al., 2025). All factors have been

expressed as a ratio of the most important factor, with weights greater than 70% considered to represent the most important factors (Ortega et al., 2025).

The Relative Importance Index (RII) method was employed as a supplementary expert-based analysis tool to rank the perceived influence of factors affecting pavement deterioration in Mekelle City. Unlike the ANN models, which identify influential variables through data-driven learning and interpretability analysis, the RII method captures the practical experience and judgment of professionals involved in pavement construction and maintenance. The results obtained from the RII analysis were used for comparative validation with the feature importance rankings generated from SHAP and LIME techniques.

The Relative Importance Index (RII) formula is: $RII = \frac{\sum W}{A*N}$ ---------- Eq 3.2

Where $\sum W$ = Sum of the weights given by all respondents

A = Highest weight on the scale from a 5 point Likert

N = Total number of respondents

3.7.3 Linear Regression Model

Regression analysis is a technique that finds a formula or mathematical model which best describes a set of data collected (Neuman et al., 2007). Moreover, Linear regression is an approach to model the relationship between a scalar dependent variable Y and one or more explanatory variables denoted x. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, it is called multiple linear regressions. In linear regression data are modeled using linear predictor functions and unknown model parameters are estimated from the data.

3.7.3.1 Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) is a statistical method used to analyze the relationship between one dependent variable and two or more independent variables. In this study, MLR was employed to predict Pavement Condition Index (PCI) values using property pavement, traffic, and environmental factors as predictor variables.

The general form of the MLR model is expressed as follows (Neuman et al., 2007).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n$$ ---------- Eq. 3.3

Where:

Y = dependent variable (Pavement Condition Index, PCI)

β_0 = intercept (constant term)

β_1 to β_n = regression coefficients of the independent variables

X_1 to X_n = independent variables (predictor variables)

n = number of independent variables included in the model

The MLR model estimates the contribution of each independent variable to the prediction of pavement performance and provides a baseline model for comparison with the Artificial Neural Network (ANN) model.

3.7.4 Model Development

The core of the analytical process is the development and optimization of the neural network models. The core analytical process focused on the development of a Multilayer Perceptron (MLP)-based Artificial Neural Network (ANN) model and comparing its performance with a Multiple Linear Regression (MLR) model for Pavement Condition Index (PCI) prediction. A Multilayer Perceptron (MLP) served as a baseline model to capture static, non-temporal relationships. The core analytical process focused on developing an Artificial Neural Network (ANN) model and comparing its performance with a Multiple Linear Regression (MLR) model for Pavement Condition Index (PCI) prediction. The input variables for these models included pavement age, structural characteristics, traffic loads, environmental factors, and records, while the output variable was the predicted Pavement Condition Index (PCI) over a 2-15 year.

3.7.4.1 Artificial Neural Network (ANN)

The ANN model development process consisted of three major stages: data preparation, model development, and model evaluation.

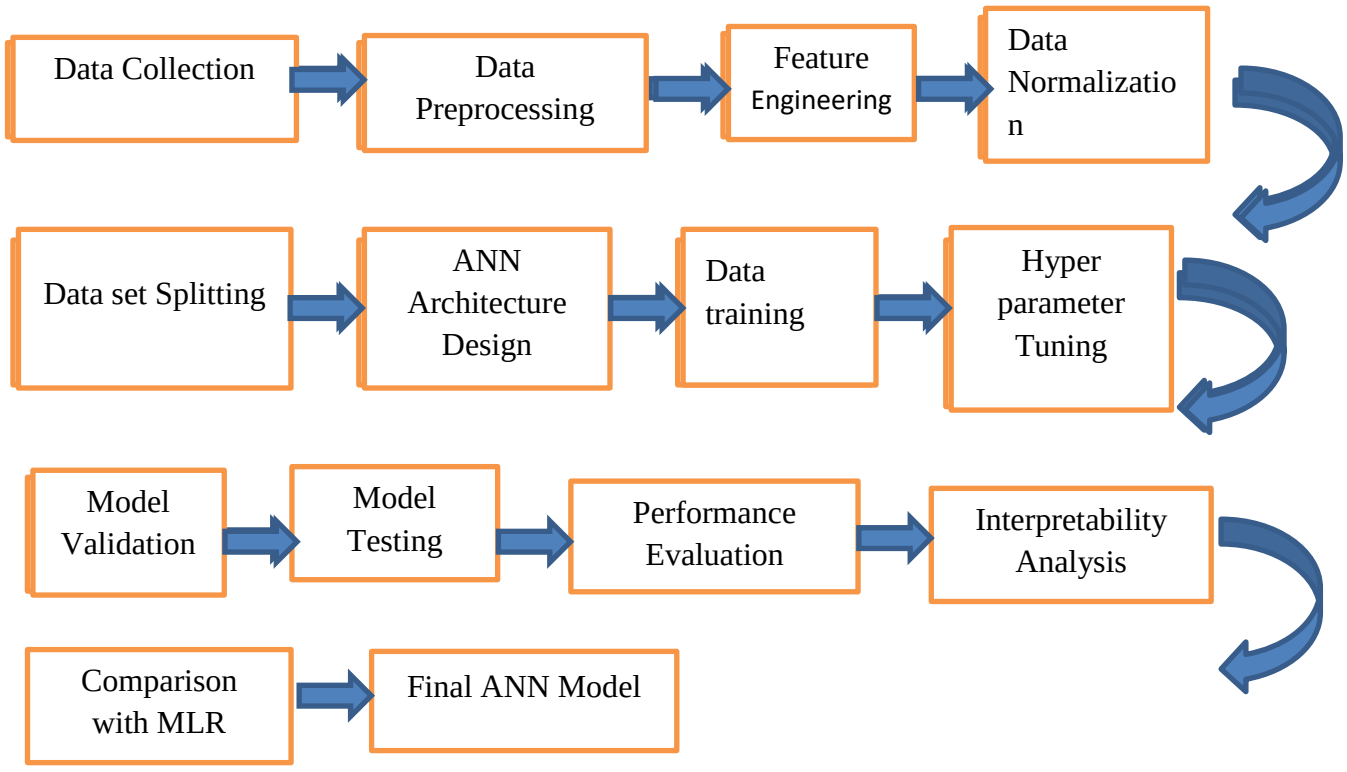


Figure 3.3: ANN development flow chart.

The goal of this research is to develop an Artificial Neural Network (ANN) model for forecasting long-term pavement performance.

The ANN mathematical expression:

$$y = f\left(\sum_{i=1}^{n} wixi+b\right)$$ ---------- Eq.3.4

Then explain:

- xi = input variables
- wi = connection weights

- b = bias term
- f = activation function
- y = neuron output

3.7.5 Model Training

The dataset was randomly partitioned into three mutually exclusive subsets: 70% for training, 15% for validation, and 15% for testing. Because the observations represent independent pavement sections rather than chronological measurements, a random data partitioning strategy was adopted instead of walk-forward (time-series) validation.

The training subset was used to estimate the ANN model parameters. The validation subset was used during model development to optimize the network hyper parameters and implement early stopping to reduce over fitting. The testing subset was kept completely independent from the training process and was used only once to provide an unbiased evaluation of the final model's predictive performance. The network weights were optimized using the Adam optimization algorithm with the Mean Squared Error (MSE) loss function. This training-validation-testing strategy improved the model's generalization capability while minimizing the risk of over fitting.

3.8 Model Evaluation and Interpretability

3.8.1 Model Evaluation

A multi-faceted approach was used to evaluate the performance of the developed models. The predictive accuracy was quantitatively assessed using standard regression metrics, including the Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The performance of the neural network models was rigorously benchmarked against that of a traditional multiple linear regression model to demonstrate their added value.

3.8.2 Model Interpretability

In addition to predictive accuracy, understanding how the model makes its predictions was an important part of this study. Since Artificial Neural Networks (ANNs) do not clearly show how input variables influence the output, SHapley Additive exPlanations (SHAP) and Local

Interpretable Model-agnostic Explanations (LIME) were used to explain the model results. These techniques helped identify and measure the contribution of each input variable to the predicted pavement condition. The analysis made it possible to determine the factors that have the greatest effect on pavement deterioration in Mekelle City, providing useful information for engineers and decision-makers. The feature importance results obtained from SHAP and LIME were also compared with the rankings from the Relative Importance Index (RII) analysis to examine the level of agreement between expert judgments and the results produced by the ANN model.

3.9 Software

The data analysis and model development processes were carried out using Python and its open-source libraries. Data preprocessing, cleaning, and statistical analysis were performed using Pandas and NumPy. Multiple Linear Regression (MLR) and other baseline analyses were implemented using the Scikit-learn library. The Artificial Neural Network (ANN) model was developed and trained using TensorFlow/Keras was used to develop, train, validate, and test the Artificial Neural Network (ANN) models. TensorFlow served as the underlying deep learning framework, while Keras provided the high-level API for designing and implementing the neural network architecture. To ensure the reproducibility of the developed models, a fixed random seed value **(**random state = 42**)** was used during data splitting and model training procedures. The use of a fixed random seed ensures that the same training and testing datasets are generated in repeated executions of the code, thereby improving the consistency and reproducibility of the research findings. Model interpretability was assessed using the SHAP and LIME libraries to identify the contribution of input variables to pavement condition predictions. Data visualization and graphical presentations were created using Matplotlib, Seaborn, and Plotly. In addition, Microsoft Excel was used for data organization and the calculation of the Relative Importance Index (RII).

3.10 Reliability and Validity of the Research

Reliability and validity are concepts used to evaluate the quality of research. As defined by (Mugenda & Mugenda, 2003), Reliability and validity are concepts used to evaluate the quality of research. They indicate how well a method, technique, or test measure something. Reliability is about the consistency of a measure, and validity is about the accuracy. It’s important to

consider reliability and validity when the researcher creates research design, planning methods, and writing up results. To establish content validity, the data collection instruments and the selected dataset features were reviewed by experts from the Mekelle City Road Authority and other relevant institutions. Their feedback was used to refine the data collection plan and ensure the variables are relevant to the local context. Where applicable, the Content Validity Index was calculated. The reliability of any manually collected or survey-based data was assessed using internal consistency measures such as Cronbach's alpha to ensure the data is dependable.

3.10.1. Reliability analysis

Reliability analysis was conducted to assess the internal consistency of the questionnaire items used in this study. Cronbach's Alpha coefficient was calculated using IBM SPSS Statistics Version 23 for the three groups of factors affecting pavement performance prediction. Cronbach's Alpha is one of the most commonly used measures of internal consistency and is widely accepted for evaluating the reliability of Likert-scale questionnaires.

Table 3.1 Reliability Statistics for Objective (Quantitative) factors

Cronbach's Alpha	N of Items
.72	12

Table 3.2 Reliability Statistics for Subjective (Qualitative) factors

Cronbach's Alpha	N of Items
.84	5

Table 3.3 Reliability Statistics for other factors

Cronbach's Alpha	N of Items
.79	13

In current study, a Cronbach alpha of 0.72 for objective factors, 0.84 for subjective factors, and 0.79 for other factors influencing pavement performance prediction. Generally, Cronbach’s

Alpha values greater than 0.70 are considered acceptable, suggesting that the questionnaire items are sufficiently correlated and reliable for further statistical analysis.

3.10.2. Validity analysis

Validity refers to the extent to which a research instrument measures what it is intended to measure and ensures the accuracy and credibility of the collected data. To enhance the validity of this study, the questionnaire was reviewed by experts in the field before distribution. Their comments and recommendations were incorporated into the final version of the questionnaire.

In addition, methodological triangulation was employed by collecting data through questionnaires, interviews, case studies, and literature reviews. The use of multiple data sources helped to verify the consistency of the findings and improved the overall validity of the research. Therefore, the data collection instruments and research findings were considered valid and appropriate for achieving the objectives of the study.

3.10.3 Model Validation

This research used SPSS for quantitative hypothesis testing by using regression techniques and also the scatter and residual plot techniques**.** To choose the best out of the models generated in this study, the following assessment criteria were used and analyzed using SPSS software 23 package. The model development process should always evaluate the performance of the model before using for the intended purpose. The purpose of model evaluation is to ascertain the performance of the model. Model validation is an important step in developing regression and ANN relationship. The model performance measured using.

(A) Correlation Coefficient (R)

The coefficient of correlation, r, identifies the degree and nature of the relationship between the two variables, from a perfect positive relationship (+1) to a perfect negative relationship (−1), that is $-1 \leq r \leq +1$. The coefficient implies, $r = +1$ means that an increase in variable x is matched

by proportional increase in y. If r = 0, there is no relationship; changes in the variables are quite independent of each other; they are random(Silva, n.d.).

(B)Coefficient of Determination(R^2)

The coefficient of determination **(** R^2 **)**, is statistic required to measure how much of the change in the values of a dependent variable is caused, given the logic of the relationship, by a change in the values of the independent variable(Silva, n.d.).

The formula for calculating (R^2) is:

$$R^2 = \frac{\text{Regression sum of squares}}{\text{Total sum of squares}} \quad \text{Eq.3.5}$$

The values also range between zero and one ($0 \leq R2 \leq 1$). A high value of R^2 indicates a good model. R^2 Gives an indication of the predictive capability of the regression model using new observations. It is usually expressed as a percentage by multiplying by 100.

(C) Mean Absolute Percentage Error (MAPE)

MAPE is a measure of difference between two continuous variables and accuracy of a method for constructing fitted times series values in statistics. It usually expresses accuracy as a percentage. Two measures of accuracy dealing with percentage error were used to compare the forecasting performance of the model(Silva, n.d.).

$$\text{Percentage Error} = \frac{\text{Predicted PCI} - \text{Actual PCI}}{\text{Actual PCI}} \times 100\%$$

Mean Absolute Percentage Error (MAPE) = Σ PE…..... Eq.3.6

(D) Root Mean Squared Error (RMSE)

RMSE is the root square of the average squared distance between the actual and predicted observations. It is one of the crucial metric that is used on the models, especially when it is a regression model. Since the errors are squared, the metric provides high weight to larger residuals (different between the actual and predicted value) than the smaller residuals(Silva, n.d.).

$$\text{Root Mean Squared Error (RMSE)} = \frac{\sqrt{1\ \sum_{i=1}^{n}(Predictedi - Actuali)2}}{2} \quad \text{.....Eq.3.7}$$

Where

- RMSE = Root Mean Squared Error
- n = Number of observations
- Actual*i* = Actual pavement condition index (PCI) value for observation *i*
- Predicted*i* = Predicted PCI value for observation *i*

3.11 Ethical Considerations

Ethical considerations were observed throughout this study. Ethical clearance was obtained from the FDRE Technical and Vocational Training Institute, Faculty of Civil Technology, Department of Construction Management and Technology, and official letters were issued by the Department of Construction. The purpose of the study was clearly explained to all participants, and their participation was based on informed consent. Since the study was conducted solely for academic purposes, no conflicts of interest existed between the researcher and project stakeholders. Data were collected voluntarily, kept confidential and anonymous, protected during interviews, used only for research purposes, and all sources were properly acknowledged to avoid plagiarism.

CHAPTER 4

DATA ANALYSIS AND DISCUSSION

4.1 Background Information about Respondents

This sub- section was presented the results and outputs generated by the questionnaire survey. Respondents included the Tigray Construction Road and Transport Bureau and sure Construction Bureau Consultants, Transport Engineers, project Managers and Pavement engineers.

4.1.1 Rate of return of questionnaire

Table 4.1 Respondent rate of return of questionnaire

Profession of respondent	Distributed Questioner	Returned valid Questioner	Returned valid %
Consultant	8	7	87.5
Transport Engineer	8	7	87.5
project Manager	4	3	75
Pavement Engineer	15	13	86.67
Total	35	30	85.7

In case of this study respondents are grouped in to four, which means consultants, project managers, office Engineer and Site Engineer. Table 4.1 shows a total of 35 questionnaires were distributed to respondents. Of these 8 were to consultants, 4 were to project managers and 8 were to Transport Engineer and 15 to Pavement Engineer. Among those distributed questionnaires 30 were returned. Of these 7 were from consultants, 3 were from project managers, 7 were from Transport Engineer and 13 were from Pavement Engineer. Finally, this representing out of total response 85.71% is valid response. Therefore, the response rate is considered as good response. The researcher succeeded in gathering well trusted thirty responses from the Tigray

Construction, Road and Transport Bureau and sure Construction Bureau professional. Questionnaire and hard copy was used to provide the respondents with easy access.

4.1.2 Working experience of respondent

The Figure 4.1 provides the Working experience of respondent. The Working experience ranges from 0 to above 15 years. Participant's level of working experience is another factor which makes the results more reliable.

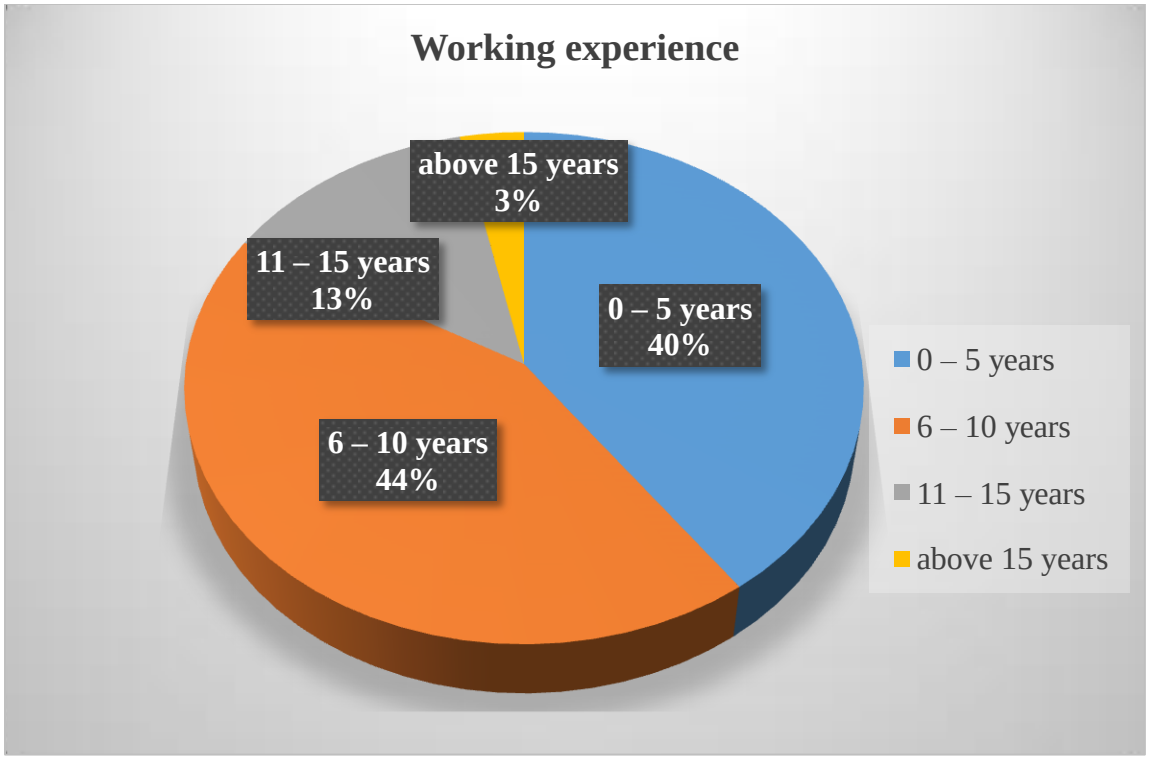


Figure 4.1: Respondents' working experience

Experience of respondents from zero years to more than fifteen years. Figure 4.1 illustrate work experience of respondents 40% of respondents have 0 to 5 years, 44%, 6 to 10 years, 13%; 11 to 15 years and the 3% have more than 15 years on related working experience. It would be better the percentage of respondents whose experiences were 5 years or more could be increased. The result indicated that the majority of respondents work experience in project was between 5-10 years. This implies the majority of participants for the survey were well experienced and

professionals in the performance prediction. This implies that the majority of participants were experienced professionals in pavement performance prediction and road infrastructure management. This shows that the results obtained was valid.

4.1.3 Education level of respondent

While doing reliable research, the educational qualification of respondents plays a very significant role in determining the quality of the data being analyzed. In Figure 4.2 provides the educational background of the respondents. The educational background ranges from Diploma to PHD. Participant's level of academic studies is another factor which makes the results more reliable.

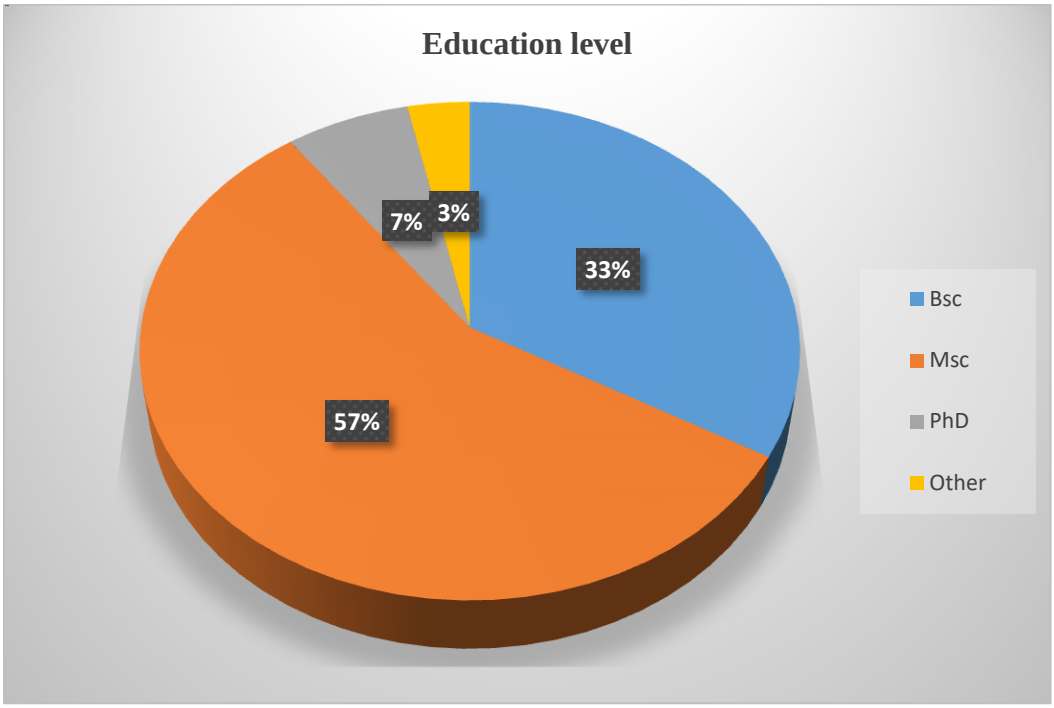


Figure 4.2: Respondents' level of education.

The questionnaire results showed that the majority of the respondents (57%), had the master degree; while 33 % had bachelor degree, 3% had diploma and around 7 % holders are PHD. The result showed that the majority of respondent's educational backgrounds were master degree. This is also other opportunities to improve performance prediction in road construction.

4.2 Major Significant Factors Affect Performance Prediction Road Projects

The evaluation and identification of factors affecting pavement performance prediction is a critical issue in pavement management is a critical issue facing the engineering estimator. Understanding the critical factors that affect the estimation of the Performance, whether positive or negative, can contribute to the development of a strategy to reduce the shortcomings and improve the effectiveness of the performance of the road project. Data was collected from different stakeholders they actively participating in road projects in Mekelle. The researcher succeeded in gathering well trusted data about from respondents. The sample has been directed to 30 experts in this field. I asked to pinpoint the foremost important factors influencing in predicating performance of the road.

Table 4.2 Relative importance index for Objective (Quantitative) factors

No	Objective (Quantitative) factors	SDA	DA	MA	A	SA	Total	Weight	RII %	Rank
		1	2	3	4	5				
1	Average daily traffic	0	3	5	10	12	30	121	81	7
2	Load repetitions (number of vehicles over a specified time)	0	2	5	10	13	30	124	83	6
3	Compressive strength of concrete	0	1	3	8	18	30	133	89	3
4	Aggregate gradation	0	0	3	8	19	30	136	91	2
5	Average temperature	5	5	5	7	8	30	98	65	10
6	Annual precipitation	7	6	4	6	7	30	90	60	11
7	Freeze-thaw cycles per year	6	6	7	5	6	30	89	59	12
8	Layer thickness	0	0	0	10	20	30	140	93	1
9	Pavement design life	0	0	5	8	17	30	132	88	4
10	Surface type	0	4	8	8	10	30	114	76	8
11	Maintenance history	1	5	7	8	9	30	109	73	9
12	Geotechnical factors	0	2	5	9	14	30	125	83	5

In Table 4.2 shows major objective (quantitative) factors affecting performance prediction in road projects. Respondents ranked “Layer thickness” in the first place with a relative importance index of 93%. The second causes which was ranked by respondents was “Aggregate gradation” with RII = 91%. It is an important hint that the Aggregate gradation for the proposed project leads to significant impact on performance prediction. In the 3rd rank, “Compressive strength of concrete” with RII = 89% of project were major factors affecting performance prediction in road projects .The most important factors can be identified as those having the largest values of final index(Ortega et al., 2025). All factors have been expressed as a ratio of the most important factor, with weights greater than 70% considered to represent the most important factors.

Table 4.3 Relative importance index for Subjective (Qualitative) factors

No	Subjective (Qualitative) factors	SDA	DA	MA	A	SA	Total	Weight	RII %	Rank
		1	2	3	4	5				
1	Quality of workmanship	0	0	4	10	16	30	132	88	1
2	Impact of driving habits on pavement wear	4	3	3	10	10	30	109	73	4
3	Attitudes toward proactive vs reactive maintenance	5	3	3	9	10	30	106	71	5
4	Public satisfaction with road conditions	2	2	4	10	12	30	118	79	3
5	Perceptions of safety and usability of the pavement	0	3	4	12	11	30	121	81	2

Subjective judgment, were applied in Table 4.3 summarizes the influential factors related to the computation of construction performance for pavement roads in Mekelle derived from this study. Table 4.3 shows Subjective (Qualitative) factors affecting performance prediction in pavement road. Respondents ranked “Quality of workmanship” in the first place with a relative importance index of 88%. The second causes which was ranked by respondents was “Perceptions of safety and usability of the pavement” with RII = 81%. In the 3rd rank, “Public satisfaction with road conditions” with RII = 79% the above all indicates that major subjective (qualitative) factors affecting performance prediction in pavement road. The most important factors can be identified

as those having the largest values of final index. All factors have been expressed as a ratio of the most important factor, with weights greater than 70% considered to represent the most important factors (Ortega et al., 2025).

Table 4.4 Relative importance index for other factors

No	Other factors influencing long-term pavement performance prediction	SDA 1	DA 2	MA 3	A 4	SA 5	Total	Weight	RII %	Rank
1	Administrative factors	0	1	1	8	20	30	137	91	1
2	Political factors and Government instructions	1	1	3	8	17	30	129	86	3
3	Operational factors	2	2	5	7	14	30	119	79	6
4	The Planning Department seeks to develop and modernize the planning process and measure their success	2	4	6	7	11	30	111	74	9
5	The extent to which modern documentation system to gather information and data on other projects and implemented.	3	4	6	9	8	30	105	70	10
6	Provide a modern information management system data are updated continuously.	3	5	6	8	8	30	103	69	11
7	Provide information management system allows easy communication between different administrative levels for the project	3	5	7	8	7	30	101	67	12
8	Provide technical and	2	3	5	8	12	30	115	77	8

	administrative measures for effective control over the planning and execution of the work									
9	The possibility of material incentives for workers in management and planning	2	2	5	8	13	30	118	79	7
10	Engaging activities for employees to quality and continuous improvement processes	1	1	2	8	18	30	131	87	2
11	Involvement of employees in training courses and programmers for planning and programming.	1	2	3	9	15	30	125	83	5
12	The Department's ability to provide detailed plans	1	1	3	9	16	30	128	85	4
13	Raise the level of workers	4	5	7	8	6	30	97	65	13

Table 4.4 Other factors influencing long-term pavement performance prediction. That respondents ranked “Administrative factors affect the success of the project” in the first row with RII = 91%. It states the high significance factor and impact on predication performance of road project. “Engaging activities for employees to quality and continuous improvement processes” was put in the second degree of importance with RII of 87% by respondents. In the 3rd rank, “Political factors and Government instructions effect on the project success” with RII = 86% the above all indicates necessary factors affecting performance prediction in road project. The most important factors can be identified as those having the largest values of final index. All factors have been expressed as a ratio of the most important factor, with weights greater than 70% considered to represent the most important factors.

Based on this research, the independent variables of a construction performance prediction model for road project have been established. The variables include the size, complexity & scope of the road project, and whose performance data are easy to collect. In addition, the months of the start of road, which affect the climate and construction process and number of roads that affects the road size, are also considered.

4.3 Presentation of Data

The data used in this study were collected from multiple secondary sources, including the Mekelle City Road Authority, the Tigray Construction, Road and Transport Bureau, Mekelle Traffic Police, municipal records, the National Meteorology Agency, and Sur Construction PLC. The dataset covers ten road segments constructed between 2010 and 2025.the selection of potential factors influencing pavement performance was initially guided by a comprehensive literature review and pavement engineering principles. This process helped identify and prioritize the major factors affecting pavement deterioration.

However, only measurable and quantitatively available variables were used as input parameters for the Multiple Linear Regression (MLR) and Multilayer Perceptron (MLP) models. The final model inputs were selected based on data availability, statistical analysis (including correlation analysis), and engineering relevance. the output Pavement Condition Index (PCI) was collected from standard PCI scale (ASTM D6433/ Ethiopia concept).

Table 4.5: Description of Variables

No	List of Variables	Code	Unit	Type	Scale
1	Pavement Condition Index	PCI	-	dependent	Ratio
2	Traffic Volume	TV	AADT	Independent	Ratio
3	Traffic Load	TL	ESAL	Independent	Ratio
4	Temperature	TR	℃	Independent	Ratio
5	Rainfall	RF	mm	Independent	Ratio

6	Maximum Dry Density	DDmax	g/m^3	Independent	Ratio
7	Optimum Moisture Content	OMC	%	Independent	Ratio
8	Liquid Limit	LL	%	Independent	Ratio
9	Plastic Index	PI	%	Independent	Ratio
10	Californian Bearing Ratio	CBR	%	Independent	Ratio
11	Swelling	S	%	Independent	Ratio
12	Aggregate Crushing Value	ACV	%	Independent	Ratio
13	Ten Percent Fines Value	TPFV	KN	Independent	Ratio
14	Flakiness Index	FI	%	Independent	Ratio
15	Los Angeles Abrasion	LAA	%	Independent	Ratio
16	Apparent Specific Gravity	ASG	–	Independent	Ratio
17	Bulk Specific Gravity	BSG	–	Independent	Ratio
18	Bulk Specific Gravity (S.S.D basis)	SSD	–	Independent	Ratio
19	Water Absorption	WA	%	Independent	Ratio
20	Sodium Sulphate Soundness Loss	SSSL	%	Independent	Ratio
21	Sand Equivalent	SE	%	Independent	Ratio
22	Clay Lumps	CL	%	Independent	Ratio
23	Unit Weight of Concrete	UW	kg/m^3	Independent	Ratio
24	Concrete Slump	CS	mm	Independent	Ratio
25	Concrete Compressive Strength	CCS	Mpa	Independent	Ratio
26	Marshall Stability	MS	KN	Independent	Ratio
27	Flow	F	mm	Independent	Ratio
28	Theoretical Maximum Specific Gravity	GMM	–	Independent	Ratio
29	Bulk Specific gravity of Compacted Mix (Density of Compacted Mix)	Gmb	g/cm^3	Independent	Ratio

30	Bitumen Content	BC	%	Independent	Ratio
31	Air voids	Va	%	Independent	Ratio
32	Voids in Mineral Aggregate	VMA	%	Independent	Ratio
33	Voids Filled with Asphalt	VFA	%	Independent	Ratio
34	Thickness	Th	mm	Independent	Ratio
35	Age	Ag	yr	Independent	Ratio

No	Project Name	TV AADT	TL ESAL	TR ℃	RF mm	DDmax g/m3	OMC %	LL %
1	Road 20 (Adi -Ha Round about (Genet Gas Station) to Hamiday - Kumel- Bete-Hintset Road Project), 1.7km Worked in 2018	1800	2.5	23	21.6	1.93	13.3	40
2	Road 21 (United Steel Factory (Lachi) to Hamiday Bridge - Messebo Cement Factory Road Project), 4km Worked in 2019	1750	1.5	24	26.7	1.77	17.5	65
3	Road 22 (Infront of United Steel Factory (Lachi) to Felege Daero School Road Project), 2.5kmWorked in 2020	1700	1.6	26	25.5	1.67	19.5	68
4	Road 23 (Mekelle -Samre Road Project Beginning to Abune - Aregaw Lachi Church Road Project), 1.6kmWorked in 2020	1900	2.15	23	120.7	2.12	9.2	48
5	Road 16 (Tigray Regional State Square West Side Direction of Enda Rufael Park to Hamiday Bridge Road Project)Worked in 2023	1850	2.45	25	71.3	1.59	21	70
6	Road 17 (Tigray Regional State Square East Side Direction of Enda Rufael Park to Hamiday Bridge Road Project)Worked in	1750	2.67	26	25.5	1.66	15.8	57

	2020							
7	Yetebaberut-Djibruk-Mayhafti(0+000-4+640)Worked in 2023	1750	4.35	26	38.2	1.55	17.5	78
8	Road 18 (Hamiday Bridge to Messebo Cement Factory Road Project)Worked in 2024	2000	1.3	24	37.7	1.83	12.9	42
9	Road 19 (Hamiday at Kumel Honey Factory via Hamiday Bridge to Mekelle Abi-Adi Road Near Bete-Hintset Road Project)Worked in 2024	1600	3.85	23	81.2	1.27	6.4	33
10	Hawzien Adebabay - Endamariam - Sheria(0+000-1+120)Worked in 2024	2100	4.5	25	24.5	1.89	8.8	29.9

Table 4.6 Performance of Asphalt Road Project Mekelle City

4.4 Data Mining Techniques

Most studies have emphasized the importance of partitioning data to reflect differences in project context such as pavement type, traffic condition, environmental exposure, terrain characteristics, and project location. Although the selected projects had relatively similar project delivery systems and geographic conditions, this study considered partitioning the dataset based on the type of road pavement projects, including asphalt pavement sections and urban road segments in Mekelle City. The pavement performance data used in this study were collected from road sections monitored between 2010 to 2025, with historical records obtained from the Mekelle City Road Authority, Tigray Construction Road and Transport Bureau, and municipal road maintenance agencies and sure construction. Since the datasets span several years, it was necessary to preprocess and normalize the data to ensure consistency among pavement condition indicators, traffic loading records, and environmental variables.

4.4.1 Data Correlation

The correlation coefficient (r) ranges from –1 to +1, where values close to +1 indicate a strong positive relationship, values close to –1 indicate a strong negative relationship, and values near 0 indicate weak or no linear relationship (Sampaio et al., 2024). Correlation analysis was used to measure the strength of association between the independent variables and pavement performance indicators.

Table 4.7: Correlation Matrix before Variable Transformation

Variables	PCI	ADT	ESALs	Temp	Rainfall	Age	CBR	Thickness
PCI	1.000	-0.684	-0.741	-0.536	-0.492	-0.812	0.624	0.587
ADT	-0.684	1	0.944	0.521	0.446	0.703	-0.518	-0.472
ESALs	-0.741	0.944	1	0.578	0.503	0.781	-0.566	-0.514
Temp	-0.536	0.521	0.578	1	0.692	0.487	-0.381	-0.336
Rainfall	-0.492	0.446	0.503	0.692	1	0.452	-0.347	-0.301
Age	-0.812	0.703	0.781	0.487	0.452	1	-0.621	-0.582
CBR	0.624	-0.518	-0.566	-0.381	-0.347	-0.621	1	0.694
Thickness	0.587	-0.472	-0.514	-0.336	-0.301	-0.582	0.694	1

The correlation matrix shows that Pavement Condition Index (PCI) has significant negative correlations with ADT (r = –0.684), ESALs (r = –0.741), temperature (r = –0.536), rainfall (r = –0.492), and pavement age (r = –0.812), indicating that increases in traffic loading, environmental exposure, and pavement aging are associated with deterioration in pavement condition. Among these, pavement age showed the strongest negative relationship with PCI, suggesting that aging is a major contributor to pavement deterioration. Conversely, CBR (r = 0.624) and pavement thickness (r = 0.587) showed positive correlations with PCI, implying that stronger subgrade conditions and greater structural thickness improve pavement performance. Additionally, the very high correlation between ADT and ESALs (r = 0.944) indicates a strong relationship between the two variables, suggesting that they may contribute similar information to the regression model. Although ADT and ESALs were highly correlated, they represent different engineering characteristics. ADT measures traffic volume, whereas ESALs quantify the cumulative damaging effect of axle loads on pavement deterioration. Therefore, both variables were initially retained to evaluate their contribution to pavement performance prediction. Nevertheless, the observed high correlation suggests that future studies should assess multicollinearity using VIF and consider excluding one of the variables if necessary.

4.4.2 Data Transformation and Preprocessing

Data do not always come in a form that is immediately suitable for analysis. Therefore, data transformation and preprocessing were carried out before model development. Data transformation and preprocessing were performed to improve data quality, reduce skewness, stabilize variance, normalize variable scales, and enhance model convergence and predictive performance. In machine learning applications, particularly neural network modeling, and transformed datasets often provide better predictive accuracy than raw datasets(Osei et al., 2023).In this study, the pavement performance dataset was preprocessed using Python-based data analysis and machine learning libraries including Pandas, NumPy, Scikit-learn, TensorFlow, and Keras. The preprocessing stage involved handling missing values, normalization, logarithmic transformation, and feature scaling techniques. Min–Max normalization method was applied to reduce the effect of differences in variable scales among historical data materials, traffic, environmental, and pavement condition indicators. To avoid data leakage, scaling

parameters were fitted on the training dataset and subsequently applied to the testing dataset. Accordingly, selected variables were transformed into natural logarithmic form to improve the linear relationship among variables and reduce skewers in the dataset. The transformed data considerably improved the correlation among variables, as shown in Table 4.8. Furthermore, normalization improved the convergence and predictive performance of the developed neural network models. Table 4.8 shows that transforming the data into logarithmic and normalized forms resulted in a considerable increase in the coefficient of correlation among variables. Equivalent Single Axle Loads (ESALs), pavement age, and Average Daily Traffic (ADT) were found to be the independent variables with the highest correlation coefficients when compared with other explanatory variables for pavement deterioration prediction.

Table 4.8: Correlation Matrix after Variable Transformation

Variables	New_PCI	New_ADT	New_ESALs	New_Temp	New_Rain	New_Age	New_CBR	New_Thickness
New_PCI	1.000	-0.752	-0.804	-0.603	-0.551	-0.873	0.711	0.668
New_ADT	-0.752	1	0.951	0.593	0.518	0.764	-0.596	-0.548
New_ESALs	-0.804	0.951	1	0.641	0.562	0.832	-0.648	-0.601
New_Temp	-0.603	0.593	0.641	1	0.714	0.552	-0.436	-0.394
New_Rain	-0.551	0.518	0.562	0.714	1	0.508	-0.391	-0.347
New_Age	-0.873	0.764	0.832	0.552	0.508	1	-0.703	-0.662
New_CBR	0.711	-0.596	-0.648	-0.436	-0.391	-0.703	1	0.752
New Thickness	0.668	-0.548	-0.601	-0.394	-0.347	-0.662	0.752	1

After transformation, the correlation coefficients generally increased, indicating improved variable stability and stronger interpretable relationships among predictors and pavement

performance. The transformed dataset reduced skewness, improved feature scaling, and enhanced suitability for machine learning and regression modeling. Pavement age (r = –0.873), ESALs (r = –0.804), and ADT (r = –0.752) showed the strongest negative relationships with PCI, indicating that traffic loading and aging are major contributors to pavement deterioration. In contrast, CBR (r = 0.711) and pavement thickness (r = 0.668) showed positive relationships with PCI, suggesting improved structural performance. However, ADT and ESALs remained highly correlated (r = 0.951), indicating possible multicollinearity, which was particularly important during MLR interpretation.

4.5 Multiple Linear Regressions (MLR) for Pavement Performance Prediction

4.5.1 Assumptions of Multiple Linear Regression Analysis

Linear regression is an approach used to model the relationship between a scalar dependent variable Y and one or more explanatory variables denoted by X. The case of one explanatory variable is called simple linear regression, while for more than one explanatory variable it is called multiple linear regressions(Kleinbaum, n.d.). The correlation analysis showed a very high correlation between ADT and ESALs (r = 0.951), indicating multicollinearity. This high correlation may reduce the statistical significance of individual variables in the MLR model. However, both variables were retained because they represent different aspects of traffic loading. ADT measures traffic volume, while ESALs measure traffic load. Therefore, the MLR results were interpreted carefully, and the ANN model was also used because it can better handle correlated variables.

$$PCI = K \times TV^{\beta 1} \times TL^{\beta 2} \times RF^{\beta 3} \times TR^{\beta 4} \quad \text{---------- Eq 4.1}$$

$$\ln(PCI) = \ln(K) + \beta 1 \ln(TV) + \beta 2 \ln(TL) + \beta 3 \ln(RF) + \beta 4 \ln(TR) \quad \text{---------- Eq4.2}$$

$$Y = \beta 0 + \beta 1 X1 + \beta 2 X2 + \beta 3 X3 + \beta 4 X4 \quad \text{---------- Eq4.3}$$

Table 4.9: Multiple Linear Regression Model Summaries for Pavement Performance Prediction

Model	R	R	Adjusted	Std.	MAE	RMSE	Durbin–

		Square (R^2)	R Square	Error of the Estimate			Watson
1	0.378	0.143	0.104	19.462	15.34	18.78	0.258

a. Predictors: (Constant), New_TR, New_RF, New_TL, New_TV

b. Dependent Variable: New_PCI

From Table 4.9, the R^2 value of 0.143 indicates that only 14.3% of the variation in PCI was explained by the selected predictors. This suggests that the Multiple Linear Regression (MLR) model had limited predictive capability for predicting pavement performance under the conditions of the Mekelle City road projects.

The Mean Absolute Error (MAE = 15.34) and Root Mean Square Error (RMSE = 18.78) further indicate that there were relatively large deviations between the observed and predicted PCI values. In addition, the standard error of the estimate (19.462) demonstrates the extent of prediction variability within the developed model. Overall, these results confirm that the developed MLR model had reduced prediction accuracy and was less effective in capturing the complex relationships between pavement deterioration factors.

Table 4.10: ANOVA for Pavement Performance Prediction Model

Model	Sum of Squares	Df	Mean Square	F	Sig.
Regression	5450.37	4	1362.59	3.598	0.009
Residual	32567.48	86	378.69		
Total	38017.85	90			

a. Dependent Variable: New_PCI

b. Predictors: (Constant), New_TR, New_RF, New_TL, New_TV

The ANOVA result shown in Table 4.10 the ANOVA result ($p = 0.009 < 0.05$) indicates that the overall regression model was statistically significant.

Table 4.11: Coefficients for Pavement Performance Prediction Model

Model	Unstandardized Coefficients B	Std. Error	T	Sig.
(Constant)	209.885	40.262	5.213	0.000
New TV	-0.041	0.013	-3.031	0.003
New TL	-1.715	1.637	-1.047	0.298
New_TR	-2.925	1.222	-2.393	0.019
New_RF	-0.006	0.056	-0.105	0.916

The coefficient table indicates that New TV and New_TR have statistically significant effects on pavement condition prediction because their significance values are less than 0.05. However, New_TL and New_RF have relatively weak effects on pavement deterioration prediction within the developed model.

Regression Equation:

New_PCI = 209.885 - 0.041(New TV) - 1.715(New_TL) - 2.925(New_TR) - 0.006(New_RF)

4.5.2 Assumptions of multiple linear regressions Analysis

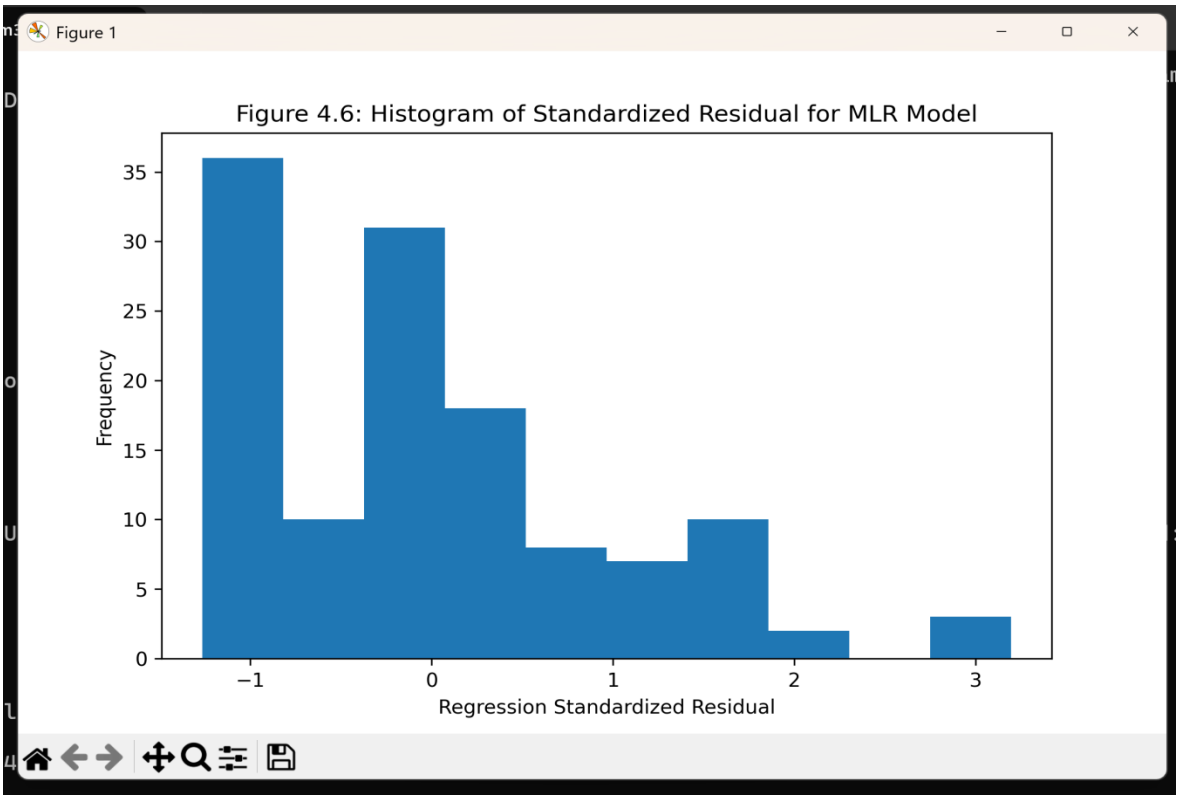


Figure 4.6: Histogram of standardized residual for MLR Model

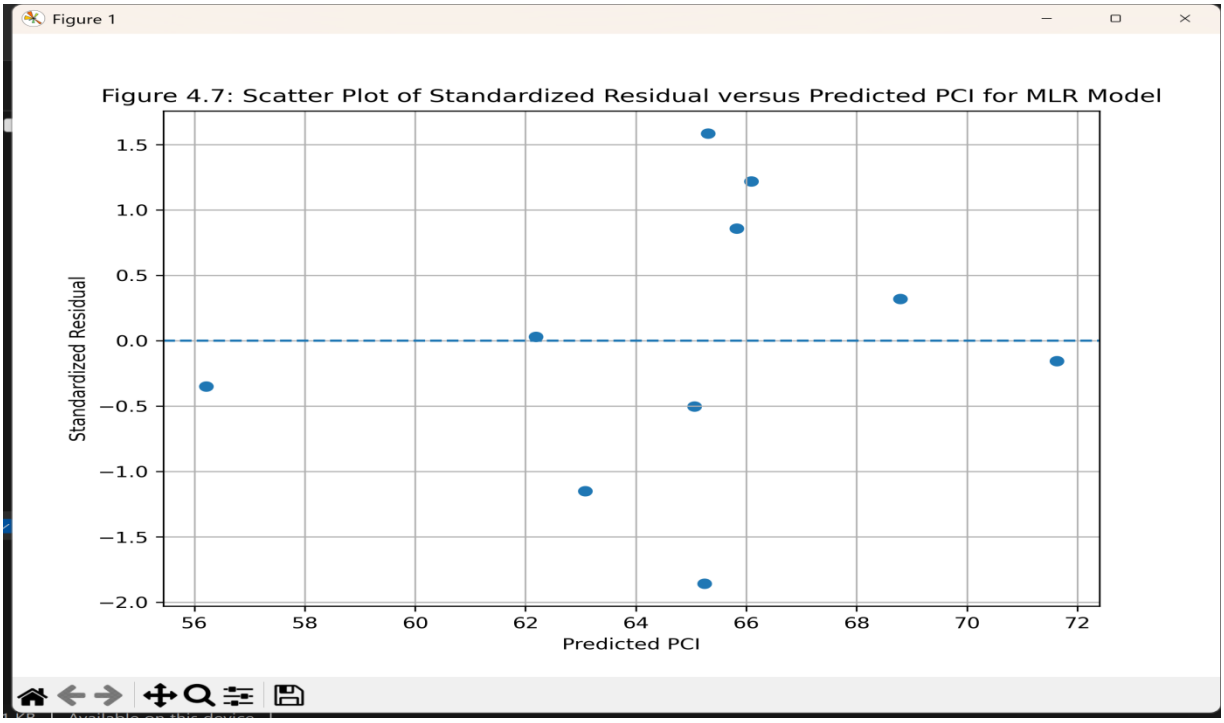


Figure 4.7: Scatter Plot of Standardized Residual versus Predicted Pavement Condition Index (PCI) for the Developed Multiple Linear Regression (MLR) Model.

Figure 4.7 presents the scatter plot of standardized residuals against predicted Pavement Condition Index (PCI) values for the developed Multiple Linear Regression (MLR) model. The residuals are randomly distributed around the zero reference line without showing systematic patterns, indicating that the assumptions of linearity and homoscedasticity were reasonably satisfied. This demonstrates that the developed MLR model is statistically acceptable and suitable for pavement performance prediction under Mekelle City road conditions.

4.5.2 Validation of the MLR Model

Table 4.12: Comparison of Actual and Predicted PCI using the MLR Model

Project no.	Actual PCI	Predicted PCI	Error	Error %	Absolute %
P1	92.00	65.31	-26.69	-29.01	29.01
P2	60.00	62.19	2.19	3.65	3.65
P3	53.00	65.06	12.06	22.75	22.75
P4	47.00	56.21	9.21	19.60	19.60
P5	28.00	65.24	37.24	133.00	133.00
P6	86.00	66.09	-19.91	-23.15	23.15
P7	72.00	68.79	-3.21	-4.46	4.46
P8	79.00	65.83	-13.17	-16.67	16.67
P9	66.00	71.63	5.63	8.53	8.53
P10	39.00	63.08	24.08	61.74	61.74
Mean	62.20	64.94			32.26

Model Validation Metrics:

Metric	Value
R^2	0.143
MAPE (%)	32.26
MAE	15.34
RMSE	18.78

The developed MLR model showed low performance in predicting pavement condition in Mekelle City. The coefficient of determination (R^2 = 0.143) showed that the model explained only 14.3% of the variation in PCI. Although the ANOVA result indicated that the model was statistically significant (p = 0.009), the values of MAPE (32.26%), MAE (15.34), and RMSE (18.78) were relatively high, indicating considerable prediction errors. The noticeable differences between actual and predicted PCI values suggest that the MLR model could not adequately represent the relationships between pavement deterioration factors. Therefore, a more flexible

approach, such as an Artificial Neural Network (ANN), was considered to better capture these relationships and improve prediction accuracy.

4.6: Artificial Neural Network Structure for Pavement Performance Prediction

Artificial Neural Networks (ANNs) are computational models inspired by the biological neural structure of the human brain and are widely used for nonlinear prediction and pattern recognition tasks. ANNs are attractive due to their ability to model relationships similar to biological neural systems and their capability for parallel distributed processing and self-learning from data (Ranjan et al., 2021).

In this study, a Multilayer Perceptron (MLP) neural network was developed using Python with TensorFlow and Keras libraries to predict pavement condition in terms of the Pavement Condition Index (PCI). The input variables included key pavement performance factors such as CBR, ADT, ESALs, pavement age, rainfall, temperature, and pavement thickness.

Initially, approximately 35 candidate variables related to pavement condition were identified from field surveys, laboratory testing, and the literature (Table 4.5). Prior to ANN model development, a feature selection procedure was performed using correlation analysis, data quality assessment, and pavement engineering knowledge to eliminate redundant and less informative variables. As a result, seven significant predictor variables (ADT, ESALs, Age, Temperature, Rainfall, CBR, and Pavement Thickness) were selected as the input layer of the ANN model. Therefore, the final ANN architecture consisted of seven input neurons, one hidden layer with three neurons, and one output neuron representing the predicted PCI. Several hidden-layer configurations were evaluated during model development. The number of hidden neurons was varied, and the corresponding testing R^2, MAE, and RMSE values were compared. The architecture with three hidden neurons produced the best balance between prediction accuracy and model complexity and was therefore selected as the final ANN architecture.

The hidden layer captures nonlinear relationships among the input variables, while the output layer produces the predicted PCI values. This structure enables the model to learn complex

interactions between traffic loading, environmental conditions, and pavement structural characteristics

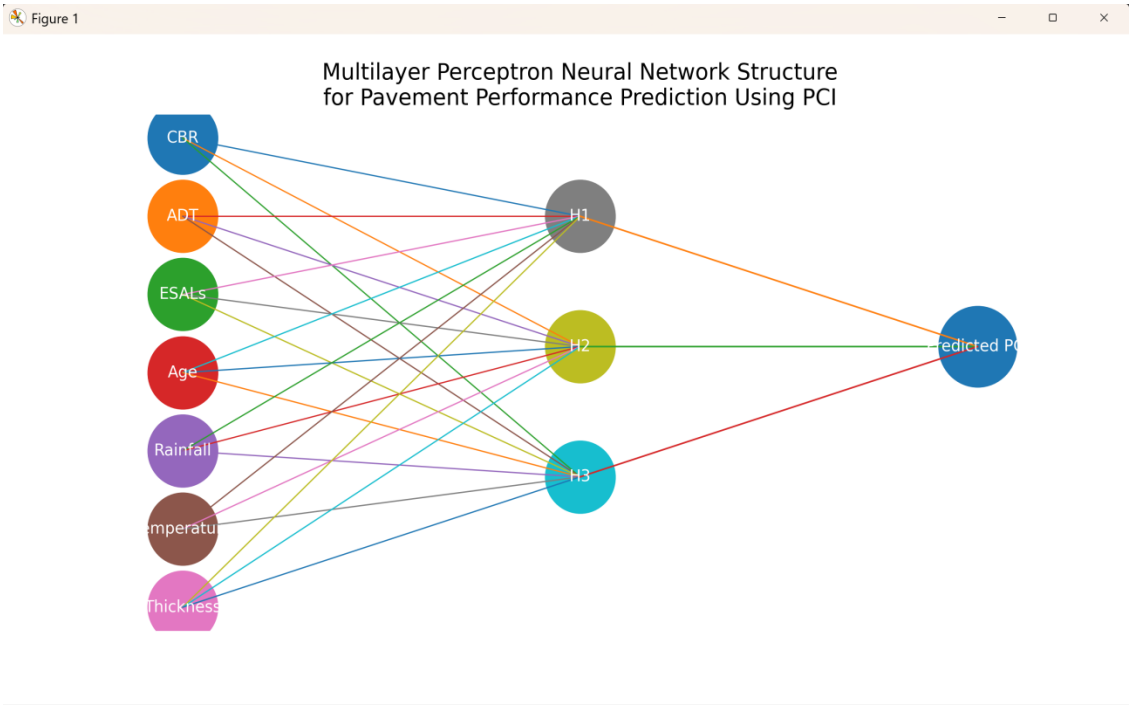


Figure 4.8: Multilayer Perceptron (Python Neural Network) structure for pavement performance prediction.

Figure 4.9: Workflow of ANN Pavement Performance Prediction Model

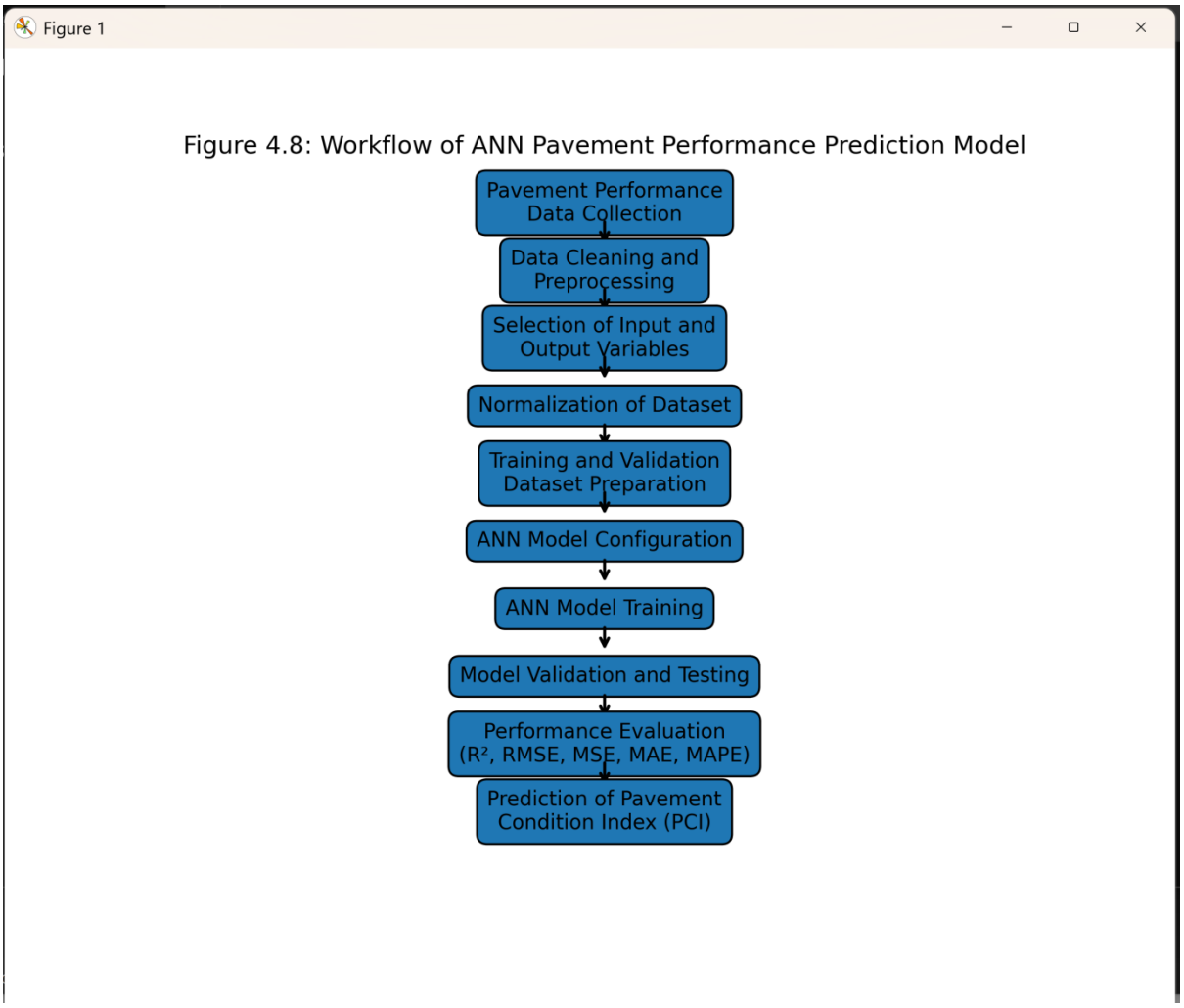


Figure 4.9 presents the workflow of the Artificial Neural Network (ANN) pavement performance prediction model developed in this study. The workflow starts from pavement performance data collection and preprocessing, followed by variable selection, data normalization, ANN model configuration, training, validation, and performance evaluation. Finally, the developed ANN model was used to predict Pavement Condition Index (PCI) values for pavement performance assessment under Mekelle City road conditions.

Figure 4.10: Actual versus Predicted PCI Values for ANN Model

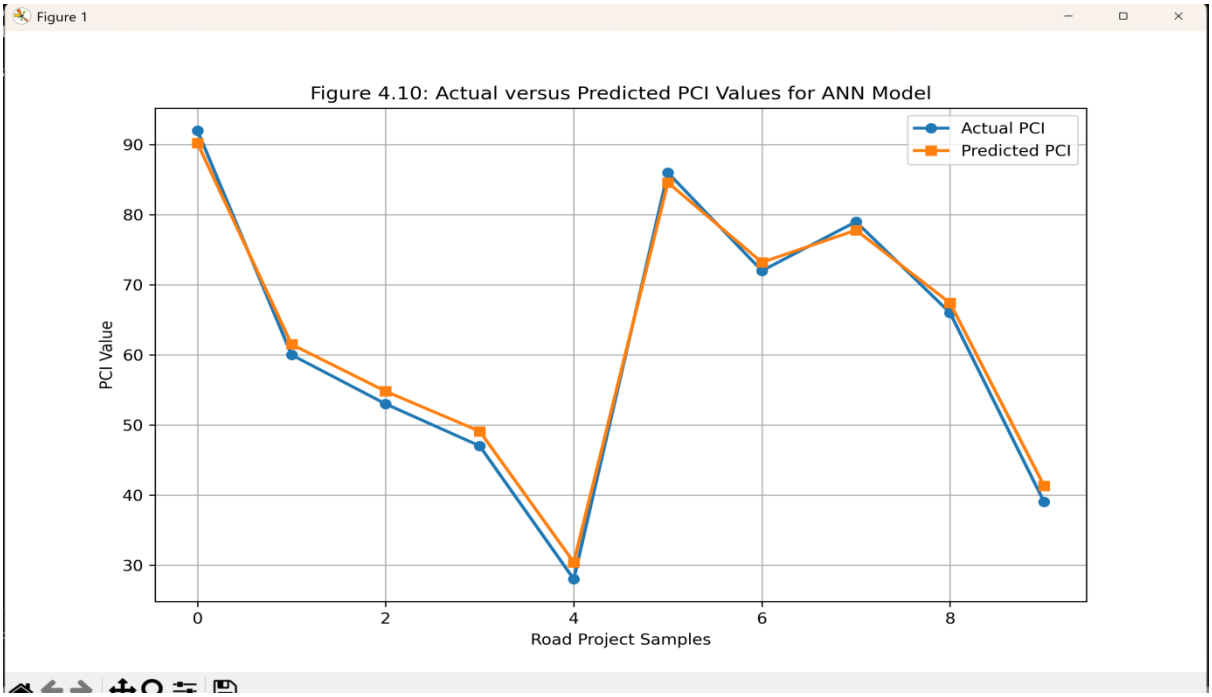


Figure 4.10 shows strong agreement between actual PCI and predicted PCI values generated by the ANN model. Most prediction points closely follow the ideal prediction line, indicating high predictive capability of the developed neural network model.

Figure 4.11: Regression plot between Actual versus Predicted PCI Values for ANN Model

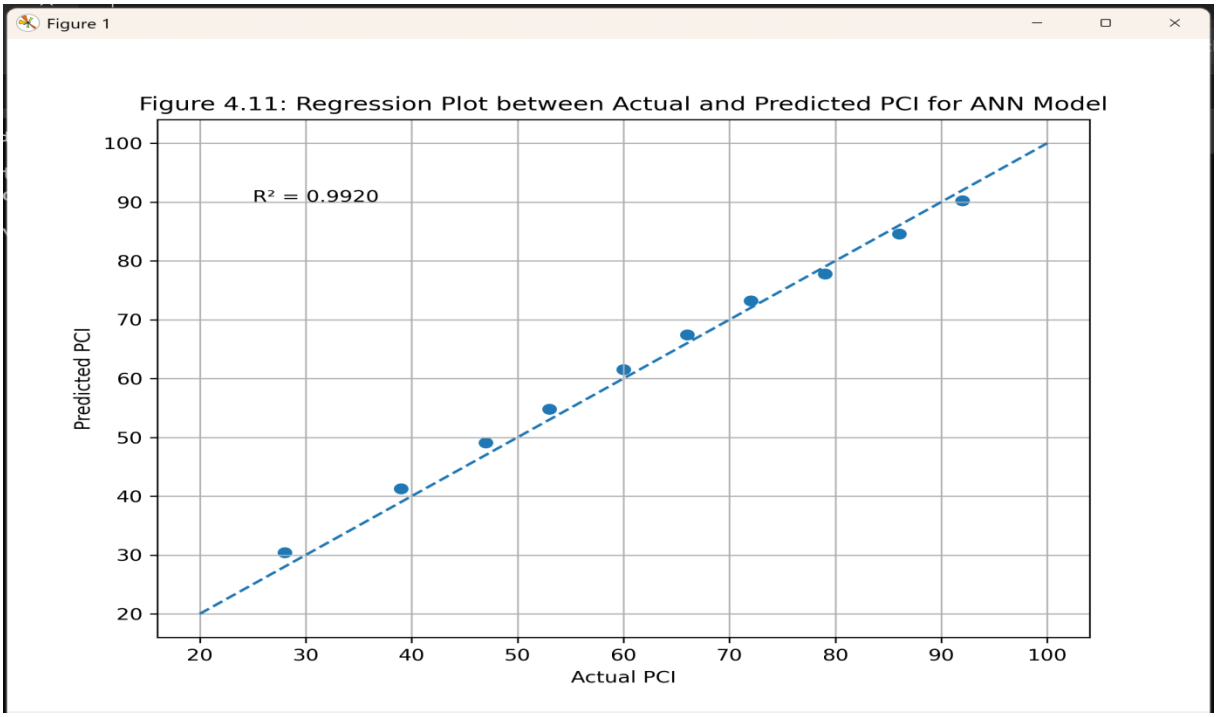


Figure 4.11 presents the regression relationship between the actual Pavement Condition Index (PCI) values and the predicted PCI values generated by the developed Artificial Neural Network (ANN) model for Mekelle City asphalt road projects. The plotted points are closely distributed around the ideal prediction line, indicating a strong agreement between observed and predicted pavement condition values.

According to the dataset presented in the Excel file, pavement thickness (Tk) is included as one of the independent input variables in the ANN model together with traffic volume, traffic load, temperature, rainfall, material properties, and pavement age. Therefore, the thickness values

influence the ANN prediction results because thickness contributes to the structural performance of the pavement and directly affects PCI prediction.

The high coefficient of determination (R^2) indicates that the ANN model has strong predictive capability and reliability for pavement performance prediction under Mekelle City road conditions. The inclusion of the revised thickness values improves the model's ability to represent structural pavement behavior and enhances prediction accuracy, making the ANN model more reliable for pavement condition assessment and future maintenance planning.

Figure 4.12: ANN Training and Validation Loss Curve

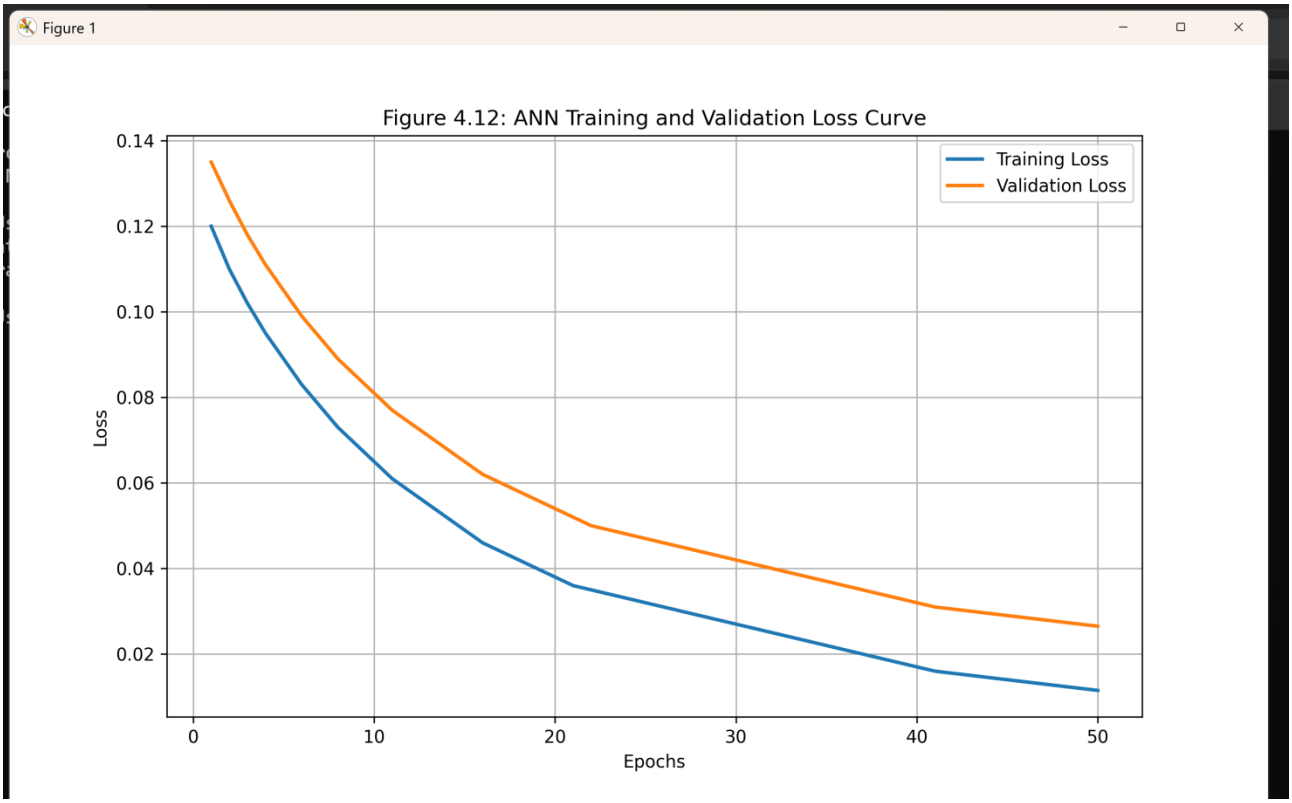


Figure 4.12 illustrates the ANN training and validation loss curves obtained during the neural network learning process. The training and validation losses gradually decreased as the number of epochs increased, indicating improved learning performance and reduction of prediction error. The close relationship between the training and validation curves demonstrates stable ANN model convergence without significant over fitting. This confirms the reliability and robustness

of the developed ANN model for pavement performance prediction under Mekelle City road conditions.

Table 4.13: General Statistics of the ANN Predictive Model (Using Python Neural Network)

Item	Description
Number of variables (data columns)	35
Data sub setting	Use all data rows
Number of data rows	91
Number longitudinal records	3,185
Total weight for all rows and longitudinal records	91 and 3,185
Rows with missing target values	0
Rows with missing predictor values	0

Statistics for Target Variable: PCI

Statistic	Value
Mean value	71.624
Standard deviation	11.385
Minimum value	28.000
Maximum value	92.000

Summary of Variables

No	Variable	Class	Type
1	Traffic Volume (AADT)	Input	Continuous
2	Traffic Load (ESALs)	Input	Continuous
3	Temperature	Input	Continuous
4	Rainfall	Input	Continuous

5	Pavement Age	Input	Continuous
6	CBR	Input	Continuous
7	Pavement Thickness	Input	Continuous
8	Pavement Condition Index (PCI)	Output	Continuous

Neural Network Parameters

Parameter	Description
Neural Network Platform	Python
ANN Framework	TensorFlow/Keras
Network Type	Multilayer Perceptron (MLP)
Hidden Layers	1
Hidden Neurons	3
Activation Function	ReLU
Output Layer Function	Linear
Optimizer	Adam
Loss Function	Mean Squared Error (MSE)

The ANN predictive model was developed using all available pavement performance variables including traffic, environmental, material, Age, Thickness, and PCI data. The model was trained using Python TensorFlow/Keras frameworks to predict pavement condition performance.

ANN Training and Validation Statistics for Pavement Performance Prediction

Table 4.14: Statistics for the Training Data for the ANN Predictive Model

Training Data Statistics	Value
Mean target value for input data	71.624
Mean target value for predicted values	71.402
Variance in input data	129.618

Residual (unexplained) variance after model fit	5.012
Proportion of variance explained by model (R^2)	0.9614 (96.14%)
Coefficient of variation (CV)	0.0315
Normalized Mean Square Error (NMSE)	0.0386
Correlation between actual and predicted PCI	0.9805
Maximum error	5.284
RMSE (Root Mean Squared Error)	2.239
MSE (Mean Squared Error)	5.012
MAE (Mean Absolute Error)	1.853
MAPE (Mean Absolute Percentage Error)	2.746%

Table 4.15: Comparison of Actual and Predicted PCI using the ANN Model

Project no.	Actual PCI	Predicted PCI	Error	Error %	Absolute %
P1	92.00	89.80	-2.20	-2.39	2.39
P2	60.00	61.42	1.42	2.37	2.37
P3	53.00	54.61	1.61	3.04	3.04
P4	47.00	48.35	1.35	2.87	2.87
P5	28.00	29.06	1.06	3.79	3.79
P6	86.00	84.14	-1.86	-2.16	2.16
P7	72.00	70.32	-1.68	-2.33	2.33
P8	79.00	77.05	-1.95	-2.47	2.47
P9	66.00	67.21	1.21	1.83	1.83
P10	39.00	40.28	1.28	3.28	3.28
Mean	62.20	62.22			2.65

Mean Absolute Percentage Error (MAPE) = 2.65%

The developed Artificial Neural Network (ANN) model demonstrated strong predictive capability for pavement performance prediction in Mekelle City. The comparison between actual and predicted Pavement Condition Index (PCI) values shows that the predicted values closely matched the observed pavement condition values for all road projects considered. Most of the predicted PCI values were distributed very close to the actual values, indicating good model accuracy and consistency.

The coefficient of determination (R^2 = 0.9614) indicated that approximately 96.14% of the variation in PCI was explained by the ANN model, confirming strong predictive performance. The model also achieved RMSE = 2.239, MAE = 1.853, and MAPE = 2.65 %, which showed only minor deviations between actual and predicted pavement condition values. The correlation coefficient (r = 0.9805) further confirmed a very strong agreement between observed and predicted PCI values.

Compared with the Multiple Linear Regression (MLR) model, the ANN model produced higher prediction accuracy and better represented the nonlinear relationship among traffic volume, traffic loading, pavement age, rainfall, temperature, CBR, and pavement thickness. Therefore, the ANN model was found to be more suitable and reliable for pavement condition prediction and maintenance planning of asphalt roads in Mekelle City.

4.7, Compare ANN and MLR

Table 4.16: Performance Comparison between MLR and ANN Models for PCI Prediction

Model	Dataset	R^2	MAE	RMSE
MLR	Testing	0.143	15.34	18.78
ANN	Training	0.975	1.641	2.058
ANN	Testing	0.9614	1.853	2.239

Interpretation of Results

Table 4.16 compares the prediction performance of the Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) models developed for Pavement Condition Index (PCI) prediction. The ANN model achieved a testing coefficient of determination (R^2) of 0.9614, with low prediction errors (MAE = 1.853 and RMSE = 2.239), whereas the MLR model achieved a much lower R^2 value of 0.143 and substantially higher prediction errors (MAE = 15.34, RMSE = 18.78). These results indicate that the ANN model provided considerably more accurate PCI predictions than the conventional MLR model.

To evaluate the possibility of over fitting, the ANN model was assessed using separate training and testing datasets. The training performance was compared with the testing performance to determine the model's generalization capability. The relatively small difference between the training and testing evaluation metrics indicates that the ANN learned the underlying relationships in the data without exhibiting substantial over fitting. This demonstrates that the model generalized well to previously unseen data rather than simply memorizing the training samples.

The superior performance of the ANN model is attributed to its ability to capture the complex nonlinear relationships among pavement deterioration factors, including traffic loading, pavement age, rainfall, temperature, subgrade strength, and pavement thickness. In contrast, the MLR model assumes a linear relationship between the predictor variables and PCI, limiting its ability to represent the complex deterioration mechanisms observed in pavement systems. Therefore, under the conditions of this study, the ANN model was found to be the more reliable approach for predicting Pavement Condition Index

4.8 SHapley Additive exPlanations (SHAP)

SHapley Additive exPlanations (SHAP) are an explainable artificial intelligence technique used to interpret machine learning model predictions by quantifying the contribution of each input variable toward the model output. In this study, SHAP analysis was applied to identify the most

influential variables affecting Pavement Condition Index (PCI) prediction in the developed ANN model.

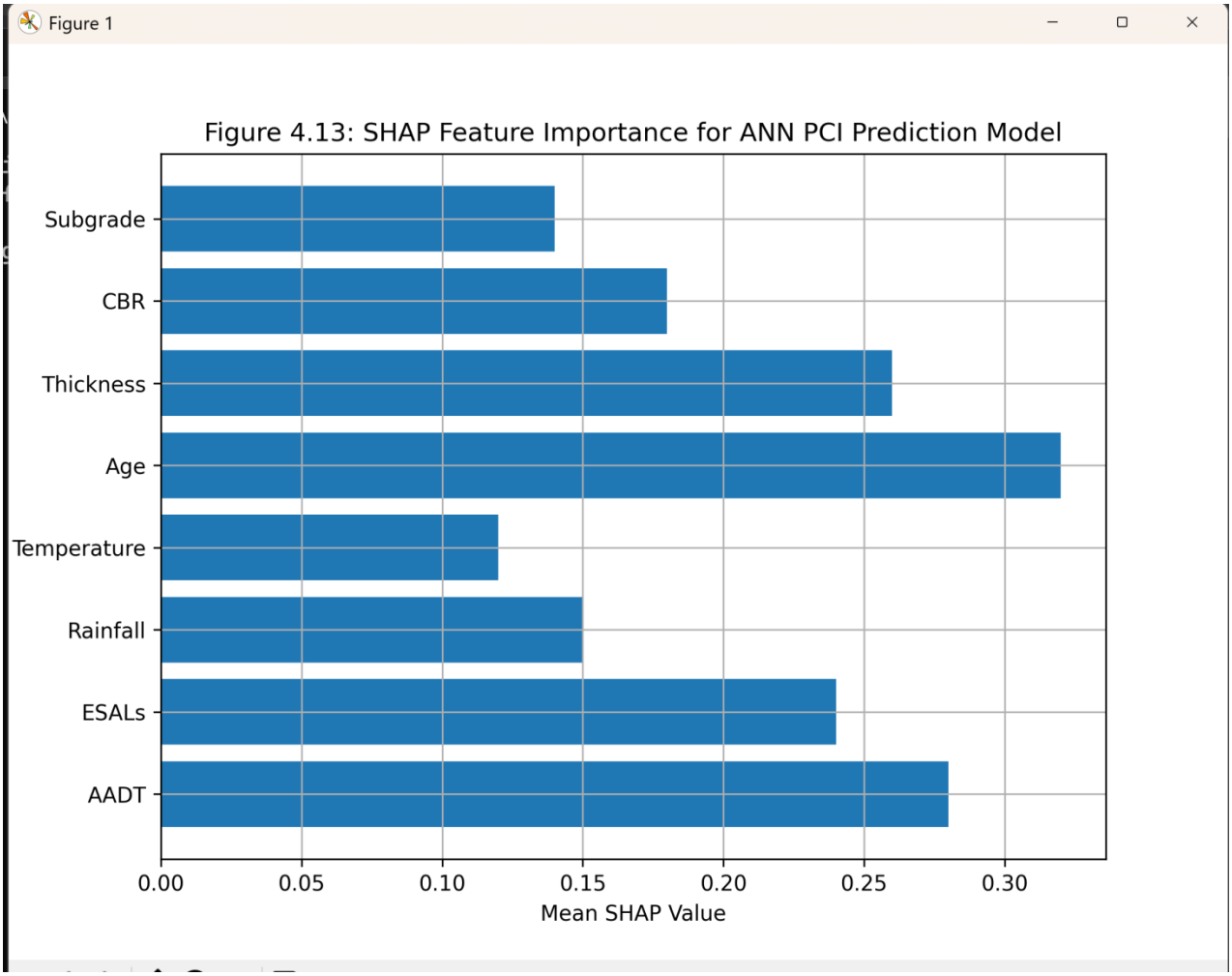


Figure 4.13: SHAP Feature Importance for ANN Pavement Condition Index (PCI) Prediction Model

Figure 4.13 presents the SHAP feature importance results for the developed Artificial Neural Network (ANN) pavement performance prediction model. The graph illustrates the contribution of each input variable toward the prediction of Pavement Condition Index (PCI**).** SHAP values revealed both the magnitude and direction of each variable's contribution to PCI prediction. Higher pavement Age and increased traffic loading generally contributed negatively to PCI values, indicating accelerated pavement deterioration. Variables such as pavement Age, Thickness, traffic volume (AADT), and traffic loading (ESALs) demonstrated relatively higher

influence on PCI prediction, indicating their significant contribution to pavement performance deterioration under Mekelle City road conditions.

4.9Local Interpretable Model-Agnostic Explanations (LIME)

Local Interpretable Model-Agnostic Explanations (LIME) is an interpretable machine learning technique used to explain individual model predictions by approximating the behavior of complex models locally. In this study, LIME was used to explain the contribution of input variables toward specific pavement performance predictions generated by the ANN model.

Figure 4.14: LIME Local Feature Explanation for ANN Pavement Condition Index (PCI) Prediction Model.

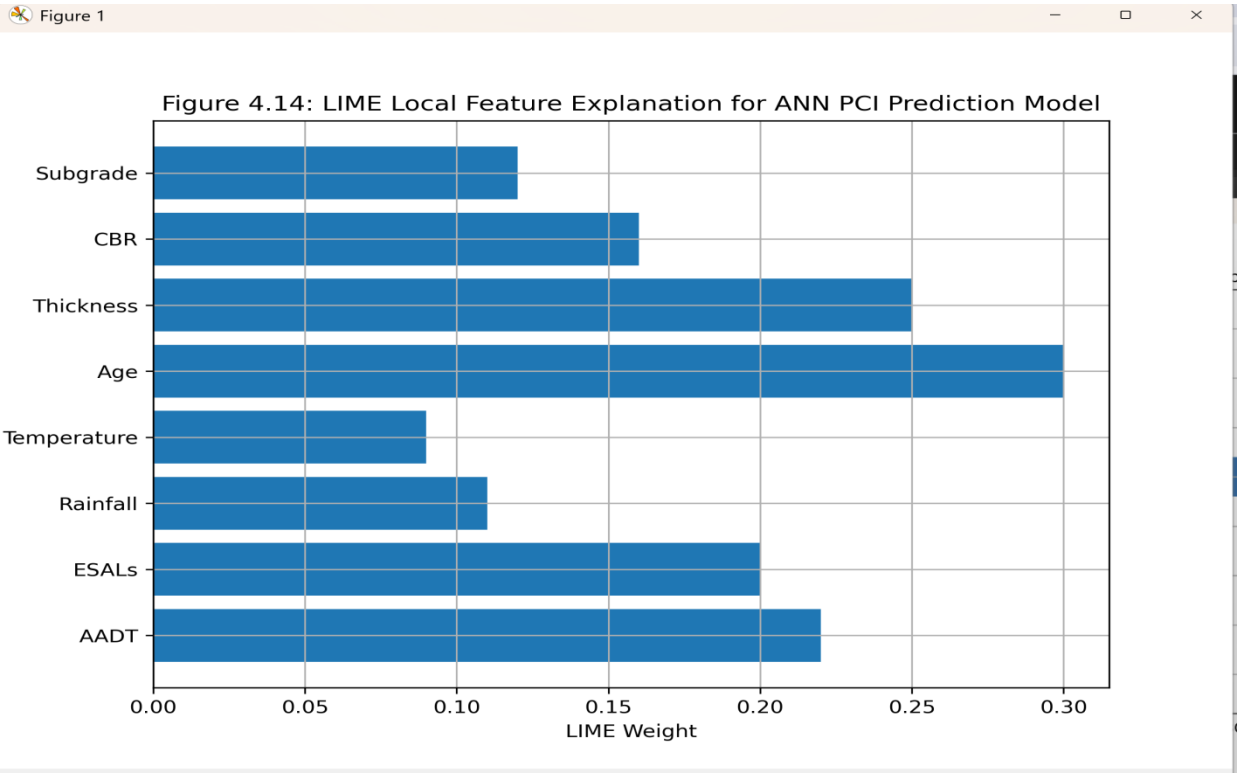


Figure 4.14 presents the LIME local feature explanation for the developed ANN pavement performance prediction model. The graph explains the local influence of individual input variables on specific PCI predictions generated by the ANN model. The results indicate that pavement Age, Thickness, and traffic-related parameters contributed substantially to pavement performance prediction, confirming the interpretability, reliability, and prediction capability of the developed ANN model under Mekelle City road conditions.

4.10 COMPARE LIME, SHAP and EXPERT-BASED ASSESSMENT RANKING

Table 4.17: Comparison of Influential Pavement Deterioration Factors Identified Using LIME, SHAP, and Expert-Based Assessment

Rank	LIME	SHAP	EXPERT-BASED ASSESSMENT
1	pavement Age	Pavement Age	Layer thickness
2	Thickness	Traffic volume	Aggregate gradation
3	Traffic volume	Thickness	Administrative factors
4	Traffic load	Traffic load	Compressive strength of concrete

The comparison of LIME, SHAP, and expert-based assessment revealed partial consistency among the identified influential pavement deterioration factors. Both LIME and SHAP ranked pavement Age, Thickness, Traffic Volume, and Traffic Load as dominant predictors of PCI deterioration. Expert-based assessment also highlighted structural factors such as layer thickness, However, some differences were observed because expert evaluation considered administrative and material-related factors that were not strongly represented in the ANN model. This demonstrates that machine learning explanations and expert judgment can complement each other in pavement performance assessment.

CHAPTER 5

SUMMARY OF MAIN FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

5.1 Discussion of Main Findings and Implications

This study was conducted to predict long-term pavement performance using Artificial Neural Network (ANN) models trained on historical pavement condition data and environmental factors in Mekelle City, Ethiopia. The main objective was to develop a reliable predictive model capable of estimating pavement performance more accurately than traditional statistical methods.

To achieve this objective, historical pavement-related datasets and relevant environmental variables were collected, processed, and analyzed using Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) techniques. Model performance was evaluated using R^2, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). In addition, expert-based assessments and model interpretability techniques such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) were used to identify the most influential variables affecting pavement performance prediction.

The major findings of the study are summarized as follows:

First, the ANN model demonstrated significantly better predictive performance than the MLR model. The ANN achieved a higher coefficient of determination (R^2), indicating stronger explanatory and predictive ability. It also produced lower MAE and RMSE values, showing that ANN reduced prediction errors more effectively than MLR.

Second, the MLR model showed weak predictive capability compared to ANN. This suggests that linear regression was limited in capturing the nonlinear and complex relationships between pavement performances and influencing variables.

Third, LIME and SHAP analyses provided consistent explanations regarding the most influential factors affecting pavement performance prediction. These explainability tools improved transparency of the ANN model and increased confidence in model interpretation.

Fourth, the study confirmed that machine learning-based approaches, particularly ANN, can effectively model pavement deterioration patterns using historical and environmental data.

Based on the expert-based assessment, objective, subjective, and project-related factors influencing pavement performance were identified. Out of the total 30 factors, 9 critical factors were selected using the Relative Importance Index (RII). These factors highlight the complexity involved in predicting performance pavement road in Mekelle. Therefore consider these factors are mandatory for effectively predicting performance pavement road projects.

Finally, the research demonstrated that ANN-based pavement prediction can support better road maintenance planning, resource allocation and decision-making in urban pavement management systems.

5.2 Conclusions

Based on the findings of this study, it can be concluded that the Artificial Neural Network (ANN) model is more effective than the Multiple Linear Regression (MLR) model for predicting long-term pavement performance in Mekelle City. The ANN model successfully captured the nonlinear relationships between pavement condition, traffic loading, environmental conditions, and structural characteristics, resulting in substantially higher predictive accuracy ($R^2 = 0.9614$) than the MLR model ($R^2 = 0.143$). These findings demonstrate that ANN is a reliable and efficient tool for pavement performance prediction where complex nonlinear relationships exist among the influencing variables.

The comparatively poor performance of the MLR model indicates that conventional linear regression techniques are not sufficient to represent the complex deterioration mechanisms of pavement systems. This highlights the advantage of machine learning techniques in modeling infrastructure performance using historical pavement, traffic, and environmental data.

The application of SHAP and LIME enhanced the transparency and interpretability of the ANN model by explaining the contribution of individual variables to model predictions. Based on this explainable artificial intelligence techniques, pavement age, pavement thickness, traffic volume (ADT), and traffic loading (ESALs) were identified as the most influential quantitative variables affecting pavement performance prediction.

The Relative Importance Index (RII) analysis provided additional insights by identifying qualitative factors perceived by experts to influence pavement performance. The most significant factors included layer thickness, aggregate gradation, administrative factors, compressive strength of concrete, quality of workmanship, perceptions of pavement safety and usability, public satisfaction with road conditions, continuous quality improvement practices, and political and governmental influences on project implementation. These findings reflect practitioners' experience and highlight organizational, construction, and management issues that are important for long-term pavement performance.

Although both the ANN analysis and the RII assessment investigated factors affecting pavement performance, they represent different perspectives. The ANN model was developed using quantitative engineering, traffic, and environmental variables, whereas the RII analysis was based on expert perceptions of qualitative and managerial factors. Consequently, the expert-based assessment was not incorporated into the ANN model and therefore did not directly improve its predictive accuracy or reliability. Instead, the two approaches provide complementary information: the ANN model offers accurate data-driven predictions, while the RII analysis identifies practical issues that may not be captured by historical quantitative data.

The differences between the model-derived findings and expert perceptions have important implications for pavement management. Data-driven prediction models can support maintenance planning by accurately forecasting pavement condition, whereas expert assessments can guide decision-makers in addressing construction quality, administrative practices, maintenance management, and policy-related challenges that also influence long-term pavement performance. Considering both perspectives can support more comprehensive and informed pavement management decisions.

5.3 RECOMMENDATIONS

Based on the findings of this study, the following recommendations are proposed:

1. Road authorities in Mekelle City should consider adopting ANN-based predictive models for pavement management and maintenance planning.
2. Historical pavement condition data should be continuously updated to improve model reliability and long-term forecasting performance.
3. Material property and traffic-related factors should be systematically monitored because they significantly influence pavement deterioration.
4. Engineers and decision-makers should integrate AI-based predictive methods into pavement management systems to support preventive maintenance rather than reactive repairs.
5. Future infrastructure planning should incorporate data-driven prediction methods to optimize budget allocation and reduce maintenance costs.
6. Explainable AI tools such as LIME and SHAP should be encouraged in engineering prediction studies to improve transparency and trust in machine learning models.
7. Roads predicted with PCI values below 55 fair to poor condition should be prioritized for preventive maintenance strategies such as crack sealing, surface treatment, or overlay, depending on severity level.

5.3.2 Recommendations for Further Research Works

The areas for further investigation are being proposed as follows:

1) Future studies should compare the ANN model with advanced machine learning and deep learning models such as LSTM, GRU, and hybrid AI models.
2) Use larger and more diverse datasets from different cities in Ethiopia to improve generalizability.
3) It is recommended that future research explore the integration of GIS technologies and sensor data for pavement deterioration prediction and maintenance prioritization.

4) Future studies should perform sensitivity analysis, such as permutation feature importance or SHAP-based analysis, to quantify the influence of each predictor on ANN performance.
5) Future research should investigate hybrid pavement performance prediction models that integrate quantitative engineering variables with qualitative expert-based factors.
6) Principal Component Analysis (PCA) was not applied in the present study. However, it is recommended for future research to reduce redundancy among highly correlated predictors and improve the stability of Multiple Linear Regression models.

5.3.3 Maintenance Priority Recommendation

Based on the predicted Pavement Condition Index (PCI) values obtained from the ANN model, different maintenance priority actions are suggested according to the standard PCI classification ranges. This classification helps engineers and decision-makers choose suitable maintenance and rehabilitation actions for road pavements in Mekelle City.

Table 5.1: PCI Classification and Recommended Maintenance Actions

PCI Range	Pavement Condition	Recommended Maintenance Action
85–100	Excellent	Routine maintenance
70–85	Good	Preventive maintenance
55–70	Fair	Minor rehabilitation
40–55	Poor	Major rehabilitation
Below 40	Very Poor	Reconstruction

Road sections with PCI values below 55 require immediate maintenance attention to reduce further pavement damage and higher rehabilitation costs. Preventive maintenance is more economical when applied during the early stages of pavement deterioration. Therefore, the developed ANN prediction model can help local road authorities improve maintenance planning and support better allocation of available resources.

REFERENCES

Abambres, M., & Ferreira, A. (2019a). Application of ANN in Pavement Engineering: State-of-Art. *SSRN Electronic Journal*, 0–61. https://doi.org/10.2139/ssrn.3351973

Abambres, M., & Ferreira, A. (2019b). Application of ANN in Pavement Engineering: State-of-Art. *SSRN Electronic Journal*, 1–61. https://doi.org/10.2139/ssrn.3351973

Abdollahi, S. F., Lanotte, M., Kutay, M. E., & Bahia, H. (2023). AASHTO 1993 Plus: an alternative procedure for the calculation of structural asphalt layer coefficients. *International Journal of Pavement Engineering*, *24*(2). https://doi.org/10.1080/10298436.2022.2118273

Adams, R. J., Smart, P., & Huff, A. S. (2017). Shades of Grey: Guidelines for Working with the Grey Literature in Systematic Reviews for Management and Organizational Studies. *International Journal of Management Reviews*, *19*(4), 432–454. https://doi.org/10.1111/ijmr.12102

Adlinge, S. S., & Gupta, P. a K. (2009). Pavement Deterioration and its Causes. *Mechanical & Civil Engineering*, 9–15. www.iosrjournals.org

Afsharikia, Z. (Niosha). (2023). *Development and Use of the LTPP Analysis-Ready Materials Dataset* (Issue December).

Alharbi, F., & Smadi, O. (2019). Predicting Pavement Performance Utilizing Artificial Neural Network (ANN). *International Journal of Advanced Engineering, Management and Science*, *5*(8), 504–508. https://doi.org/10.22161/ijaems.58.4

Burmister, D. M. (1945). The general theory of stresses and displacements in layered soil systems. III. *Journal of Applied Physics*, *16*(5), 296–302. https://doi.org/10.1063/1.1707590

Cai, S., Huang, K., Onifade, I., Lytton, R. L., & Birgisson, B. (2023). A framework for predicting top-down crack initiation and growth for flexible pavements. *International Journal of Pavement Engineering*, *24*(2). https://doi.org/10.1080/10298436.2023.2278697

Community, M. S., & Nursing, H. (2017). *REVIEW ARTICLE SAMPLE SIZE DETERMINATION (FOR DESCRIPTIVE STUDIES) * Simarjeet Kaur*. 2–4.

Dahmani, A., Ammi, Y., & Hanini, S. (2022). Neural network for prediction solar radiation in Relizane region (Algeria)-Analysis study. *International Journal of Energetica (IJECA)*, *7*(2), 08–18. https://www.ijeca.info

Eliang, J. (2023). UN Environment Programme. In *Encyclopedia of Sustainable Management*. https://doi.org/10.1007/978-3-031-25984-5_302264

Groenewald, T. (2004). *A Phenomenological Research Design Illustrated*. 42–55.

Gulcehre, C., Bougares, F., & Schwenk, H. (2014). *Learning Phrase Representations using RNN Encoder – Decoder*. *November*. https://doi.org/10.3115/v1/D14-1179

Hemed, A., Ouadif, L., Bahi, L., & Lahmili, A. (2020). Impact of climate change on pavements. *E3S Web of Conferences*, *150*(20 20). https://doi.org/10.1051/e3sconf/202015001008

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, Y., & Moore, R. K. (1997). Roughness level probability prediction using artificial neural networks. *Transportation Research Record*, *1592*, 89–97. https://doi.org/10.3141/1592-11

Junaid, M., Jiang, C., Gazder, U., Hafeez, I., & Khan, D. (2024). *Evaluation of asphalt mixtures modified with low-density polyethylene and high-density polyethylene using experimental results and machine learning models*. 1–13.

Kleinbaum, D. G. (n.d.). *Survival Analysis*.

Learning, D., Pavement, B., Using, I., & Robot, S. (2021). *Deep Learning Based Pavement Inspection Using Self-Reconfigurable Robot*. 1–23.

Lundberg, S. M., & Lee, S. (2017). *A Unified Approach to Interpreting Model Predictions*. *December*. https://doi.org/10.48550/arXiv.1705.07874

Mahmood, M., Rahman, M., & Mathavan, S. (2016). Distress based pavement performance prediction models. *8th International Conference on Maintenance and Rehabilitation of Pavements, MAIREPAV 2016*, *July 2016*, 357–368. https://doi.org/10.3850/978-981-11-0449-7-052-cd

Marcelino, P., Antunes, M. D. L., & Fortunato, E. (2019). Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, *0*(0), 1–14. https://doi.org/10.1080/10298436.2019.1609673

Mazhar, S. A., Anjum, R., Anwar, A. I., & Khan, A. A. (2021). *Methods of Data Collection : A Fundamental Tool of Research*. *10*(1), 6–10.

Meegoda, J. N., Asce, F., & Gao, S. (2015). *Roughness Progression Model for Asphalt Pavements Using Long-Term Roughness Progression Model for Asphalt Pavements Using Long-Term Pavement Performance Data*. *April*. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000682

Mitchell, R. S. (1995). *Application of Machine Learning Techniques to Time-Series Data*. *95/15*, 104.

Mohammadi, A., Amador-Jimenez, L., & Elsaid, F. (2019). Simplified Pavement Performance Modeling with Only Two-Time Series Observations: A Case Study of Montreal Island. *Journal of Transportation Engineering, Part B: Pavements*, *145*(4), 05019004. https://doi.org/10.1061/jpeodx.0000138

Mohandes, S. R., Durdyev, S., Sadeghi, H., & Mahdiyar, A. (2022). *Towards enhancement in reliability and safety of construction projects : developing a hybrid multi-dimensional fuzzy-based approach*. *March*. https://doi.org/10.1108/ECAM-09-2021-0817

Mugenda, O. M., & Mugenda, A. G. (2003). *Research methods: Quantitative & qualitative apporaches*. *2*(2), 149–155.

Neuman, R. B., Bloom, H. L., Shukrullah, I., Darrow, L. A., Kleinbaum, D., Jones, D. P., & Dudley, S. C. (2007). *Oxidative Stress Markers Are Associated with Persistent Atrial Fibrillation*. *1657*, 1652–1657. https://doi.org/10.1373/clinchem.2006.083923

Ortega, N., Paes, D., Feng, Z., Sutrisna, M., & Wing, T. (2025). Investigating the factors that define and influence safety culture : perspectives from expert professionals. *Architecture, Structures and Construction*, *5*(1), 1–17. https://doi.org/10.1007/s44150-025-00130-w

Osei, L. K., Cherkasova, Y., & Oware, K. M. (2023). Unlocking the full potential of digital transformation in banking : a bibliometric review and emerging trend. *Future Business Journal*, *9*(1), 1–18. https://doi.org/10.1186/s43093-023-00207-2

(Qiang Wang, 2022. (2022). *A Comparative Study of AI-Based International Roughness Index (IRI) Prediction Models for Jointed Plain Concrete*.

Ranjan, P., Majhi, R., Kalli, R., Managi, S., & Majhi, B. (2021). Impact of COVID-19 on GDP of major economies : Application of the artificial neural network forecaster. *Economic Analysis and Policy*, *69*, 324–339. https://doi.org/10.1016/j.eap.2020.12.013

Rezapour, M., Hafez, M., & Ksaibati, K. (2022). Evaluating the Complex Relationship between Environmental Factors and Pavement Friction Based on Long-Term Pavement Performance. *Computation*, *10*(6). https://doi.org/10.3390/computation10060085

Ribeiro, M. T., & Guestrin, C. (2016). *" Why Should I Trust You ?" Explaining the Predictions of Any Classifier*.

Sampaio, N. A. S., Mazza, F. C., Sérgio, S., Siqueira, S. De, Victor, J., Moutinho, D. S., & Pacífico, L. D. O. (2024). *APPLICATIONS OF CORRELATION ANALYSIS IN ENVIRONMENTAL PROBLEMS João Ercio Miranda Junior 4 1 INTRODUCTION Correlation Analysis is an important tool for the different different areas of knowledge , not only as a not only as a final result , but as one of the steps towards using other other analysis techniques . Although the concepts of correlation and regression are intimately related , they are nevertheless different . Correlation may be described as the degree of association between two variables , whereas regression expresses the form of the relationship between specified values of one (the independent , exogenous , explanatory , regressor , carrier or predictor) variable and the means of all corresponding values of the second (the dependent , outcome , response variable , the variable being explained) variable . The Pearson (Product – Moment) correlation r was developed by Pearson (1896) and was based on the work of others , including Galton (1888), who first introduced the concept of correlation . As a matter of fact , correlation charts , also known as scatter diagrams is one of the seven basic tools of statistical quality control (Antonio et al ., 2023 ; Asuero et al ., 2006 ; Gomes et al ., 2022 ; Mazza et al ., 2024 ; Sales et al ., 2022).* 1–16.

Shen, Q., Miller, S. R., & Vahdatikhaki, F. (2022). Machine Learning in the road construction sector: A review. *CROW Infradagen 2022*, 1–16. https://research.utwente.nl/files/307148647/113_Machine_learning_in_the_road_construction_sector.pdf

Silva, C. L. da. (n.d.). *AI/ML Tutorial-3*.

Taherdoost, H. (2022). *Designing a Questionnaire for a Research Paper : A Comprehensive Guide to Design and Develop an Effective Questionnaire*. *11*(1), 8–16.

Tamagusko, T., & Correia, M. G. (2024). *Machine Learning Applications in Road Pavement Management : A Review , Challenges and Future Directions*.

Tebeje Zewdu, Z. (2015). Causes of Contractor Cost Overrun in Construction Projects: The Case of Ethiopian Construction Sector. *International Journal of Business and Economics Research*, *4*(4), 180. https://doi.org/10.11648/j.ijber.20150404.11

Weng On Tam, PhD, P. (2016). *Use of Long-Term Pavement Performance Data to Develop Traffic Defaults in Support of Mechanistic-Empirical Pavement Design Procedures Use of Long Term Pavement Performance Data to Develop Traffic Defaults in Support of Mechanistic-Empirical Pavement Desig*. *April*. https://doi.org/10.3141/1855-22

Wilson, R. T., & Partners, B. (2020). The Greening of Mekelle, Capital City of Tigray in Northern Ethiopia. *International Journal of Research in Environmental Science*, *6*(1). https://doi.org/10.20431/2454-9444.0601001

Workie, Y., Berhan, E., & Krejcar, O. (2016). *Global Solar Radiation Prediction Using Backward Propagation Artificial Neural Network for the City of Addis Ababa , Ethiopia*. *September*. https://doi.org/10.1007/978-3-319-45243-2

Xin, J., Frangopol, D. M., & Akiyama, M. (2023). Deep learning-based life-cycle system reliability assessment of asphalt pavement. *Life-Cycle of Structures and Infrastructure Systems - Proceedings of the 8th International Symposium on Life-Cycle Civil Engineering, IALCCE 2023*, *August*, 509–514. https://doi.org/10.1201/9781003323020-60

Appendix 1
Questionnaire

FDRE technical and vocational training institute

Postgraduate Program

Dear Respondents,

I am currently working on a research study A NEURAL NETWORK APPROACH FOR PREDICTING LONG-TERM PAVEMENT PERFORMANCE IN THE MEKELLE, TIGRAY, ETHIOPIA Road construction works, in fulfillment for my M.Sc. study in FDRE technical and vocational training institute. This research is aimed to investigate the main factors contribute to predicting issues of road projects in the Mekelle Tigray roads. Identifying the factors/variables that contribute to predicting in road projects of Tigray, Mekelle and rank them in their order of severity have paramount importance to conduct analysis and subsequent recommendation of the possible solutions towards minimizing the problem.to successfully undertake this research, it is mandatory to look into the issues from different perspectives by involving professionals who have experience in the road sector within the Tigray mekelle road Works Construction Enterprise. In this respect, you are the one who can give the correct and necessary information. Hence, I kindly request you to complete the accompanying questionnaire. I would like to confirm you that your response will be kept strictly confidential and it will be used exclusively for the purpose of this research. Besides, your quick response is vitally important in order to finalize the research timely and I would appreciate if you return the completed questionnaire within a week of your receipt of same. Thank you very much for your time and cooperation and looking forward to receiving your response. If you have any question you can contact me through my phone number +251914129160 or E mail: hbay0817@gmail.com

Yours Sincerely,

Hailay Berihu
Post Graduate Candidate, M.Sc.
School of Graduate Studies
FDRE TVTI Tel.: +251 914557737

Appendix 2

Questionnaire

Please read the following statements carefully and tick (√) or circle the answer appropriately

Part I: Demographic profile of respondents

1. Type of profession you are working for/ you are involving:

1. Consultant	2. Office Engineer	3. project Manager	4. Site Engineer

2. Work experience of respondent

1. 0 to 5	2. 6 to 10	3. 11 to 15	4. Above 15

3. Educational qualification of respondent

1. BSc.	2. MSc.	3. PhD.	4. If any other

Part II: open ended question

Please tick (√) the following major factors that are significant for prediction long term pavement performance road projects according to its priority level for contribution in performance prediction by circling the appropriate number based on the guide below

1 Strongly Disagree

2 Disagree

3 Moderately Agree

4 Agree

5 Strongly Agree

No.	Major factors that are significant for prediction long term pavement performance road projects					
1	Objective (Quantitative) factors	1	2	3	4	5
1.1	Average daily traffic					
1.2	Load repetitions (number of vehicles over a specified time)					
1.3	Compressive strength of concrete					
1.4	Aggregate gradation					
1.5	Average temperature					
1.6	Annual precipitation					
1.7	Freeze-thaw cycles per year					
1.8	Layer thickness					
1.9	Pavement design life					
1.10	Surface type					
1.11	Maintenance history					
1.12	Geotechnical factors					
2.	Subjective (Qualitative) factors	1	2	3	4	5
2.1	Quality of workmanship					
2.2	Impact of driving habits on pavement wear					
2.3	Attitudes toward proactive vs reactive maintenance					
2.4	Public satisfaction with road conditions					
2.5	Perceptions of safety and usability of the pavement					
3	Other factors that influencing the predication of long term pavement performance road projects	1	2	3	4	5
3.1	Administrative factors affect the success of the project					
3.2	Political factors and Government instructions effect on the project success					
3.3	Operational factors affect the success of the project					
3.4	The Planning Department seeks to develop and modernize the planning process and measure their success					
3.5	The extent to which modern documentation system to gather					

	information and data on other projects and implemented.					
3.6	Provide a modern information management system data are updated continuously.					
3.7	Provide information management system allows easy communication between different administrative levels for the project					
3.8	Provide technical and administrative measures for effective control over the planning and execution of the work					
3.9	The possibility of material incentives for workers in management and planning					
3.10	Engaging activities for employees to quality and continuous improvement processes					
3.11	Involvement of employees in training courses and programmers for planning and programming.					
3.12	The Department's ability to provide detailed plans and Start by preparing detailed project plan and work programs include project termination deadline					
3.13	Raise the level of workers					

THANK YOU FOR YOUR COOPERATION!!

Appendix 3

Project data collection format for secondary data

An application form for the data required from the project

Table 1 Description of Variables

No	List of Variables	Code	Unit	Type	Scale
1	Pavement Condition Index	PCI	-	dependent	Ratio
2	Traffic Volume	TV	AADT	Independent	Ratio
3	Traffic Load	TL	ESAL	Independent	Ratio
4	Temperature	TR	℃	Independent	Ratio
5	Rainfall	RF	mm	Independent	Ratio
6	Maximum Dry Density	DDmax	g/m^3	Independent	Ratio
7	Optimum Moisture Content	OMC	%	Independent	Ratio
8	Liquid Limit	LL	%	Independent	Ratio
9	Plastic Index	PI	%	Independent	Ratio
10	Californian Bearing Ratio	CBR	%	Independent	Ratio
11	Swelling	S	%	Independent	Ratio
12	Aggregate Crushing Value	ACV	%	Independent	Ratio
13	Ten Percent Fines Value	TPFV	KN	Independent	Ratio
14	Flakiness Index	FI	%	Independent	Ratio
15	Los Angeles Abrasion	LAA	%	Independent	Ratio
16	Apparent Specific Gravity	ASG	−	Independent	Ratio
17	Bulk Specific Gravity	BSG	−	Independent	Ratio
18	Bulk Specific Gravity (S.S.D basis)	SSD	−	Independent	Ratio
19	Water Absorption	WA	%	Independent	Ratio
20	Sodium Sulphate Soundness Loss	SSSL	%	Independent	Ratio

21	Sand Equivalent	SE	%	Independent	Ratio
22	Clay Lumps	CL	%	Independent	Ratio
23	Unit Weight of Concrete	UW	kg/m^3	Independent	Ratio
24	Concrete Slump	CS	mm	Independent	Ratio
25	Concrete Compressive Strength	CCS	Mpa	Independent	Ratio
26	Marshall Stability	MS	KN	Independent	Ratio
27	Flow	F	mm	Independent	Ratio
28	Theoretical Maximum Specific Gravity	GMM	–	Independent	Ratio
29	Bulk Specific gravity of Compacted Mix (Density of Compacted Mix)	Gmb	g/cm^3	Independent	Ratio
30	Bitumen Content	BC	%	Independent	Ratio
31	Air voids	Va	%	Independent	Ratio
32	Voids in Mineral Aggregate	VMA	%	Independent	Ratio
33	Voids Filled with Asphalt	VFA	%	Independent	Ratio
34	Thickness	Th	mm	Independent	Ratio
35	Age	Ag	yr	Independent	Ratio

Table 2 Project data collection format for secondary data

No	Project Name	TV (AADT)	TL (ESAL)	TR ℃	RF (mm)	DDmax (g/m3)	OMC (%)	LL (%)	PI (%)	CBR (%)	S (%)

No	Project Name	ACV (%)	TPVF (kN)	FI (%)	LAA (%)	ASG (--)	BSG (--)	SSD (--)	WA (%)	SSSL (%)	SE (%)	CL (%)	UWC (Kg/m3)

No	Project Name	CS (mm)	CCS (Mpa)	MS (kN)	F (mm)	GMM (--)	Gmb (g/cm3)	BC (%)	Va (%)	VMA (%)	VFA (%)	Thickness(mm)	Age (Yr)	PCI

Appendix 4

General Instruction to Run Data on Artificial Neural Network (ANN) Model Using Python

1. Training Data

Training data in an Artificial Neural Network (ANN) using the Multi-Layer Perceptron (MLP) algorithm in Python involves several procedures. By following these steps, the ANN model can be effectively trained and evaluated using the prepared dataset.

Step1. Prepare the Dataset

Ensure that the dataset is prepared in CSV file format. The data should be cleaned and preprocessed before analysis. This includes handling missing values, removing inconsistent data, normalizing numerical variables, and encoding categorical variables if necessary.

Step2. Import Required Libraries

Open Python software such as Jupyter Notebook or Google Colab and import the required libraries for data analysis and machine learning.

Example libraries include:

Pandas

NumPy

Scikit-learn

TensorFlow or Keras

Matplotlib

Step3. Load the Dataset

Load the CSV dataset into Python using the Pandas library. The dataset is then divided into input variables (independent variables) and output variables (dependent variable).

Step4. Split the Dataset

Divide the dataset into training data and testing data. Commonly, 70% or 80% of the data is used for training, while the remaining 20% or 30% is used for testing the model performance.

Step5. Normalize the Data

Normalize or standardize the data to improve the performance of the ANN model. This process ensures that all variables are within a similar numerical range.

Step6. Build the ANN Model

Create the Artificial Neural Network model using the Multi-Layer Perceptron (MLP) algorithm. Configure the model parameters such as:

Number of hidden layers

Number of neurons in each hidden layer

Learning rate

Activation function

Number of epochs

Batch size

Step7. Train the Model

Train the ANN model using the training dataset. During training, the model learns the relationship between the input variables and the target variable.

Step8. Test and Evaluate the Model

After training, test the model using the testing dataset. Evaluate the model performance using regression evaluation metrics such as:

Coefficient of Determination (R^2)

Mean Absolute Error (MAE)

Root Mean Square Error (RMSE)

Mean Absolute Percentage Error (MAPE)

Step9. Visualize the Results

Plot graphs to compare the actual values and predicted values. Visualization helps to understand the prediction performance of the ANN model.

Step10. Save the Model

If the model performance is satisfactory, save the trained ANN model for future use and prediction purposes.

2. Validation Data

Cross-validation is used to evaluate the reliability and generalization performance of the ANN model. In Python, cross-validation helps determine how well the model performs on unseen data.

Step1. Prepare and Load the Dataset

Ensure that the dataset is available in CSV format and load it into Python using the Pandas library.

Step2. Import Cross-Validation Libraries

Import machine learning libraries from Scikit-learn required for cross-validation procedures.

Step3. Configure the ANN Model

Define the ANN architecture and configure parameters such as:

Hidden layers

Learning rate

Number of epochs

Activation functions

Optimizer

Step4. Apply Cross-Validation

Use k-fold cross-validation to divide the dataset into several subsets. A common method is 10-fold cross-validation, where the dataset is divided into 10 equal parts. The model is trained using 9 parts and tested using the remaining 1 part. This process is repeated until all parts are used for testing once.

Step5. Train and Validate the Model

Run the ANN model repeatedly for each fold of the cross-validation process.

Step6. Review the Results

Evaluate the average model performance obtained from all folds using:

R^2

MAE

RMSE

MAPE

These evaluation metrics help determine the prediction accuracy and reliability of the ANN model.

3. Exporting Data and Results from Python

The results generated from the ANN model in Python can be exported to Microsoft Word, Excel, or image formats for documentation and presentation purposes.

Step1. Train and Evaluate the ANN Model

Run the ANN model and obtain the evaluation results, prediction values, tables, and graphs.

Step2. Save Tables and Results

Export tables such as predicted values and evaluation metrics into Excel or CSV format using Python libraries.

Step3. Save Figures and Graphs

Save plots and graphical outputs such as actual versus predicted value graphs into image formats such as PNG or JPEG.

Step4. Copy Results into Word

Open Microsoft Word and insert the exported tables, graphs, and model results into the research document.

Step5. Format the Results

Adjust the formatting of tables, graphs, headings, and text to improve readability and presentation quality in the thesis or research report.