ABDULLAHI SAID
ABDULLAHI

A Predictive Modeling of Dynamic Modulus in Asphalt Mixtures: A Machine Learning Approach

NEU
2024

**NEAR EAST UNIVERSITY**
**INSTITUTE OF GRADUATE STUDIES**
**DEPARTMENT OF CIVIL ENGINEERING**

# A Predictive Modeling of Dynamic Modulus in Asphalt Mixtures: A Machine Learning Approach

**M.Sc. THESIS**

**Abdullahi Said ABDULLAHI**

**Nicosia**

**June, 2024**

**NEAR EAST UNIVERSITY**
**INSTITUTE OF GRADUATE STUDIES**
**DEPARTMENT OF CIVIL ENGINEERING**

# A Predictive Modeling of Dynamic Modulus in Asphalt Mixtures: A Machine Learning Approach

**M.Sc. THESIS**

**Abdullahi Said Abdullahi**

**Supervisor**

**Assoc.Prof.Dr. Shaban Ismael Albrka**

**Nicosia**

**June, 2024**

# Approval

We certify that we have read the thesis submitted by **Abdullahi Said Abdullahi titled "A Predictive Modeling of Dynamic Modulus in Asphalt Mixtures: A Machine Learning Approach"** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Educational Sciences.

| Examining Committee | Name-Surname | Signature |
|---|---|---|
| | Assist. Prof. Dr. Mustafa Alas | ..... |
| | Assist. Prof. Dr. Suleiman Ibrahim | ..... |
| | Supervisor: Assoc. Prof. Dr Shaban Ismael Albrka | ..... |

Approved by the Head of the Department

.23./.09./2024

Prof. Dr. Kabir Sadeghi

Prof. Dr. Kabir Sadeghi

Head of Civil Engineering Department.

Approved by the Institute of Graduate Studies

...../...../2024

Prof. Dr. Kemal Hüsnü Can Başer

Head of the Institute of Graduate Studies.

## Declaration

I confirm that all the content, documents, research, and findings included in this thesis have been collected and presented in compliance with the academic guidelines and ethical standards established by the Institute of Graduate Studies at Near East University. Thus, I confirm that I have thoroughly recognized and credited all external sources and data utilized in this work, as per the norms and requirements.

**Abdullahi Said Abdullahi**

03/07/2024

# Acknowledgments

First of all, we are very grateful to Allah the Almighty for giving us the Strength and the courage to complete this thesis. also, I want to thank our parents for their prayers and encouragement. Several people who have been a backbone to complete and without their help and guidance that this thesis would have not been completed. I would like to express our deepest gratitude for helping us during our years as engineering students.

Assoc.Prof.Dr. Shaban Ismael Albrka my advisor has been guiding us through every step of this thesis; he has overseen the progress of thesis, modeling, and proper implementation of design codes. In addition, he was literally making sure that I am on the right track from the very beginning to the very end of the thesis. Thank you for giving us the delight of working as engineers, and being patient for supporting us to achieve the thesis. Thank you for your time and effort.

I want to thank my appreciation to all of the professors and instructors at Near East University for spreading knowledge and offering sincere and valuable support during the course.

Also, I want to thank everyone who had a hand in providing us with what I needed to complete the work from outside the university in different domains, and of course, our families who stood by us all the time and applied all the effort to help us reach where I am now.

# Abstract

**A Predictive Modeling of Dynamic Modulus in Asphalt Mixtures: A Machine Learning Approach**

**Abdullahi Said Abdullahi, Assoc.Prof.Dr. Shaban Ismael Albrka**
**MSc, Department of Civil Engineering, Faculty of Civil and Environmental Engineering, Near East University, Nicosia.**
**June, 2024,**

Machine learning methods have been progressively utilized in the field of civil engineering, namely for forecasting the behavior of asphalt mixes. This study is centered around creating prediction models for the Dynamic Modulus (E*) of asphalt mixes. The Dynamic Modulus is an essential characteristic that impacts the performance of flexible pavements under different loading conditions and temperatures. Traditionally, the task of calculating E* using empirical methods has been lengthy and occasionally inaccurate; hence, there is a want for more advanced techniques. The present study employs data-driven machine learning (ML) methods to improve prediction accuracy, specifically by utilizing random forests, k-nearest neighbors, and extreme gradient boosting (XGBoost) regression. The primary objectives are to understand the essential attributes of E*, assess the predictive capacity of the ML models, and compare their performance in terms of precision, accuracy, and efficiency. The research also emphasizes certain constraints about the effect of external elements, such as differences in materials or climatic conditions, on the potential influence on model performance. Research findings suggest that machine learning models may significantly improve the accuracy of E* forecasts, therefore providing a substantial contribution to the efficiency and longevity of pavement construction. The objective of this project is to enhance the field of pavement engineering through the application of machine learning techniques. The objective is to establish a robust basis for subsequent academic pursuits and real-world implementations in the fields of road maintenance and construction.

**Key Words:** Machine Learning (ML), Dynamic Modulus (E*), Asphalt Mixtures.

# Contents

# LIST OF TABLES

# LIST OF FIGURES

## List of Abbreviations

| | |
|---|---|
| **ANN:** | Artificial Neural Network |
| **KNN:** | K-Nearest Neighbors |
| **E*:** | dynamic Modules |
| **Log E*:** | logarithmic dynamic modules |
| **RMSE:** | Root Mean Square Error |
| **MSE:** | Mean Square Error |
| **MAE:** | Mean Average Error |
| **$R^2$ :** | R (square) |
| **\|G*\|:** | dynamic shear modulus |
| **Va:** | air voids |
| **T:** | Temperature |
| **A:** | aging |
| **Vbeff:** | effective voids in bituminous mixture. |
| **δb:** | phase angle or binder content |
| **r34, r38, r4 & r200:** | aggregate size |
| **Ac:** | asphalt content |
| **VTS:** | viscosity temperature susceptibility |
| **Fc:** | frequency |

# CHAPTER I

# Introduction

## 1.1 Background

Machine learning (ML) algorithms have been more prominent over the past few decades. The primary reason for this is their consistent ability to effectively model intricate and non-linear interactions. Currently, several industries, such as science (Au et al 2022). and more especially civil engineering (Kim & Jung 2024), are increasingly utilizing machine learning (ML) in their approaches to pavement engineering. Data analysis has been an integral part of pavement engineering since its beginning. Machine learning approaches have been used to discover important features and evaluate pavement performance. However, these models have become less successful when used as forecasting tools (Lontara 2020).

Asphalt mixtures are essential for constructing road networks (Zaman 2018). state that the preservation of pavements is important. The complex modulus characterizes the viscous and elastic characteristics of linear viscoelastic materials, consisting both real and imaginary portions ( Kalash and Walczak 2003). discovered that the modulus | E*| represents the magnitude of the Dynamic Modulus (E*). Consistency is an essential characteristic of asphalt mixing. The parameter in question is responsible for determining the performance of flexible pavements under different loading circumstances and temperatures, as outlined in the Mechanistic-Empirical Pavement Design Guide (Xu et al2022).

The dynamic modulus is theoretically defined as the highest value of dynamic shear stress (τ) when combined with the recoverable shear strain, as described by Kalash and Walczak in 2003. Traditionally, the determination of E* values has relied on empirical approaches, which are laborious and require significant effort. Therefore, there is a demand for sophisticated strategies in developing enhanced E* prediction models. Machine learning approaches have increasingly been employed to address common civil engineering challenges in data processing and optimization (Xu et al 2022). This study presents a methodology that uses data analysis to estimate the performance of asphalt mixes. The objective of this research is to thoroughly investigate the Dynamic Modulus (E*). As an illustration, it computes and assesses the predictive capabilities

of five well-known models, including XGB Regression, Random Forest, and KNN K-nearest neighbors.

## 1.2 problem statement of study

Precisely predicting the dynamic modulus (E*) of asphalt mixture is a crucial difficulty in pavement engineering, since it directly affects the design performance and longevity of flexible pavements. Conventional empirical techniques for determining E are time-consuming and lack accuracy, resulting in inefficiencies in pavement design and maintenance. The introduction of machine learning presents an opportunity to improve the accuracy and efficiency of these prediction by employing sophisticated data-driven models. This work aims to overcome the drawbacks of traditional methods by creating and assessing machine learning models, including random forest, K-Nearest (KNN), and XGBoost, to predict the dynamic modulus of asphalt mixes.

### 1.3 objectives of the study

The main aims of this study are:

- To assess and critically analyze how well the 3 models— XGBoost, (KNN) and Rforest.
- The models for the prediction task all work well, such as Random Forest, KNN and XGB regression,
- Comparing their effectiveness, one can determine which of the machine learning methods is at most precision, accuracy, and efficiency.

## 1.4 scope of study

This study applies the above to predict the Dynamic Modulus (E*) of the asphalt mixes the configurator can be used to customize different types of machine learning models. This study then compares the effectiveness of a Random forests, (KNN), XGBoost.

## 1.5 Limitation

In the study, a number of limitations are acknowledged. First, in order to ascertain whether or not they fully serve all forecasts, intricate sophisticated machine learning models may require special attention when used and understood. although the models are not entirely controlled in the study, their performance may be impacted by extraneous variables such as material variability and ambient circumstances.

# CHAPTER II

## Essential ideas and background information of machine learning

### 2.1 Foundational Ideas in Machine Learning

Artificial intelligence (AI) is a potent modeling tool extensively utilized in several scientific disciplines (Krzywanski 2022). Creating mechanisms that allow an algorithm to gather and use data is the focus of machine learning (ML). This learning technique is based on identifying statistical patterns or other data-related aspects rather than requiring consciousness. Because of this, many machine learning algorithms deviate greatly from how humans learn items. Machine learning algorithms can provide useful insights into the different complexity levels present in different learning setups (Oladipupo 2010).

A fundamental idea in education is the model, which holds the knowledge that has been acquired and is employed in forecasting. Models are often created for a specific purpose. In a medical setting, machine learning (ML) may be employed to predict the likelihood of a specific illness. The most crucial idea is model training, which uses information as input to teach the model. Typically, machine learning models are employed for predictions after a single training session (Sivaramakrishnan & Dhandapani 2019). Conventional machine learning falls into one of two categories based on an algorithm's ability to improve its prediction accuracy with training. Supervised Learning These methods make use of the target data as well as a particular set of pre-labeled input variables (training data). Using input variables, it develops a mapping function to connect inputs with the required outputs. Adjust the algorithm's parameters and modification strategies are continued until the system shows an acceptable level of accuracy with respect to the training data (Fawzy & Jasem2020).

### 2.1.1 Oriented Learning

These algorithms work on a particular set of pre-labeled input variables called a training data other than the target data. It develops a mapping system to relate inputs with the needed to use the input variables to generate results. The configurations and of the algorithm stay the same until the system shows an acceptable degree of truth of the training data (Fawzi & Jassem 2020).

Supervised learning methods require a substantial quantity of labeled training data in order to produce models with improved prediction performance (Balanchine 2013). The regression algorithm seeks to identify the most optimal function that closely aligns with the data points in the training dataset. Regression-based approaches may be categorized into three main groups: Polynomial regression, multiple regression, and linear regression. Classification algorithms provide correct class selection for given data by assigning each input to the relevant class. The prediction function generates a categorical result, which is assigned a value corresponding to one of the specified classes (Directions 2023). Regression analysis is employed to address problems involving the prediction of continuous variables, whereas support vector machines (SVMs) are utilized for algorithmic classification. The Rforest approach is used to address regression and classification problems (Directions 2023).

Supervised learning has been widely used in the engineering sectors, including civil engineering. Supervised learning has a variety of uses in civil engineering. For example, it is used to examine sensor data from different infrastructures, such buildings and bridges, in structural health monitoring. Its main objective is to assess the structure's health by identifying deterioration and forecasting structural breakdowns. Supervised learning also has the potential to forecast the seismic collapse of frame constructions. Early warnings of possible problems are made possible by supervised learning algorithms that are taught to recognize patterns in sensor data utilizing vibration data, lamb waves, electromechanical impedance, acoustic emission, etc. (Amezquita & Sanchezetal2020). learning is employed (Directions,2023). Moreover, (Mein 2023). state that supervised learning is used to forecast the characteristics material of the buildings, such as the strength of concrete. One of crucial aspects of building project management is supervised learning. It is widely used in many construction-related fields, such as on-site operation monitoring,

safety, and bridge inspection (Xu et al 2021). In addition, soil composition data and past performance data are analyzed using supervised learning models in geotechnical engineering to provide predictions for soil settlement, slope stability, shallow and piling foundations, and other geotechnical factors (Bahmanietal 2022).

### 2.1.2UnmonitoredEducation

In contrast to supervised learning, there is no definitive or erroneous answer associated with each input item. Unsupervised learning is characterized by a higher level of unpredictability compared to supervised learning in the context of acquiring knowledge. Unsupervised learning seeks to uncover the inherent structure or distribution pattern of the data in order to get a more profound understanding of it. As the algorithm learns on its own hence strives for the most accurate representation of a particular pattern that it detected in the input space. while incorporating it into the broader input pattern framework. Consequently, the extracted Qualities of each input item are categorized and linked to multiple inputs. These association and It is also with this kind of problems on clustering where unsupervised learning applies the solution. For example, unsupervised learning can be used to analyses topographical mapping and terrain. jobs in civil engineering. To handle elevation data collected by remote sensing devices such as light detection and ranging (LiDAR), for example. According to Garnierite (et al 2022). this study helps to produce detailed topographic maps, a very critical requirement of most civil engineering applications Like site planning and a useful tool for leveling applications.

Unsupervised learning is also designed in the field of computer engineering, with tons of applications. For instance, it could be used within systems or hardware for the detection of anomalies like: memory failure in the CPU, or it can still be used to recognize abnormal activities or security breach in Human computational networks and systems (Usama  2019).

### 2.1.3 Reinforcement-Based Learning

This method uses a strong computer to map actions to certain decisions, producing reward or feedback signals as a consequence. Using incentives and penalties based on

prior experiences, the system automatically teaches itself to identify significant positive behaviors (Fawzy&Jasem,2020).

Learning agents have predetermined goals and are somewhat able to sense the state of their surroundings. As a result, people are empowered to act to alter the environment and get closer to their designated goals. On the other hand, information is obtained by direct interaction with the issue environment in reinforcement learning (Directions, 2023).

The field of civil engineering makes extensive use of reinforcement learning. Using real-time data from loop detectors, vehicle counts, and traffic cameras, reinforcement learning systems can regulate traffic flow and control traffic lights (Tan 2022). That is the technique that is also necessitated driverless cars. Thanks to it, they can recognize streets, navigate turns, and even make enables them to make informed decisions about their path (Directions 2023). Reinforcement learning in the Geotechnical engineering can help in optimizing the process of drilling and excavation by the approach consisted of using geotechnical equipment data, such as soil samples or geophysical sensors (Coelho and Smyrniou 2023). It has also been applied to a vast application range, in network In Cellular Data Analytics, Sensor Energy Management, Mobile Application Predictions, and Building Energy Optimization is an important issue in which reinforcement learning is used in irrigation control. Many advantages, such as improved agricultural production and quality, reduced energy expenses, and a reduced These applications can produce an impact on the environment (Ding & Du 2022).

Reinforcement learning is also a technique that is used in virtually every technical field, for example, computer engineering so that robots can learn, improve, adapt, and repeat tasks using specific Instruments such as robotic arms, grippers, cameras, and torque sensors are being used for that purpose now (Kormus).

### 2.1.4 Teaching under Partial Super

The objective of a learning process is to utilize both labeled and unlabeled data in order to achieve certain objectives. Semi-supervised learning is a specific topic within the

broader field of machine learning. According to van (Angeleno and Hoops 2020), it is positioned in a theoretical middle ground between supervised and unsupervised learning.

Machine learning, often known as ML, commonly differentiates between two essential tasks: supervised learning and unsupervised learning. Supervised learning involves the provision of a dataset containing input values (x) and their matching output values (y). The objective of this job is to develop a classifier or regressor that can predict the output value for inputs that have not been previously encountered. In unsupervised learning, an output value is not provided. Rather, people try to extract an underlying framework from the provided information. The goal of unsupervised aggregation is to create a mapping between groups and the inputs (real number vectors, for example) so that similar inputs are grouped together. A branch of machine learning called semi-supervised learning aims to combine these two objectives. By utilizing knowledge that is frequently linked to one of these two tasks, semi-supervised learning algorithms seek to improve performance in either of them in support of categorization when dealing with a classification Inclusion of extra data points with unknown labels may further make it difficult. On the other hand, by making use of the knowledge that certain data points belong to the same class, clustering algorithms would help improve learning (van Angeleno & Hoops 2020). Some of the most common strategies for semi supervised learning are expectation maximization (EM) with generative mixture models and self-supervised learning Transudative support vector machines and graph-based algorithms have also been presented by ( Zhan 2009). in order to emphasize a category of pavement condition-evaluation variables: apparent surface stress, paved layer strength, and pavement roughness. Building the prediction model involves combining labeled data from hand surveys with unlabeled data from many, accelerometer, surface testers, lasers, and road pictures. The model can also be trained further with labeled data to enhance the results, though unlabeled data from the same training dataset can also enhance the model (Liu  2021).

in addition to the recognition related to high-quality water and determination of likeliness of contamination Semi-supervised learning can be used for monitoring and detection in water treatment supply (Yuan & Jia 2016).

### 2.1.5 learning transfer

This perspective views transfer learning as a methodology for recycling machine-learning models that were previously educated about newly manufactured models. General information can only be shared among models if they are meant to achieve the same goal. The machine-learning-based method reduces the amount of resources and annotated data required for training a new model (Weiss et al 2016).

Transfer learning is used that utilizes knowledge from the same but other domain to train a make them human-like subjects before they mature in one. Because there are very little target training data, transfer learning is such a situation may be necessary for any one of a multitude of reasons, not the least of which is that the data are actually out, the related costs, or the high cost of inaccessibility. The proliferation of the availability of enormous data repositories makes transfer learning solutions attractive when the aim is to use pre-existing sets of loosely but significantly related content datasets to address a topic of interest. For instance, Transfer learning has proven to be quite useful for many machine learning applications, be it in text sentiment categorization (Weiss et al 2016). It has also found successful use in the areas of contexts from image classification and human behavior to software defects and multilingual text.

Using earlier pretrained models instead of starting completely from scratch is known as transfer learning. As an illustration, the pre-trained object identification or video analysis models can be fine-tuned further to advance safety monitoring of construction sites. This modification entails being aware potential hazards and ensure that safety protocols are being observed. In addition, there is recognition models to verify if workers are wearing helmets and safety vests properly informing employees of any hazardous circumstances (Lee & Lee 2023).

Pre-trained models for smart traffic light management, congestion monitoring, and traffic flow Prediction can be adapted for the traffic management and optimization using transfer learning (Krishnakumar 2018). By reusing models or data from past BIM projects, transfer learning can aid in accelerating the speed with which the digital representations for future construction or renovation projects (Zabin et al 2022). Transfer learning that includes reuse of learned models prior to Design Rule Checking

or Fault Analysis, can speed up the development and verification process of integrated circuits and hardware components in Computer Engineering ( Pan and colleagues, 2021).

## 2.2 Machine Learning's Evolution

### 2.2.1 The Earlier Times

The foundational research on artificial intelligence was done in 1943 by Warren McCulloch and Walter Pitts. They used a variety of theories about computing, logic, and brain function to build the prototype for artificial neurons. This is an emulator of a neuron, marked "on" and "off," that reacts to input from nearby neurons. McCulloch and Pitts showed that networks of these artificial neurons could calculate any computable function. They also demonstrated how these structures may be used to create logical operations such as AND, OR, and NOT. What's more, they proposed that these kinds of networks may be taught. Hebbian learning—a guideline for modifying the connections between neurons—was first presented by Donald Hebb in 1949 and continues to have a significant impact on neural network research. This device replicated a network of forty neurons using vacuum tubes and parts taken from a B-24 aircraft. Even though his doctoral committee was first skeptical, Minsky continued researching universal computing in neural networks. Alan Turing had a significant impact in shaping AI starting in 1947. Important ideas including machine learning, reinforcement learning, genetic algorithms, and the Turing Test were first introduced in Alan Turing's 1950 paper "Computing Machinery and Intelligence". He proposed the "Child Programmed," which calls for a change in focus in AI research from simulating adult intelligence to simulating a child's learning process (Huang 2010).

### 2.2.2 Deep Learning's Ascent

The Turing Test the creation of the first computer game of checkers, and the emergence of artificial intelligence in the 1950s lay the groundwork for further developments. The first industrial robot, Perceptron, Decision Trees, and the Chain Rule Method—all

groundbreaking ideas that are now fundamental to machine learning—were unveiled in the 1960s. Neural networks and pattern recognition were advanced in the 1970s with the introduction of key Methods like KNN, backpropagation, and the Recognition. Deeper network designs were made possible in the 1980s with the development of Artificial Neural Networks (ANNs), the restricted Boltzmann machine (RBM), Explanation-Based Learning, and the Backpropagation Algorithm. behavior-based robotics (BBR) first appeared in the 1990s, and Deep Blue showed how artificial intelligence (AI) could beat human champions in challenging games. Large-scale datasets like ImageNet were created in the 2000s, which fueled the deep learning revolution. Other developments like the Deep belief network, Deep Boltzmann machine, and Deep Neural Networks (DNNs) also represented a crucial turning point. Game-changing models like Alex Net, Generative Adversarial Networks (GANs), U-Net for medical image segmentation, and the legendary AlphaGo's triumph against human Go champions emerged in the 2010s. With the release of Denoising Auto Encoders (DAE), Nash learning, breakthroughs in reinforcement learning, and Deep inverse approaches in 2020, deep learning continues to develop and shape 's future AI (Directions 2023).

### 2.2.3 Current Developments

The tremendous advancements in digital technology for cyber-physical systems in recent years have presented significant challenges for applications in academia, industry, and services. As a result of the widespread IoT use there is a lot of data, noisy interference, incomplete and inconsistent information, and complicated data with several dimensions. Artificial intelligence models based on machine learning (ML) have become very effective instruments for process improvement and data analytics in a variety of scientific domains. Since 2012, machine learning (ML) technologies have developed, demonstrating their usefulness in resolving challenging industrial issues. Applications cover a broad spectrum of purposes, including as process optimization, job planning, improvement of quality. To identify and resolve inefficiencies in various activities, machine learning is a cutting-edge technology that is widely used in general science, medical, manufacturing, and service (Wang 2022).

## 2.3 Customary Approaches in Civil Engineering

### 2.3.1 Appraisal Techniques

In civil engineering, empirical techniques make conclusions and build structures based on historical facts, tests, and observations. These techniques, which are frequently grounded in real-world application, are especially helpful when working with complicated, non-linear, or variable situations or when theoretical knowledge is lacking. In civil engineering, some instances of empirical methods include: assessing the strength and stability of materials and structures through load testing; evaluating the performance of structures through field testing and monitoring; and estimating variables such as soil bearing capacity, concrete strength, and structural stability using empirical equations and formulas (Jafari & Constantin ides 2022).

### 2.3.2 Methods of Analysis

In civil engineering, analytical approaches include the analysis and design of structures using mathematical theories and equations. These techniques essentially rest on mathematical, mechanical, and physical concepts. Analytical techniques are used by engineers to ascertain how constructions will behave under different loads and circumstances. Typical analytical techniques include the following: structural analysis, which computes forces, stresses, and deformations in structures using the concepts of statics and dynamics. In order to design water and wastewater systems, including pipelines, pumps, and channels, hydraulic and fluid dynamics calculations are used. Geotechnical analysis is used to analyze soil qualities and forecast slope stability, bearing capacity, and settlement (Jafari & Constantin ides, 2022).

### 2.3.3 Algebraic Techniques

In civil engineering, numerical methods refer to the application of numerical techniques and computers to solve complicated engineering issues. These techniques are especially useful for handling complex geometries or non-linear dynamics. Among the numerical techniques are computational fluid dynamics (CFD), which is used to model fluid flow in hydraulic structures or environmental assessments, numerical

modeling of traffic flow and transportation systems for road network and traffic management optimization, and structural design software, which uses numerical techniques to ascertain the dimensions and specifications of structural components. (Jafari & Constantin ides 2022).

## 2.4 General Engineering Usefulness

A subfield of artificial intelligence called machine learning is gaining popularity in the data analysis and computing industries. It entails creating models and algorithms that let programs behave intelligently. This is where exploratory data is entered in detail into the machine learning process processing. Designing an intelligent real-time system depends largely on machine learning models. engineering applications that can process the acquired data and solve practical problems emerging from these data (Haveri 2024).

### 2.4.1 Applications of Machine Learning in Other Engineering Fields

Machine Learning is very critical to computer engineering as it equips autonomous robots. This allows them to see and consider the surroundings, which make them act upon decisions that are thoroughly thought of. Machine learning techniques also give birth to understanding robots. human language and gesture and act in response to it (Martinez 2020). planning, machinery and equipment. Mechanical engineering also employs machine learning to make accurate predictions in maintenance planning for machinery and equipment, with the objective of minimizing Downtime and ensures efficiency. Machine learning makes certain to be able to use machine make new, smarter mechanical parts and systems designed with the help of learning algorithms by considering a wide range of limiting limitations and criteria by performance (Guo 2021). Machine learning is applied in electrical engineering with the capability to predict demands, Optimal distribution network development, and grid fault diagnosis. Chemical engineers Utilize machine learning techniques to optimize performance and control process chemicals reactivity for bettering product quality and reducing the use of energy. Machine learning algorithms identify abnormalities within electrical systems, which is very essential in ensuring that the detection of these faults and hence the associated hazards with these defects have been detected in a real-time basis

(Prajwal 2021). Machine learning is used to predict the properties and interactions of chemical compounds, which facilitates drug development techniques (Gao 2022). Data from a range of sensors used to monitor and assess air quality, water quality, and other environmental parameters may be analyzed using machine learning models in environmental engineering. It is also used for life cycle analysis, enhancing treatment methods, locating and describing pollution sources, and evaluating the condition of wastewater and water systems. It is promising that machine learning models may be used to predict natural disasters like hurricanes and earthquakes since doing so would significantly enhance preparedness and response to disasters (Zhong 2021). This is because material engineers are always developing new materials that call for certain characteristics that machine learning (ML) can predict with accuracy. This industry benefits from the use of machine learning (ML) in producing very complex content, as the graphs that were produced made clear. To make sure that the materials are as specified, machine learning may also be helpful in the quality control inspection of industrial operations (Stergiou 2023). argue that mining activities, material Machine learning is used in a design field, in addition to materials characterization and even job prediction. Additionally, in aerospace engineering, the best flight control systems can be better made by. machine learning (Caliche 2023). say, "Machine learning models can estimate" needs of maintenance through the processing of data from in-service sensors installed on board aircraft.

### 2..4.2 Machine learning in civil engineering

Through the incorporation of machine intelligence into a wide range of applications and procedures, the discipline of civil engineering is undergoing a dramatic transformation. learning (ML), to increase accuracy and productivity in different fields. ML Algorithms In structural engineering for detecting structural decay and for load-bearing prediction capacity for quantitative predictions using sensor data. Machine learning models in the field of pavement engineering are used in the prediction of pavement deterioration, which will in turn assist in the formulation of effective maintenance techniques and lifespans. Further, machine learning (ML) aids in predicting material characteristics and helps develop high performance materials by analyzing large datasets to enable optimization of compositions to have the desired

features. Such developments save costs and time, and at the same time improve the degree of reliability and safety in civil engineering Projects.

### 2.4.3 Road Surface Predictive Maintenance

Machine learning is applied in predicting the pavement that has become worn down and, in the improvement, pertaining to upkeep schedules in the field of predictive maintenance. Engineers who are responsible for doing extensive data analysis on information acquired from a variety of sources, including the Long-Term Pavement Performance (LTPP) database, for example, can predict rutting caused by the repetitive traffic environmental factors. Metrics such as RMSE and R-squared is a good indicator of strong predictions Gaussian Process Regression has shown time and again with its ability to be outstandingly accurate. This approach reduces the cost for maintenance and improves roads' safety and performance by thereby making repair operations fast and inexpensive.

A study in 2023 developed rut prediction models using the LTPP data using a variety of machine learning techniques the following techniques were utilized Ground Penetrating Radar (GPRs), Support Vector Machines, and Regression Trees, GPRs being the greater accuracy for predicting rut depth. The best model had an $R^2$ value of 0.70 and an RMSE of 1.96. More importantly, transportation authorities can now properly schedule maintenance works, hence ensuring that resources are allocated in an optimized manner. This is only possible with the use of this excellent degree of Accuracy.

The application of such algorithms in machine learning predictive maintenance has the potential of delivering a Proactive approach to pavement management. To strictly enhance the life and utmost it is recommended to plan proactively for maintenance of pavements to prevent infliction. This can happen through the correct predication of time and place of pavement wear. This reduces costs simultaneously with its advantage of lower road closures and Increasing user comfort and safety.

### 2.4.4 Pavement condition monitoring

This way, AI and machine learning technology have greatly improved the assessment. and surveillance of pavement problems. These technologies make exact, effective, Full

road inspection is very important in any maintenance and repair requirements concerning the road surface. Traditional pavement monitoring has been undertaken by laborious, time-consuming methods that are prone to human error—primarily visual inspection. Modern automated systems have embraced these techniques various monitoring, and continuous machine learning sensor devices of the roads. The systems collect real-time information on pavement failure through the application Data from intelligent sensors deployed in vehicles, drones, and stationary platforms. It is primarily for this reason that convolutional neural networks (CNNs) can be applied to interpret this data such that, by Identify and classify many kinds of roadbed damage, including cracks, potholes, Rutting. Analytical techniques for high-resolution pavement photographs and video image processing based on artificial intelligence have become some of the recent applications. This, in a way, will allow the use of timely and effective inspections by correctly identifying surface abnormalities. Also, the GPS data help in identifying This thus pinpoints the major source of failure to the reported issues hence allowing focused maintenance efforts Operations (MDPI).
latest cutting-edge approach utilizes data collected from single-lane inspection of the applying multiple linear regression (MLR) models and artificial neural network (ANN) models to Able to predict pavement conditions over multi-lanes, this enormously decreases the time and financial resources needed for a detailed road inspection. By carrying out a criteria-based analysis, including traffic direction, location and lane features these models can be able to predict correctly condition of the whole road. It makes the maintenance process more efficient. (MDPI). Furthermore, the prediction capabilities of machine learning models are enhanced by including Different data sources including traffic patterns, environmental conditions, and historical maintenance records. The ability of models to predict enables proactive planning of maintenance on the rate of pavement deterioration under different traffic loading and weather conditions. It makes the pavement more durable and sustainable, and it also helps ensure that optimum allocation of the maintenance budgets is achieved. and Resources.

### 2.5.1 Pavement Material Characterization

Machine learning applied to pavement material analysis makes our understanding performance in different conditions and thus allow to develop more efficient

approaches to road construction and maintenance. Using modern methods that include machine learning methods and multi-scale numerical simulations, the scientists can enhance their Ability to predict accurately the behavior of materials. Performing numerical simulations at various scales is important in properly characterizing pavement materials. The simulations encompass a wide range of sizes, from miniscule and nanoscale features to large-scale structural responses. Macroscopically, In the process of front-end design, simulations may be utilized to do an analysis of the structural and dynamic response of a whole pavement system to traffic loading. It is possible to examine the mechanical characteristics of asphalt mixes at the meso level through the use of simulations. In this regard, an inquiry is carried out into the action of stones and asphalt binding agents at small and tiny levels.

The applications of the growing techniques in machine learning, e.g artificial neural networks (ANNs) to predict complex relationships among a host of material properties and performance results. These models can effectively handle vast datasets derived from field observations and laboratory tests to make precise predictions on parameters such as fatigue life, strength, durability and many more. For example, the Artificial Neural Networks (ANNs) can also be trained with past performance data to predict how pavement materials will function in the future, considering different traffic and environmental conditions. By enhancing material composition and construction techniques, it is feasible to design pavements that have increased durability and reduced construction costs. This can be achieved by using the Predictive Performance of MDPI. Further improvement of prediction accuracy in these models can be achieved by using data from another sources, such as the historical maintenance record, traffic patterns, and environmental factors. This will ensure that repairs are carried out in proactive method rather than the reactive method. That approach makes certain an optimal utilization of budgets and maintenance resources but at the same time extend the life of the pavement.

### 2.5.2 Traffic Pattern Analysis

Machine learning for transport involves running many applications like data analytics, road problem detection, congestion forecasting, and predictive maintenance. However,

them Importance of emergency (Silva ML) is further enhanced in this domain as it facilitates effective-routing of traffic during (et al 2020).

The traffic pattern study is significant for the development of robust and sustainable pavement systems. Surely, engineers accomplish the task of saving the performance of pavements and their life span through proper design optimization. based on a rigorous understanding of traffic patterns and loads. To be able to predict pavement behavior under diverse traffic conditions, and also to analyze complex data, machine Learning has made tremendous strides in this field. There are several of the growing requirements of the methodologies proposed for the technologies in traffic prediction in the form of predicting congestion scenarios. In this respect, various machine learning models have been employed so far, including Neural Networks (NNs), Support Vector Machines (SVMs), and regression analysis, in order to Predict traffic congestion, as they can handle large data sets very well in effective learning extract patterns out of data. For example, Yusheng et al. developed a new deep-learning framework to predict the traffic flow in varied urban road networks basing on recent developments in However, this technique is aware of the way traffic flow moves and shifts in time and space. They used a stacked autoencoder (SAE) model and analyzed its performance relative to other common models, such as neural networks or support vector machines. They measured each model's performance using three criteria: MAE, RMSE, and MRE. On short-term predictions (15 minutes), they have reported a good performance within their new model with an RMSE value equal to 50, however, it doesn’t perform well in the long term where this error increases towards a score of 138 (Yasir et al 2022). used weather information like rain, humidity, and temperature to help predict traffic jams. They started with a complex model that looked at 54 different factors. Then they simplified their model by removing less important factors, ending up with 10 key variables. Six of these variables were specific days of the week, and four were related to the weather. This approach was pretty accurate to 75.5% of the cases. However, they missed to include an important Detail: Also, the time of the day plays an important role when it comes to traffic, which is very different (Yasir et al., 2022). The one that Akbar et al. used was a mashup of complex event processing and machine learning. approaches to predict when and where congestion will be 2. A formula called Yasir et al. presented Adaptive Moving Window Regression in 2022.

Pavement design traditionally considered as conventional load analysis along with average daily traffic loads and sometimes underperforms. However, with the advancement of machine learning and now it is possible for conducting a much comprehensive analysis. Models that simulate the the interactive investigation of dynamic responses of pavements in actual traffic scenarios between a car, its types and the road. Employing these models permits for enhancement accuracy of pavement performance prediction by considering factors like axle design vehicle speed, and surface roughness. By incorporating these comprehensive evaluations into pavement design, better able to predict and address problems such as fatigue cracking and rutting (MDPI). These approaches enable appropriate planning of maintenance and rehabilitation programs through the evaluation of effect of different traffic patterns on pavement deterioration. Machine-learning methods can Predict pavement life under different traffic loads by using weigh-in-motion (WIM). data. This feature assists transportation authorities in ensuring the most effective scheduling jobs and extending the life of their road infrastructure (MDPI, RUcore).

Further, it is such machine-learning models that provide valuable information on the key elements that pavement performance and that enable transportation agencies to, therefore, more effectively allocate them resources and prioritize maintenance jobs. Random survival forest models are used to obtain a deeper understanding of how the pavement will last based on traffic load. This adds to minimizes unforeseen pavement problems and hence allows for planning of repair activities. Adopting such a proactive strategy in pavement management does not only Improves pavement durability and safety while reducing maintenance Costs (MDPI, RU).

### 2.5.3 Improving Traffic Safety

To improve traffic safety, it's important to predict the road hazards using an intensive research. Machine learning (ML) methods have played a widespread role in other disciplines of safety prevention and risk identification. This include operations such as mitigating fraudulent traffic, preventing the unauthorized channeling of data in Internet of Things networks to ensure information security, and enhancing transportation safety. More importantly, ML is much more flexible than classical statistical methods and does not require strict prior assumptions on the bivariate

relationship between independent and dependent variables. ML approaches have shown good accuracy in road safety modeling, especially with the appearance of high dimensional big datasets that serve as a basis for a description of this research area in the field of traffic safety. Through the algorithms, risk prediction models were implemented such as neural networks, support vector machines (SVMs), and random forests (RFs). However, the challenge of interpretability is unveiled due to the absence of an explanation of the inner causal relationship in the black box, which is revealed in complex machine learning models. Consequently, this hampers their usefulness in evaluations. Explaining tree models such as LightGBM, AdaBoost, and eXtreme Gradient Boosting (XGBoost) have been explored in certain studies using Shapley Additive explanation (SHAP) technology. All can be used to look into multiple factors and identifications of contributors concerning traffic safety (Qi et al 2022).

### 2.5.4 Detection of Structural Damage

All engineering structures tend to decompose and decay internally over a period of time. The detection of damage is an essential part of the work principles, since it provides opportunity to assess damages as soon as possible. This, in turn, increases security and provides control over the modern with high performance and reliability. The goal of machine learning in (SHM) is to produce models or representations that create a relationship between patterns generated from sensor data and targets for damage assessment at diverse levels. While common machine learning models are effective, they cannot efficiently interpret huge volumes of unorganized sensor information. Therefore, the process often requires detailed engineering and expert knowledge to pull out elements from raw data that hint at damage. Retrieved features are finally fed into a fitting machine-learning model (Yuan 2020). One approach is to teach a neural network how to distinguish between the frequency responses of an intact structure and those of structures with varied levels of damage. Afterward of training the neural network, it has been able to identify every precise damage and its position and intensity of it. (Fang 2005) On the basis of the deep learning approach in the Structural Health Monitoring (SHM) applications – an assessment of the algorithm implementation has pointed a great potential to develop the end-to-end systems reviewing the algorithms without a need for great prices in preliminary signal

processing. These deep learning models can be tailored to various SHM tasks: damage detection, concentrated-damage, and range of the injury. As a result, many neural network architectures have been investigated for SHM such as Millpond RNN and CNN. As in the case of most other networks, the hyperparameters used by this network are usually determined through techniques that include grid search and random choice. The use of MLPs can be exemplified by one application, the identification of defective rotation machinery components in SHM. A One-Dimensional Conventional Neural Network (1DCNN) is a deep learning architecture specifically tailored for the processing of time series data. In its perspective, the convolution processes are applied on sequences of data points with a view of extracting informative characteristics. This makes it a perfect tool where one is required to analyze sequentially organized data in SHM. Recurrent Neural Networks (RNN) are designed to find patterns in time and thus an effective tool of analysis when one is analyzing sensor measurements. However, standard RNNs had a problem of performing on long sequential data with challenges such as gradient explosion or vanishing. In solving these challenges, researchers came up with custom RNN structures like LSTM and GRU. The LSTM and GRU models have great capacities in handling long-range associations among variates within sequential sensor data, hence are very applicable for the case of Structural Health Monitoring (SHM) (Dang 2021).

### 2.5.5 Prediction of Structural Failure

Machine learning (ML) helps in structural failure prediction, data is used to make sense of it and understand the patterns used in the prediction of problems with structures, predict the best time for any maintenance work, additionally to detect its risk assessment (Zaprorid 2023). However, the inputs for predicting the response of structures are conditions at which loads occur and parameters associated with structure composition include characteristics of loadings, geometry as boundary-conditioned factors material's properties. The results are the field variables that we want to identify, which include displacements, stresses, and strains. In the research by the extreme learning machine was used for modeling both the load-carrying capability and the failure manner of a beam-column joint connection are taken into consideration. Li et al. applied NN and various conventional ML approaches to predict the consequences of gas blows (Li et al 2023).

# CHAPTER III

# Methodology

## 3.1 OVERVIEW

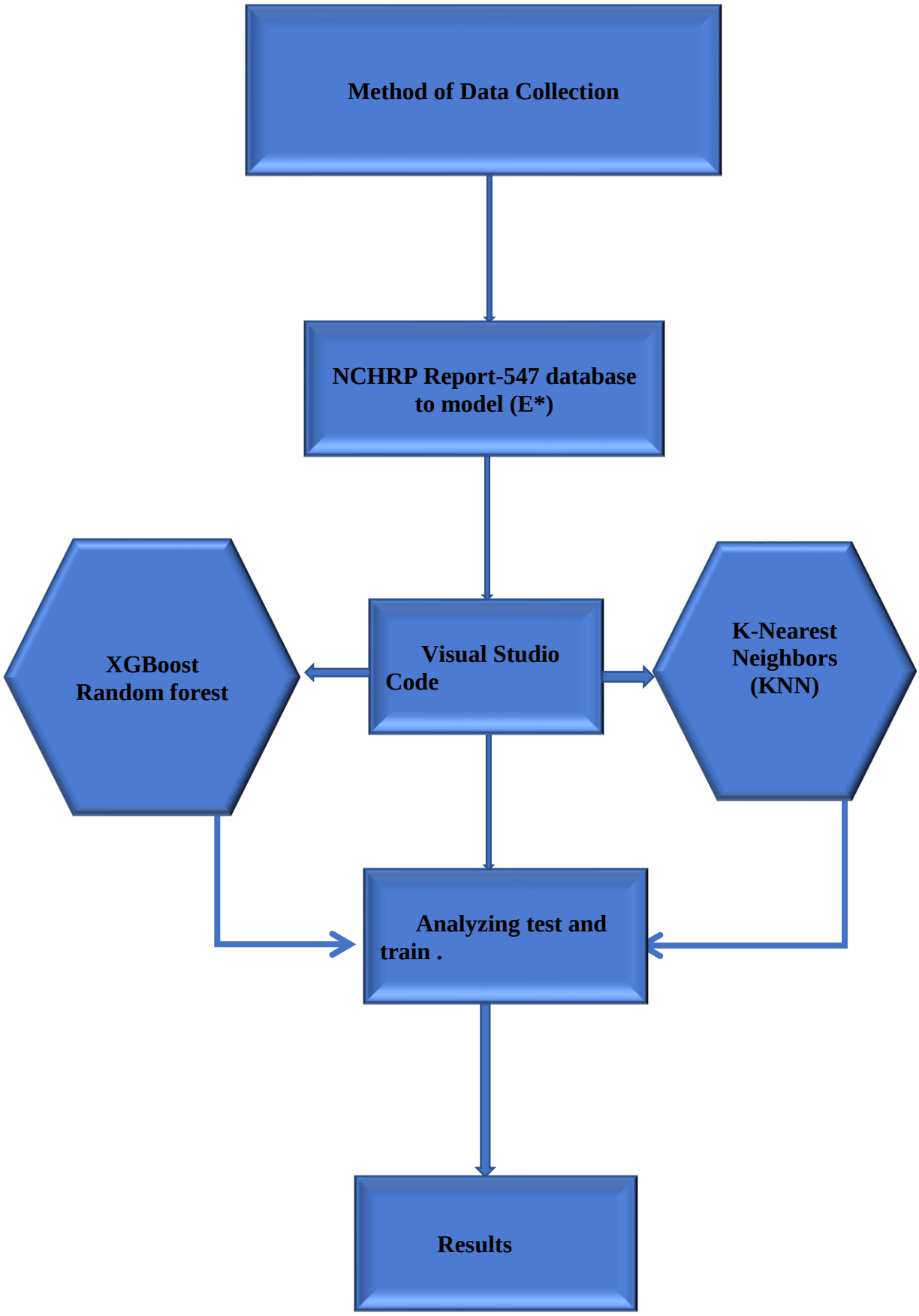
This chapter describes the methodologies employed in this thesis, including machine learning models, feature selection, and their significance.

## 3.2 Data Collection

Data collection starts with sourcing information available on the NCHRP Report-547 database, which carries some essential parameters needed for modeling. The data collected thus forms the basis for analysis and is used in modeling the parameter dynamic modulus (E*) The data is processed in one of the coding environments familiar to the modern computer world, Visual Studio Code, and analyzed. This is implemented by putting in the three models, including XGBoost, RForest, (KNN), and ANN The powers in which these models are built come in with the typologies in hand and are either way pronounced from them to derive new solid, reliable predictions about the outcome. For the purpose of carrying out a job that involves the assessment of models, the data will be separated once more into training data and testing data. The models will be trained on the data from the training set, and then they will be evaluated on the other testing set. the results are then analyzed and tallied to collect accuracy and other performance metrics. The analysis finally concludes with the critical insights and conclusions based on the data from the modeled methods. dynamic modulus (E*) specifically measures the viscoelastic properties of asphalt mixture which in turn affects the overall performance of asphalt pavements. This dataset refers to asphalt mixture properties in asphalt pavements. These parameters are essentially required to understand the mechanical behavior of mixtures, where the visco-elastic properties are among those that will influence the performance of pavements the most.

**Figure 1**

**Follow chart of methodology**



### 3.3 Input Parameters

**r34, r38, r4, r200:** Sieving data that indicates the gradation of the aggregate in the asphalt mix. **A (aging):** Represents the aging condition of the asphalt mixture **AC:** Asphalt content, indicating how much bitumen is present in the mix**. Va:** Air voids, showing the void space in the compacted mixture. **Vbeff:** Effective bitumen content, representing the bitumen that effectively coats the aggregates. **T:** Temperature at which the properties are measured. **fc:** Loading frequency, relevant for dynamic testing. **A and VTS:** Parameters related to the bitumen's properties, such as aging and viscosity changes with temperature. **db:** Bulk density of the mix, which can affect performance. **|G|*:** Complex shear modulus, a measure of the mixture's resistance to shear deformation. **(E*)**: dynamic modulus of asphalt mixture.

Each parameter in the dataset plays a crucial role in determining the performance characteristics of asphalt mixtures used in asphalt pavements. By understanding and accurately measuring these parameters, engineers can design asphalt mixes that balance durability, flexibility, and resistance to various forms of distress, ensuring long-lasting and reliable pavement structures. The combination of these parameters helps in modeling the dynamic modulus (|E*|), which is essential for predicting how the material will perform under different traffic and environmental conditions.

## 3.4 Machine Learning Model

a machine learning model is built through an orderly process that starts with the collection of relevant data, then cleaning, and transformation for analysis. Then one can go ahead to select a proper algorithm based on the task at hand i.e., classification versus regression or clustering. A model is trained with some fraction of data so that it comes to learn about the patterns and relationships. Then, it is tested and validated over another subset of data to verify how efficiently the model is working. Now, a well-scoring model can be put into real-world practice where it applies its learned patterns to new data which never saw before and makes predictions or decisions about it. Almost always for the accuracy of model over time continuous monitoring as well as updating is needed since with constant flow of data in most real-world applications many things change out there.

### 3.4.1 XGBoost

The software includes an integrated functionality that accelerates the training process while working with a substantial dataset. The classification and regression type modeling challenges are resolved using a concatenated gradient-boosting framework technique. In the XGBoost algorithm, a less powerful base learner (decision tree) is incorporated with additional learners and is sufficiently effective in rectifying the errors generated by previous models during prediction. XGBoost enhances accuracy by employing a differentiable loss function known as mean squared error and an optimization technique called gradient descent. The approach involves the utilization of the first and second-order terms from the Taylor's series expansions. In order to manage the intricacy of the expanded word, a regularization term is incorporated. The regularization term controls the occurrence of over-fitting and helps to balance the final learning weight w. XGBoost use fitting approaches to discover and predict the underlying pattern in order to reduce overfitting. It achieves this by using sampling methods for both rows and columns. The technique effectively combines parallel processing (node division) and distributed computing (multi-threading) to demonstrate the practicality of the fastest model (Prakash 2023).

### 3.4.2 Random forest

Forest is a highly favored approach in the field of ensemble machine learning and data science. The RF algorithm effectively addresses classification and regression problems due to its fundamental structure and minimal computational demands. This combination of such weaker models as decision trees creates a robust, complete model. In Random Forest technique, there are decision trees that are being trained through a random subset of characteristics and data points. It is applicable to complicated datasets involving both categorical and continuous variables centred around specific tasks. As an ensembling tree model it uses the bagging principle in parallel and boosting principle in consecutive mode to train the basic learner. Random Forest reduces overfitting, problem of decision trees. Balancing of bias and volatility is through averaging of several tree projections to give results that are more reliable and better. Random Forest is one of the most powerful machine learning tools because this algorithm behaves proper under different conditions and it randomly chooses data,

along with attributes. Therefore, this makes the model apparent all across data specific disciplines (Prakash 2023).

### 3.4.3 K-Nearest Neighbors (KNN)

The KNN algorithm is a simple and efficient technique. the algorithm functions by determining the 'k' nearest data points to a given query point, utilizing a designated distance measure, such as Euclidean distance. The method assigns the most frequent class among these neighbors to the query location for categorization.

KNN is a non-parametric method that uses instances to learn. The straightforwardness of this design facilitates its implementation and comprehension. Nevertheless, the K-nearest neighbors (KNN) algorithm may be demanding in terms of CPU resources and is also susceptible to the size of features and irrelevant data. Therefore, it is crucial to do meticulous feature scaling and selection. Although KNN has several disadvantages, it continues to be widely used because of its versatility and efficiency in many real-world applications, including recommendation systems, pattern recognition, and anomaly detection.

## 3.5 Evaluation of model performance

There is four primary metrics. RMSE is a precise accuracy metric obtained by taking the square root of the average of the squared discrepancies between projected values and actual values. A lower root means square error (RMSE) value suggests a greater degree of precision in a model. The Mean Absolute Error (MAE) measures the difference between the predicted values produced by a model and the actual values, providing a clear indication of how much the predictions differ from the real data. R-squared (R2) is a statistical measure that quantifies the proportion of variance in the dependent variable that can be predicted by the independent variables. It ranges from 0 (no explanatory power) to 1 (perfect fit), and represents the quality of the model's fit. Finally, the Mean Bias Error (MABE) is a statistic used to measure the average bias in a model's predictions. Although it is not commonly used as the only indicator of model error, as it may not accurately represent large individual prediction mistakes, Mean Bias Error (MBE) is important for identifying and measuring the average bias in the model's outputs. A positive bias in a variable, such as wind speed, suggests that

the data from datasets is being overestimated, whereas a negative bias indicates that the data is being underestimated. Similarly, in the case of directional variables such as wind direction, a positive bias signifies a deviation in the clockwise direction, whereas a negative bias indicates a divergence in the counterclockwise direction. Assessing the performance of MBE in conjunction with other measures, such as correlation coefficients, enhances our overall comprehension of model effectiveness. Smaller error values and larger correlation coefficients, especially for variables that have a certain direction, suggest a better level of accuracy in the model.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(Y_{\mathrm{Pred},i} - Y_{\mathrm{Obs},i}\right)^2}{N}} \qquad \text{Eq 1}$$

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| Y_{\mathrm{Obs},i} - Y_{\mathrm{Pred},i} \right| g \qquad \text{Eq 2}$$

$$\mathrm{MBE} = n1 \sum_{i=1n} (P_i - O_i) \qquad \text{Eq 3}$$

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} \qquad \text{Eq 4}$$

Where $\mathrm{SSR} = \sum \left(Y_{\mathrm{Pred},i} - \overline{Y}\right)^2$, $\mathrm{SST} = \sum \left(Y_{\mathrm{Obs},i} - \overline{Y}\right)^2$, $\overline{Y}$ is the mean of y value; N = number of observed value, $Y_{\mathrm{Predi}}$ predicted value $Y_{\mathrm{Obi}}$ =observed value.

# CHAPTER IV

## Results and discussions

This part summarizes the findings to ascertain the efficacy and comprehensibility of three sophisticated machine learning models: (KNN), XGBOOST, Rforest in predicting the characteristics of dynamic modulus of asphalt mixture (E*). The model outcomes are situated on characteristics of Asphalt mixture (E*). The results are elaborated in detail through graphs and tables that lead to the perceptiveness of the predicted accuracy. The user's text is a single period. This provides a comprehensive analysis of how these models may be utilized to forecast the behavior of asphalt mixes.

### 4.1 A performance comparison of three models

**Table 1**

Comparative analysis of estimation performance for 3 models: evaluation of test and train data on asphalt mixes the user's text is "(E*)".

Header titles for each column: The table is partitioned into two primary portions, one dedicated to training data and the other to testing data. Every segment is accompanied by the following metrics:

$R^2$ is a statistical measure that quantifies the proportion of the variation in the dependent variable that can be accurately predicted by the independent variables. RMSE: A lower RMSE value suggests a greater degree of precision in a model. The MAE measures the difference between the predicted values produced by a model and the actual values, providing a clear indication of how much the predictions differ from the real data. MABE is important for identifying and measuring the average bias in the model's outputs. A positive bias in a variable, such as wind speed, suggests that the data from datasets is being overestimated, whereas a negative bias indicates that the data is being underestimated.

KNN (K-Nearest Neighbors)

Rforest (Random Forest)

XGPR (XGBoost for Regression).

The table in the image displays performance metrics for different machine learning models, both on training and testing datasets. (The table 1 as shown below).

| Models | Train | | | | Test | | | |
|---|---|---|---|---|---|---|---|---|
| | $R^2$ | RMSE | MAE | MAPE | $R^2$ | RMSE | MAE | MAPE |
| KNN | 0.857 | 587472.33 | 3.48E+05 | 3.08E-01 | 0.794 | 7.04E+05 | 4.31E+05 | 3.79E-01 |
| Rforest | 0.994 | 1.23E+05 | 7.45E+04 | 9.20E-02 | 0.954 | 3.32E+05 | 2.03E+05 | 2.57E-01 |
| XGP | 0.998 | 6.55E+04 | 4.50E+04 | 1.28E-01 | 0.98 | 2.20E+05 | 1.37E+05 | 2.33E-01 |

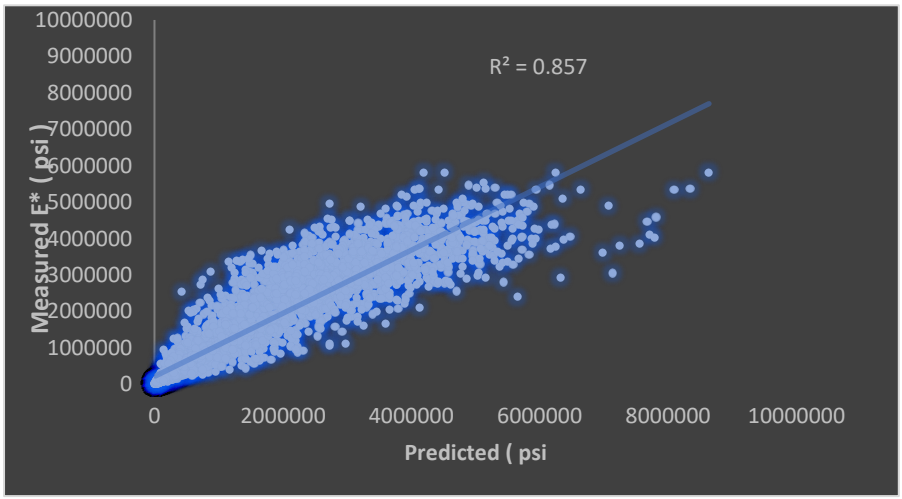
### 4.1.1 K-Nearest Neighbors (KNN)

Comparing Measured and Predicted Values of E*(psi) with KNN algorithm on both the training (a) and testing (b) datasets.

The graph has shown below in figure 2 is scatter plot that compares the predicted values against the measured (actual) values of a certain property, denoted as E* in psi (pounds per square inch). The data seems to have been generated from a model, specifically a KNN (A- Train) Axes: X-axis (Predicted psi): this represents the predicted Values produced by the KNN model. (Y-axis (Measured E* psi): this represents the actual measured values of E*, which might a material's strength or another physical property in psi units. Data points represent individual observations on the graph, where the projected value (obtained from the KNN model) is plotted against the actual observed value. The proximity of the dots to the diagonal line indicates the degree of accuracy in the model's predictions compared to the actual measurements. Diagonal Line: symbolizes the optimal situation in which the projected values precisely correspond to the measured values. The points on this line represent the model's perfect predictions, while the scatter of points around the line indicates the

model's accuracy. The R square value of 0.857, displayed in the upper right of the graph, is the coefficient of determination. It measures how well the predicted values from the KNN model explain the variability in the measured data. (Figure 2 is shown below).

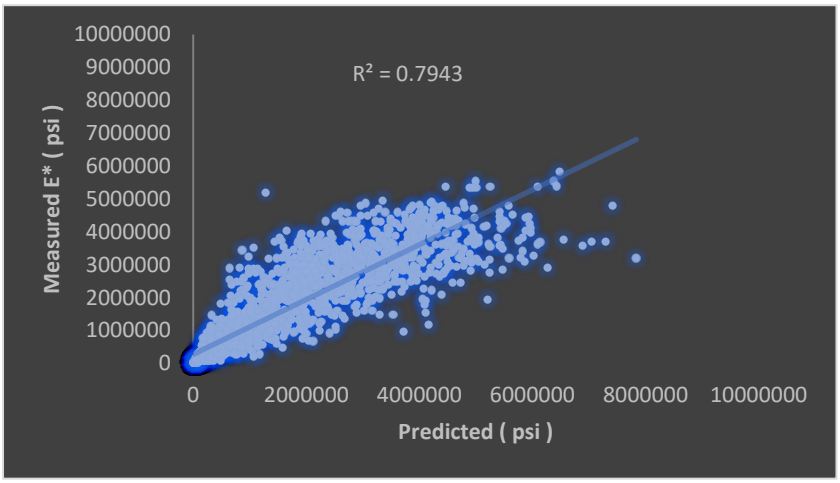
**Figure 2**
**A: train**




The data visualization that is displayed below in figure 3 is a scatter plot that represents a comparison between the anticipated values and the measured (actual) values of a certain property. This property is represented as E* in psi, which stands for pounds per square inch. The information appears to have been derived from a model, more precisely a KNN (B test in this case). The axes: The X-axis, also known as the Predicted psi, is a representation of the predicted values that are generated by the KNN model. (Y-axis (Measured E* psi): this shows the actual measured values of E*, which may be the strength of a material or another physical property measured in psi units in the physical property unit system. Data points are individual observations that are displayed on the graph. The projected value, which is produced from the KNN model, is compared against the actual value that was seen. The degree to which the model's predictions are accurate in comparison to the actual data is shown by the closeness of the dots to the diagonal line. The ideal scenario is represented by the diagonal line, which is the point at which the predicted values exactly match to the values that have

been measured. This line contains points that reflect the model's flawless predictions, and the dispersion of points that surround the line is an indication of how accurate the model is. The value of 0.7943 for the R square statistic, which can be seen in the upper right corner of the graph, represents the coefficient of determination. It determines the degree to which the projected values derived from the KNN model adequately explain the variability in the data that was observed. (The illustration of Figure 3 may be seen below).

**Figure 3**
**B: test**




### 4.1.2 Random forest

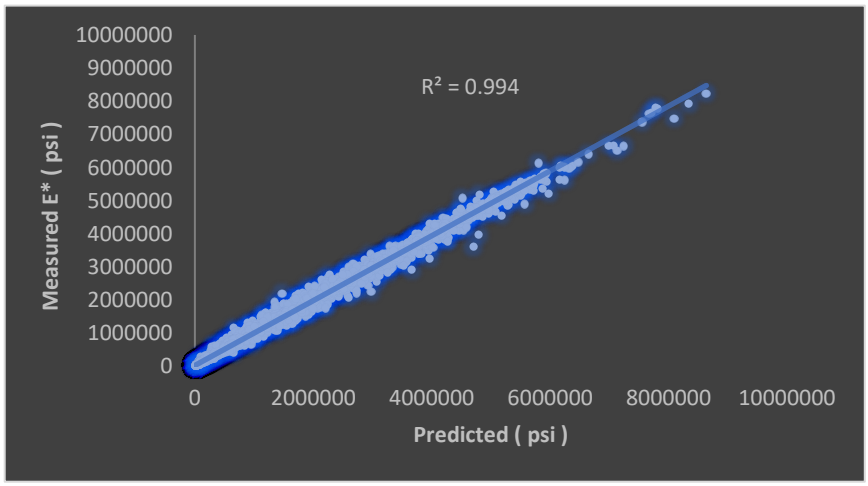
Comparison of predicted and measured E*(ψ) values using random forest Algorithm for training (a) and testing (b) datasets**.**

The visualization that is displayed below in figure 4 is a scatter plot that represents a comparison between the anticipated values and the measured (actual) values of a certain property. This property is represented as E* in psi, which stands for pounds per square inch. A model, more precisely a Random forest algorithm (also known as Random forest A- Train), appears to have been responsible for the generation of the data. One of the axes is the X-axis, which stands for the expected psi. This axis reflects the values that are predicted by the Rforest model. (Y-axis (Measured E* psi): this shows the actual measured values of E*, which may be the strength of a material or another physical property measured in psi units in the physical property unit system. On the graph, the data points represent unique observations, and the projected value,

which was produced from the Random forest model, is shown against the actual value that was seen. The degree to which the model's predictions are accurate in comparison to the actual data is shown by the closeness of the dots to the diagonal line. Diagonal Line: symbolizes the optimal situation in which the projected values precisely correspond to the measured values. The points on this line represent the model's perfect predictions, while the scatter of points around the line indicates the model's accuracy. The R square value of 0.994, displayed in the upper right of the graph, is the coefficient of determination. It measures how well the predicted values from the Rforest model explain the variability in the measured data. (figure 4 is shown below)

**Figure 4**

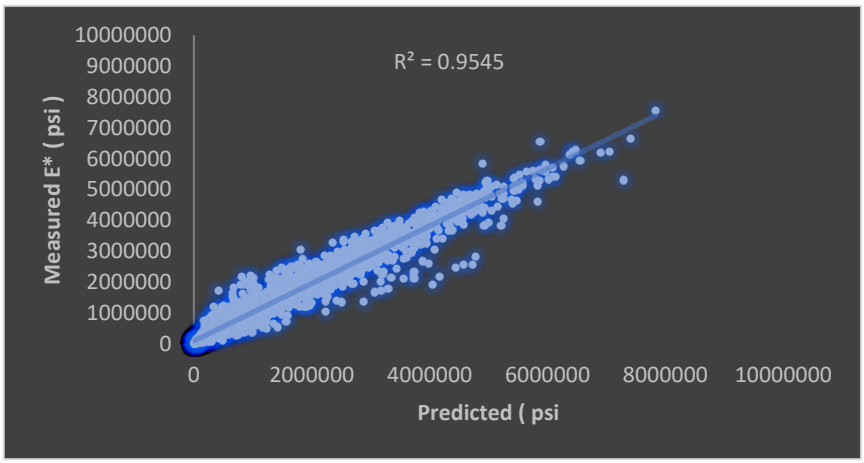
**(A) train**




The graph has shown below in figure 5 is scatter plot that compares the predicted values against the measured (actual) values of a certain property, denoted as E* in psi (pounds per square inch). The data seems to have been generated from a model, specifically a Random forest) algorithm (referred to as Random forest B- test) Axes: X-axis (Predicted psi): this represents the predicted Values produced by the Rforest model. (Y-axis (Measured E* psi): this represents the actual measured values of E*, which might a material’s strength or another physical property in psi units. Data points represent individual observations on the graph, where the projected value (obtained from the Random forest model) is plotted against the actual observed value. The proximity of the dots to the diagonal line indicates the degree of accuracy in the

model's predictions compared to the actual measurements. Diagonal Line: symbolizes the optimal situation in which the projected values precisely correspond to the measured values. The points on this line represent the model's perfect predictions, while the scatter of points around the line indicates the model's accuracy. The R square value of 0.954, displayed in the upper right of the graph, is the coefficient of determination. It measures how well the predicted values from the Rforest model explain the variability in the measured data. (Figure 5 is shown below)

**Figure 5**
**B: test**



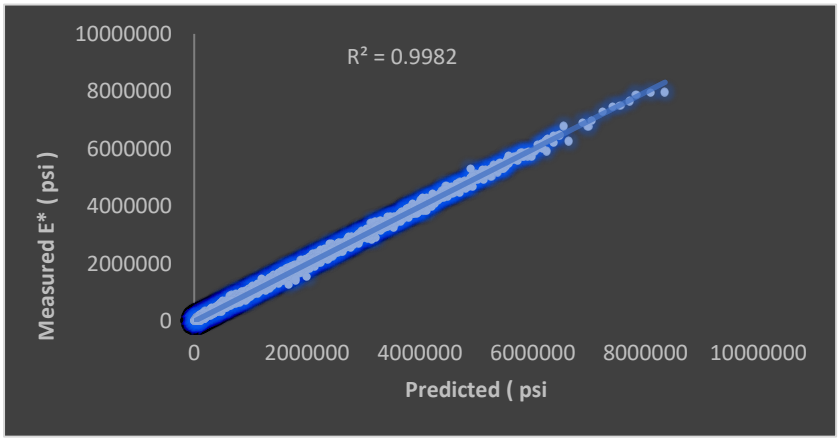

### 4.1.3XGB

Comparing Measured vs. Predicted E* (psi) for the XGB regression for (a) train and (b) test datasets.

The graph has shown below in figure 6 is scatter plot that compares the predicted values against the measured (actual) values of a certain property, denoted as E* in psi (pounds per square inch). The data seems to have been generated from a model, specifically a XGB) algorithm (referred to as XGB A- Train) Axes: X-axis (Predicted psi): this represents the predicted Values produced by the XGB model. (Y-axis (Measured E* psi): this represents the actual measured values of E*, which might a material’s strength or another physical property in psi units. Data points represent

individual observations on the graph, where the projected value (obtained from the XGB model) is plotted against the actual observed value. The proximity of the dots to the diagonal line indicates the degree of accuracy in the model's predictions compared to the actual measurements. Diagonal Line: symbolizes the optimal situation in which the projected values precisely correspond to the measured values. The points on this line represent the model's perfect predictions, while the scatter of points around the line indicates the model's accuracy. The R square value of 0.999, displayed in the upper right of the graph, is the coefficient of determination. It measures how well the predicted values from the XGB model explain the variability in the measured data. (Figure 6 is shown below).

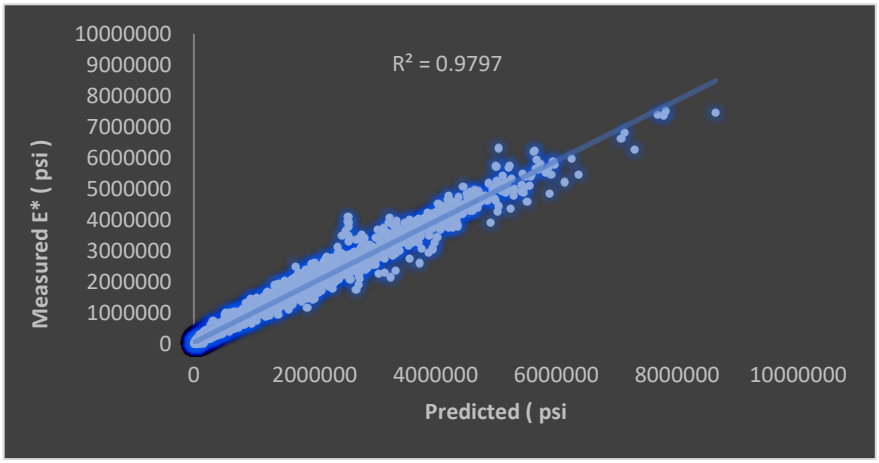
**Figure 6**
**A test**




The graph has shown below in figure 7 is scatter plot that compares the predicted values against the measured (actual) values of a certain property, denoted as E* in psi (pounds per square inch). The data seems to have been generated from a model, specifically a XGB) algorithm (referred to as XGB B-test) Axes: X-axis (Predicted psi): this represents the predicted Values produced by the XGB model. (Y-axis (Measured E* psi): this represents the actual measured values of E*, which might a material's strength or another physical property in psi units. Data points represent individual observations on the graph, where the projected value (obtained from the XGB model) is plotted against the actual observed value. The proximity of the dots to the diagonal line indicates the degree of accuracy in the model's predictions compared to the actual measurements. Diagonal Line: symbolizes the optimal situation in which the projected values precisely correspond to the measured values. The points on this

line represent the model's perfect predictions, while the scatter of points around the line indicates the model's accuracy. The R square value of 0.979, displayed in the upper right of the graph, is the coefficient of determination. It measures how well the predicted values from the XGB model explain the variability in the measured data (Figure 7 is shown below).

**Figure 7**
**(b) test**




In order to comprehensively provide the discussion of the results in the table, it is crucial to provide a concise explanation of each statistic and to provide an overall assessment of the characteristics of the models under consideration. There are three models in this case: K Nearest Neighbors, Random Forest, and Extreme Gradient Boosting. The Coefficient of Determination, Root Mean Square Error, Mean Absolute Error, and Mean Absolute Percentage Error would be used to grade these models. Each of these measures indicates the predicted accuracy of each model on both the training and testing datasets. The table focuses on determining the dynamic modulus of asphalt mixes, denoted as E*. This parameter is crucial in defining the material qualities and performance of asphalt under various situations.

The R Square value is the Coefficient of Determination; thus, this indicates how well-predictive the model is for the actual data. An R Square value of 1 would mean perfect prediction by the model, while on the other hand, an R Square of 0 would say that the model did no better than if it had used a mean of the target variable. In table never

The R square values for the KNN, Reforest, and XGP are 0.857, 0.994, and 0.998 on the training dataset, respectively. This shows that XGP has precisely captured the variance of the training data, closely followed by RForest, while KNN is way off. However, when testing these same using test data, the R square values become 0.794 for KNN, 0.954 for RForest, and 0.98 for XGP. This clearly shows that XGP displays superior performance on unseen data, thus excellent generalization, while RForest has good generalization but not ideally. Much drop in R square value was observed for KNN, suggesting possible overfitting and poorer generalization compared to the other models.

The RMSE metric measures the average size of prediction errors, providing insight into a model's accuracy. Thus, lower values of RMSE mean that the model is better. Again, the table has KNN with quite a high RMSE on training data at 587472.33 and improves in test data to 704000. This rise is vast, meaning KNN has problems generalizing since its predictive error is significant in unseen data. In contrast, RForest presents a considerably smaller RMSE of 123,000 for the training data and 332,000 for test data for better accuracy and generalization. XGP did even better with the RMSE values since with the lowest value, it gives 65500 on training data and 220000 on the test data, showing a much better accuracy and more robust generalization.

Another measure of prediction error is the MAE, which averages the absolute differences between predicted and actual values. The MAE for KNN is already very high, about 348000 on the training set, and then further rises to 431000 for the test set alone, resoundingly establishing it as a poor generalizer. RForest has a MAE of 74500 on the training data and 203000 on the test data. Again, XGP shows the best performance with respect to MAE, getting values of 45000 for training and 137000 for test, that are very close to each other and hence correspond to quite accurate predictions and good generalization.

MAPE is valuable because it can be used to compare models at different sizes, as it measures the percentage difference between predictions. A lower Mean Absolute Percentage Error (MAPE) indicates a superior model performance. The table above indicates that the Mean Absolute Percentage Error (MAPE) is high for K-Nearest Neighbors (KNN) on both the training data (0.308) and the test data (0.379), suggesting that the predictions made by KNN are erroneous. Regarding this matter, the RForest model exhibited superior performance, achieving a Mean Absolute

Percentage Error (MAPE) of 0.092 on the training data and 0.257 on the test data. In addition, the XGP model demonstrates superior accuracy compared to both models, with MAPE values of 0.128 on the training data and 0.233 on the test data, expressed as percentages. KNN is a simple, instance-based learning algorithm that classifies new cases based on their similarity to stored instances. From the performance metrics of KNN, one can infer that although it fits the training data rather delicate, it does not generalize for the test part, as can be seen from the enormous drops in R square and rises in RMSE, MAE, and MAPE. Indicating overfitting means that the model captures noise in the training data rather than underlying patterns, hence poor performance in new data. While KNN is simple and easy to use, its effectiveness decreases with broader and more diversified datasets. It may be computationally expensive and memory-intensive because all its training instances are stored and distances are computed for each prediction.

Random Forest is a combination learning method that constructs many decision trees during the learning stage and takes output as the average prediction value over all those trees. The performance metrics for RForest suggest that the fitting was excellent for the training data and quite good in generalization for the test data. Almost close to 1, the R square value of its RMSE, MAE, and MAPE values are all significantly less compared to KNN towards the training and test datasets. It shows that RForest captures complex patterns in the data well without overfitting the data or giving more trust in prediction tasks. It allows for big data, and high-dimensional data to be handled and is robust to overfitting—making RForest one of the most versatile and widely used models in many domains.

Extreme Gradient Boosting combines the strengths of Gradient Boosting and Gaussian process regression into a robust and scalable framework for predictive modeling. The performance metrics of the XGP model are better than the models with maximal R-squares and minimal RMSE, MAE, and MAPE. That is, it seems that XGP manages to capture the complex patterns in the data with excellent reliability and, at the same time, generalizes quite well to new data, making that model the one with the best performance in this comparison. Thus, XGP works, reducing boosting procedure errors, and yet integrates in a probabilistic way the predictions from Gaussian processes, providing accuracy with quantification of uncertainty.

In general, the analysis of Knn, RForest, and XGP models shows the strengths and weaknesses of the subsequent models. KNN, being simple and interpretable, thrives on good generalization and can be computationally quite expensive. RForest is a balanced technique that is both highly accurate and robust, fitting most problems. Still, XGP shows superiority in performance and generalization, so it fits the problems where tasks demand high accuracy and high reliability. All in all, the choice of model depends on the specific application, type of data, and available computing resources. XGP is preferred for high precision purposes, whereas RForest is our preference for those applications that require good generalization capacity. On more straightforward problems, we can use KNN or take it into account as a baseline model, given its simplicity in building and interpretation.

Practically, the modeling application to estimate the dynamic modulus of asphalt mixtures ($E*$) has very significant weight. An accurate estimate of $E*$ sheds light on material properties and their performance under various conditions, which often directly impacts pavement design, maintenance, and longevity. The superior performance of XGP in this context suggests that the model can provide trustworthy predictions of E, which can significantly assist in better decision-making for infrastructure projects. The next best alternative is RForest, which is again robust if computational efficiency and interpretability become essential considerations. The next one is the KNN model, with overall lower accuracy values, and that would also be helpful when limited computational resources should be considered.

Hence, all of the above lead to the conclusion that the comprehensive analysis of the KNN, RForest, and XGP models considered in evaluation metrics such as R Square, RMSE, MAE, and MAPE gives extensive information on their predictability power and practical prospects. XGP comes out as the most efficient one with good accuracy and high robusticity and generalization ability that makes it a fit for tasks experiencing refined and accurate or confident predictions. RForest also performs well and finds a strong balance between accuracy and interpretability. While KNN is simple and interpretable, it lacks generalization and can be computationally expensive. It depends on the model choice for a particular application, the nature of the data, and computational resources; XGP has put ahead not for high-precision tasks over RForest and generalizability, while KNN is simple and easy to handle.

# CHAPTER V

## Conclusion and recommendation

## 5.1 conclusion

In conclusion, three machine learning models are analyzed: K-Nearest Neighbors, Random Forest, and Extreme Gradient Boosting. Each of these approaches has specific strengths and weaknesses. Although KNN is very simple and easy to interpret, it suffers from a high computational cost. It does not generalize very well, which makes this algorithm less effective in conditions of more extensive and more diversified datasets. On the other hand, random forest is one of the robust and balanced techniques. It performs well for big data for cases with many dimensions; it is resistant to overfitting, which gives good generalization for test data in complex patterns. This makes it versatile for many applications. It turns out that XGBoost is incomparable to the rest of the models in terms of accuracy and reliability. Only then does it combine gradient boosting and Gaussian process regression to achieve very high-performance metrics, such as R-squared, RMSE, MAE, and MAPE. It is especially good at capturing complex patterns in data and offers excellent generalization to new data, so it would be very appropriate for tasks that require high precision and reliability.

The excellent performance of XGBoost in practical applications, especially in estimating asphalt mixtures' dynamic modulus (E*), is critical. E* prediction is paramount to understanding material properties and, hence, their performances under a wide range of conditions, impacting directly on pavement design, maintenance, and endurance. Whereas Random Forest is also robust and preferred when computational efficiency and interpretability are essential, KNN may be more suitable for more straightforward problems or when computational resources are low owing to its lower accuracy. Therefore, the choice of a model depends on the specific application, type of data, and computing resources available. The preference is based on tasks requiring high precision with XGBoost, good generalization with Random Forest, and sometimes KNN classifications owing to their simplicity during the build and interpretation. Based on the evaluation of the machine learning models—K-Nearest Neighbors, Random Forest, and Extreme Gradient Boosting—or E* dynamic modulus prediction of asphalt mixtures, a few recommendations come up for future research studies and practical applications.

## 5.2 Recommendation

It is recommended to prioritize the use of XGBoost for high-precision tasks due to its superior performance in terms of accuracy and reliability. Compared to the other two models, this has returned the highest R-squared values and the lowest errors, hence very effective in complex patterns and generalizing well to new data. This makes XGBoost especially useful for projects involving critical infrastructure where exact predictions of material behavior under varying conditions are necessary. Researchers and practitioners should integrate this XGBoost into their predictive modeling workflow to improve the reliability and accuracy of results. While high in precision, XGBoost could not beat Random Forest with regard to robust performance, especially when handling high-dimensional data and reducing overfitting. This makes it balanced for applications where both accuracy and interpretability are important because of its ability to fit training data well and generalize to test data. In these cases, practitioners are encouraged to use random forests in scenarios where computational efficiency and model interpretability are essential. Such a model can turn out to be excellent in applications where a good tradeoff between accuracy and the capability to explain model decisions is required—for instance, in preliminary feasibility studies or in an environment where stakeholder understanding of the project is very crucial. Although less accurate than the other two models, the K-Nearest Neighbors algorithm has some advantages due to its simplicity and ease of implementation. It can be used as a baseline model to which more sophisticated algorithms should be compared, for example, KNN. This will make it quite effective in smaller datasets or when the interpretability of the model is essential. It would often be a good solution for simple problems or at least be used as a baseline model in the development process. KNN will lead researchers to consider it in cases where computational resources are limited, and a fast, interpretable solution is required. Evaluation and continuous optimization of machine learning models are crucial to improving their performance over time.

This can, to a considerable extent, be improved by regularly updating the training data and tuning the hyperparameters, together with the addition of new features. A powerful framework of validation should be implemented to test model performance on unseen data to prevent overfitting and ensure that the models generalize well. Researchers should consider the iterative process of monitoring and adjusting models using

individual performance metrics to guarantee their relevance and accuracy in the long run. One such application where real-time monitoring can drive home the benefits of an integrated machine learning model with sensor networks is SHM. In this approach, continuous data will be collected and updated for modeling calibration to predict the structural performance/possible failures more accurately and timely. Optimization algorithms should be applied at least in designing efficient sensor networks and developing effective damage detection methods. A practitioner will want to consider integrating real-time data collection with predictive modeling to advance their systems' accuracy and responsiveness. Lastly, the interpretability and usability of machine learning models are critical factors for their practical applications. While model accuracy comes first, it alone is not good enough when it is at least equally important to ensure that models have clear and understandable insights that drive predictions. Feature importance analysis, partial dependence plots, decision trees, and many other tools and techniques may help improve complex models such as Random Forest and XGBoost regarding their interpretability. Developing models that are user-friendly and accessible to non-experts is very critical to ensure their adoption and effective use in different fields. Research should aim at developing such models; however, besides being well-performing, they should also be meaningful and relevant—able to provide actionable insights and be as easy as possible to use by practitioners.

# References

Ruggieri, S., Cardellicchio, A., Nettis, A., Renò, V., & Uva, G. (2022). Using machine learning approaches to perform defect detection of existing bridges. Procedia Structural Integrity, 44(2022), 2028–2035. https://doi.org/10.1016/j.prostr.2023.01.259

Samadi, H., & Samadi, L. (2022). Soil Classification Modelling Using Machine Learning Methods. Research Gate, March. https://www.researchgate.net/publication/359056095

Schwartz, Y., & Raslan, R. (2013). Variations in results of building energy simulation tools, and their impact on BREEAM and LEED ratings: A case study. Energy and Buildings, 62, 350–359. https://doi.org/10.1016/j.enbuild.2013.03.022

Shanmugasundaram, N., Praveenkumar, S., Gayathiri, K., & Divya, S. (2022). Prediction on compressive strength of Engineered Cementitious composites using Machine learning approach. Construction and Building Materials, 342(PA), 127933. https://doi.org/10.1016/j.conbuildmat.2022.127933

Silva, P. B., Andrade, M., & Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review. Journal of Traffic and Transportation Engineering (English Edition), 7(6), 775–790. https://doi.org/10.1016/j.jtte.2020.07.004

Singh, P., Ahmad, U. F., & Yadav, S. (2020). Structural Health Monitoring and Damage Detection through Machine Learning approaches. E3S Web of Conferences, 220. https://doi.org/10.1051/e3sconf/202022001096

Sparks, T. D., Kauwe, S. K., Parry, M. E., Tehrani, A. M., & Brgoch, J. (2020). Machine Learning for Structural Materials. Annual Review of Materials Research, 50, 27–48. https://doi.org/10.1146/annurev-matsci-110519-094700

Sun, H., Burton, H. V., & Huang, H. (2021). Machine learning applications for building structural design and performance assessment: State-of-the-art review. Journal of Building Engineering, 33(July 2020), 101816. https://doi.org/10.1016/j.jobe.2020.101816

Tayefeh Hashemi, S., Ebadati, O. M., & Kaur, H. (2020). Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. SN Applied Sciences, 2(10), 1–27. https://doi.org/10.1007/s42452-

020-03497-1

Truong, V. H., Pham, H. A., Huynh Van, T., & Tangaramvong, S. (2022). Evaluation of machine learning models for load-carrying capacity assessment of semi-rigid steel structures. Engineering Structures, 273, 115001. https://doi.org/10.1016/j.engstruct.2022.115001

Wu, K., Mengiste, E., & García de Soto, B. (2023). a Machine Learning Framework for Construction Planning and Scheduling. June, 391–399. https://doi.org/10.3311/ccc2023-052

Yang, H., Yan, W., & He, H. (2016). Parameters identification of moving load using ANN and dynamic strain. Shock and Vibration, 2016. https://doi.org/10.1155/2016/8249851

Yasir, R. M., Nower, D. N., & Shoyaib, D. M. (2022). Traffic Congestion Prediction Using Machine Learning Techniques. http://arxiv.org/abs/2206.10983

Yuan, F.-G., Zargar, S. A., Chen, Q., & Wang, S. (2020). Machine learning for structural health monitoring: challenges and opportunities. 1137903(April 2020), 2. https://doi.org/10.1117/12.2561610

Zaparoli Cunha, B., Droz, C., Zine, A. M., Foulard, S., & Ichchou, M. (2023). A review of machine learning methods applied to structural dynamics and vibroacoustic. Mechanical Systems and Signal Processing, 200. https://doi.org/10.1016/j.ymssp.2023.110535

Zhang, J., Hu, J., Li, X., & Li, J. (2020). Bayesian network-based machine learning for design of pile foundations. Automation in Construction, 118(March). https://doi.org/10.1016/j.autcon.2020.103295

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. Eco-Environment and Health, 1(2), 107–116. https://doi.org/10.1016/j.eehl.2022.06.001

Hegazy, T., Ayed, A.: Neural Network Model for Parametric Cost Estimation of Highway Projects, Journal of Construction Engineering and Management, 124 (1998) 3, pp. 210-218.

Use the "Insert Citation" button to add citations to this document

Tian, M. (2022, June). Size Detection System of Building Rebar Based on LightGBM Algorithm. In International Conference on Applications and

Techniques in Cyber Intelligence (pp. 176-183). Cham: Springer International Publishing.

Gharineiat, Z., Tarsha Kurdi, F., & Campbell, G. (2022). Review of Automatic Processing of Topography and Surface Feature Identification LiDAR Data Using Machine Learning Techniques. Remote Sensing, 14(19). https://doi.org/10.3390/rs14194685

Guo, K., Yang, Z., Yu, C. H., & Buehler, M. J. (2021). Artificial intelligence and machine learning in design of mechanical materials. Materials Horizons, 8(4), 1153–1172. https://doi.org/10.1039/d0mh01451f

Huang, L. F. (2010). Artificial intelligence. In 2010 The 2nd International Conference on Computer and Automation Engineering, ICCAE 2010 (Vol. 4). https://doi.org/10.1109/ICCAE.2010.5451578

Jafari, A., Sotoudeh Bahreini, R., & Yazdanian, R. (2015). Empirical, analytical and numerical approaches to failure analysis of a frictional power transmission composite roller. Engineering Failure Analysis, 52, 61–74. https://doi.org/10.1016/j.engfailanal.2015.02.013

Jhaveri, R. H., Revathi, A., Ramana, K., Raut, R., & Dhanaraj, R. K. (2022). A Review on Machine Learning Strategies for Real-World Engineering Applications. Mobile Information Systems, 2022. https://doi.org/10.1155/2022/1833507

Kormushev, P., Calinon, S., & Caldwell, D. G. (2013). Reinforcement learning in robotics: Applications and real-world challenges. Robotics, 2(3), 122–148. https://doi.org/10.3390/robotics2030122

Krishnakumari, P., Perotti, A., Pinto, V., Cats, O., & Van Lint, H. (2018). Understanding Network Traffic States using Transfer Learning. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018-Novem, 1396–1401. https://doi.org/10.1109/ITSC.2018.8569450

Krzywanski, J. (2022). Advanced AI Applications in Energy and Environmental Engineering Systems. Energies, 15(15), 15–17. https://doi.org/10.3390/en15155621

Le Clainche, S., Ferrer, E., Gibson, S., Cross, E., Parente, A., & Vinuesa, R. (2023). Improving aircraft performance using machine learning: A review. Aerospace Science and Technology, 138, 108354. https://doi.org/10.1016/j.ast.2023.108354

Lee, J., & Lee, S. (2023). Construction Site Safety Management: A Computer Vision and Deep Learning Approach. Sensors, 23(2). https://doi.org/10.3390/s23020944

Liu, C., Wu, D., Li, Y., & Du, Y. (2021). Large-scale pavement roughness measurements with vehicle crowdsourced data using semi-supervised learning. Transportation Research Part C: Emerging Technologies, 125(February), 103048. https://doi.org/10.1016/j.trc.2021.103048

Martinez-Martin, E., Cazorla, M., & Orts-Escolano, S. (2020). Machine learning techniques for assistive robotics. Electronics (Switzerland), 9(5), 9–11. https://doi.org/10.3390/electronics9050821

Mohtasham Moein, M., Saradar, A., Rahmati, K., Ghasemzadeh Mousavinejad, S. H., Bristow, J., Aramali, V., & Karakouzian, M. (2023). Predictive models for concrete properties using machine learning and deep learning approaches: A review. Journal of Building Engineering, 63(PA), 105444. https://doi.org/10.1016/j.jobe.2022.105444

Mosavi, A., & Varkonyi-Koczy, A. R. (2017). Integration of machine learning and optimization for robot learning. Advances in Intelligent Systems and Computing, 519, 349–355. https://doi.org/10.1007/978-3-319-46490-9_47

Mostoufi, N., & Constantinides, A. (2022). Applied Numerical Methods for Chemical Engineers. Applied Numerical Methods for Chemical Engineers, 1–492. https://doi.org/10.1016/C2019-0-04646-8

Oladipupo, T. (2010). Types of Machine Learning Algorithms. New Advances in Machine Learning. https://doi.org/10.5772/9385

Pan, J., Low, K. L., Ghosh, J., Jayavelu, S., Ferdaus, M. M., Lim, S. Y., Zamburg, E., Li, Y., Tang, B., Wang, X., Leong, J. F., Ramasamy, S., Buonassisi, T., Tham, C. K., & Thean, A. V. Y. (2021). Transfer Learning-Based Artificial Intelligence-Integrated Physical Modeling to Enable Failure Analysis for 3 Nanometer and Smaller Silicon-Based CMOS Transistors. ACS Applied Nano Materials, 4(7), 6903–6915. https://doi.org/10.1021/acsanm.1c00960

Prajwal, K., Amitasree, P., Vamshi, G. R., & Devi, V. S. K. (2021). Electricity Consumption Forecasting Using Machine Learning. Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSEC 2021, 1–8. https://doi.org/10.1109/ICOSEC51865.2021.9591889

Stergiou, K., Ntakolia, C., Varytis, P., Koumoulos, E., Karlsson, P., & Moustakidis, S. (2023). Enhancing property prediction and process optimization in building materials through machine learning: A review. Computational Materials Science, 220(December 2022), 112031. https://doi.org/10.1016/j.commatsci.2023.112031

Tan, J., Yuan, Q., Guo, W., Xie, N., Liu, F., Wei, J., & Zhang, X. (2022). Deep Reinforcement Learning for Traffic Signal Control Model and Adaptation Study. Sensors, 22(22), 1–21. https://doi.org/10.3390/s22228732

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. L. A., Elkhatib, Y., Hussain, A., & Al-Fuqaha, A. (2019). Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges. IEEE Access, 7(September 2017), 65579–65615. https://doi.org/10.1109/ACCESS.2019.2916648

van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. Machine Learning, 109(2), 373–440. https://doi.org/10.1007/s10994-019-05855-6

Wang, C. C. (2022). Special Issue on Recent Advances in Machine Learning and Applications. Processes, 10(11), 1–4. https://doi.org/10.3390/pr10112411

Weiss, K., Khoshgoftaar, T. M., & Wang, D. D. (2016). A survey of transfer learning. In Journal of Big Data (Vol. 3, Issue 1). Springer International Publishing. https://doi.org/10.1186/s40537-016-0043-6

Xu, Y., Zhou, Y., Sekula, P., & Ding, L. (2021). Machine learning in construction: From shallow to deep learning. Developments in the Built Environment, 6(April 2020), 100045. https://doi.org/10.1016/j.dibe.2021.100045

Yuan, Y., & Jia, K. Bin. (2016). A semi-supervised approach for water quality detection based on IoT network. Journal of Information Hiding and Multimedia Signal Processing, 7(4), 858–866.

Zabin, A., González, V. A., Zou, Y., & Amor, R. (2022). Applications of machine learning to BIM: A systematic literature review. Advanced Engineering Informatics, 51(April 2021). https://doi.org/10.1016/j.aei.2021.101474

Zha, Z. J., Mei, T., Wang, J., Wang, Z., & Hua, X. S. (2009). Graph-based semi-supervised learning with multiple labels. Journal of Visual Communication and Image Representation, 20(2), 97–103. https://doi.org/10.1016/j.jvcir.2008.11.009

Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., Ma, X., Marrone, B. L., Ren, Z. J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B. M., Xiao, X., Yu, X., Zhu, J. J., & Zhang, H. (2021). Machine Learning: New Ideas and Tools in Environmental Science and Engineering. Environmental Science and Technology, 55(19), 12741–12754. https://doi.org/10.1021/acs.est.1c01339

Alkaissy, M., Arashpour, M., Golafshani, E. M., Hosseini, M. R., Khanmohammadi, S., Bai, Y., & Feng, H. (2023). Enhancing construction safety: Machine learning-based classification of injury types. Safety Science, 162(March 2022), 106102. https://doi.org/10.1016/j.ssci.2023.106102

Aydın, Y., Işıkdağ, Ü., Bekdaş, G., Nigdeli, S. M., & Geem, Z. W. (2023). Use of Machine Learning Techniques in Soil Classification. Sustainability (Switzerland), 15(3). https://doi.org/10.3390/su15032374

Barkhordari, M. S., Armaghani, D. J., & Asteris, P. G. (2023). Structural Damage Identification Using Ensemble Deep Convolutional Neural Network Models. CMES - Computer Modeling in Engineering and Sciences, 134(2), 835–855. https://doi.org/10.32604/cmes.2022.020840

Bekdaş, G., Yücel, M., & Nigdeli, S. M. (2021). Estimation of optimum design of structural systems via machine learning. Frontiers of Structural and Civil Engineering, 15(6), 1441–1452. https://doi.org/10.1007/s11709-021-0774-0

Ben Chaabene, W., Flah, M., & Nehdi, M. L. (2020). Machine learning prediction of mechanical properties of concrete: Critical review. Construction and Building Materials, 260, 119889. https://doi.org/10.1016/j.conbuildmat.2020.119889

Bharti, Redhu, P., & Kumar, K. (2023). Short-term traffic flow prediction based on optimized deep learning neural network: PSO-Bi-LSTM. In Physica A: Statistical Mechanics and its Applications (Vol. 625). https://doi.org/10.1016/j.physa.2023.129001

Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Science of the Total Environment, 721, 137612. https://doi.org/10.1016/j.scitotenv.2020.137612

Choi, H. W., Kim, H. J., Kim, S. K., & Na, W. S. (2023). An Overview of Drone Applications in the Construction Industry. Drones, 7(8). https://doi.org/10.3390/drones7080515

D'Amico, B., Myers, R. J., Sykes, J., Voss, E., Cousins-Jenvey, B., Fawcett, W., Richardson, S., Kermani, A., & Pomponi, F. (2019). Machine Learning for Sustainable Structures: A Call for Data. Structures, 19, 1–4. https://doi.org/10.1016/j.istruc.2018.11.013

Dang, H. V., Raza, M., Nguyen, T. V., Bui-Tien, T., & Nguyen, H. X. (2021). Deep learning-based detection of structural damage using time-series data. Structure and Infrastructure Engineering, 17(11), 1474–1493. https://doi.org/10.1080/15732479.2020.1815225

Fang, X., Luo, H., & Tang, J. (2005). Structural damage detection using neural network with learning rate improvement. Computers and Structures, 83(25–26), 2150–2161. https://doi.org/10.1016/j.compstruc.2005.02.029

Fitzsimmonsa, J., Hong, Y., & Brilakis, I. (2020). Improving construction project schedules before execution. Proceedings of the 37th International Symposium on Automation and Robotics in Construction, ISARC 2020: From Demonstration to Practical Use - To New Stage of Construction Robot, October, 1144–1151. https://doi.org/10.22260/isarc2020/0157

Ghobadi, F., & Kang, D. (2023). Application of Machine Learning in Water Resources Management: A Systematic Literature Review. Water (Switzerland), 15(4). https://doi.org/10.3390/w15040620

Ghoroghi, A., Rezgui, Y., Petri, I., & Beach, T. (2022). Advances in application of machine learning to life cycle assessment: a literature review. International Journal of Life Cycle Assessment, 27(3), 433–456. https://doi.org/10.1007/s11367-022-02030-3

Gomez-Cabrera, A., & Escamilla-Ambrosio, P. J. (2022). Review of Machine-Learning Techniques Applied to Structural Health Monitoring Systems for Building and Bridge Structures. Applied Sciences (Switzerland), 12(21). https://doi.org/10.3390/app122110754

Gulghane, A., Sharma, R. L., & Borkar, P. (2023). Quantification analysis and prediction model for residential building construction waste using machine learning technique. Asian Journal of Civil Engineering, 24(6), 1459–1473. https://doi.org/10.1007/s42107-023-00580-x

Hassani, S., & Dackermann, U. (2023). A Systematic Review of Optimization Algorithms for Structural Health Monitoring and Optimal Sensor Placement.

Sensors, 23(6). https://doi.org/10.3390/s23063293

Khan, Y., & See, C. S. (2016). Predicting and analyzing water quality using Machine Learning: A comprehensive model. 2016 IEEE Long Island Systems, Applications and Technology Conference, LISAT 2016, 1–6. https://doi.org/10.1109/LISAT.2016.7494106

Korup, O., & Stolle, A. (2014). Landslide prediction from machine learning. Geology Today, 30(1), 26–33. https://doi.org/10.1111/gto.12034

Kovacevic, M., Ivaniševic, N., Petronijevic, P., & Despotovic, V. (2021). Construction cost estimation of reinforced and prestressed concrete bridges using machine learning. Gradjevinar, 73(1), 1–13. https://doi.org/10.14256/JCE.2738.2019

Kunverji, K., Shah, K., & Shah, N. (2021). A Flood Prediction System Developed Using Various Machine Learning Algorithms. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3866524

Kuradusenge, M., Kumaran, S., & Zennaro, M. (2020). Rainfall-induced landslide prediction using machine learning models: The case of ngororero district, rwanda. International Journal of Environmental Research and Public Health, 17(11), 1–20. https://doi.org/10.3390/ijerph17114147

Li, Q., Wang, Z., Li, L., Hao, H., Chen, W., & Shao, Y. (2023). Machine learning prediction of structural dynamic responses using graph neural networks. Computers and Structures, 289(September), 107188. https://doi.org/10.1016/j.compstruc.2023.107188

Machine Learning-Based Modelling of Soil Properties for Geotechnical.pdf. (n.d.).

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. Water (Switzerland), 10(11), 1–40. https://doi.org/10.3390/w10111536

Najah Ahmed, A., Binti Othman, F., Abdulmohsin Afan, H., Khaleel Ibrahim, R., Ming Fai, C., Shabbir Hossain, M., Ehteram, M., & Elshafie, A. (2019). Machine learning methods for better water quality prediction. Journal of Hydrology, 578(May). https://doi.org/10.1016/j.jhydrol.2019.124084

Naseri, H., Jahanbakhsh, H., Hosseini, P., & Moghadas Nejad, F. (2020). Designing sustainable concrete mixture by developing a new machine learning technique. Journal of Cleaner Production, 258, 120578.

https://doi.org/10.1016/j.jclepro.2020.120578

Nguyen Van, T., & Nguyen Quoc, T. (2021). Research Trends on Machine Learning in Construction Management: A Scientometric Analysis. Journal of Applied Science and Technology Trends, 2(03), 96–104. https://doi.org/10.38094/jastt203105

Pham, V. H. S., & Nguyen, N. L. (2023). Estimating the volume of civil construction materials by machine learning models. Asian Journal of Civil Engineering, 24(7), 2077–2088. https://doi.org/10.1007/s42107-023-00625-1

Poh, C. Q. X., Ubeynarayana, C. U., & Goh, Y. M. (2018). Safety leading indicators for construction sites: A machine learning approach. Automation in Construction, 93(March), 375–386. https://doi.org/10.1016/j.autcon.2018.03.022

Qi, H., Yao, Y., Zhao, X., Guo, J., Zhang, Y., & Bi, C. (2022). Applying an interpretable machine learning framework to the traffic safety order analysis of expressway exits based on aggregate driving behavior data. Physica A: Statistical Mechanics and Its Applications, 597, 127277. https://doi.org/10.1016/j.physa.2022.127277

Rozos, E. (2019). Machine learning, urbanwater resources management and operating policy. Resources, 8(4). https://doi.org/10.3390/RESOURCES8040173.

Use the "Insert Citation" button to add citations to this document.

Amezquita-Sanchez, J. P., Valtierra-Rodriguez, M., & Adeli, H. (2020). Machine learning in structural engineering. Scientia Iranica, 27(6 A), 2645–2656. https://doi.org/10.24200/sci.2020.22091

Baghbani, A., Choudhury, T., Costa, S., & Reiner, J. (2022). Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. Earth-Science Reviews, 228(March), 103991. https://doi.org/10.1016/j.earscirev.2022.103991

Bianchini, M., Maggini, M., & Jain, L. C. (2013). Handbook on Neural Information Processing. Intelligent Systems Reference Library, 49, 215–216. https://doi.org/10.1007/978-3-642-36657-4

Coelho, B. Z., & Smyrniou, E. (2023). Optimisation of in-situ soil investigation with Deep Reinforcement Learning Optimisation of in-situ Reinforcement Learning soil investigation with Deep. August.

Dhandapani, C., & Sivaramakrishnan, R. (2019). Implementation of Machine

Learning (ML) in Mechanical Engineering Application using Artificial Intelligence (AI). VIII(X), 93–99.

Ding, X., & Du, W. (2022). DRLIC: Deep Reinforcement Learning for Irrigation Control. Proceedings - 21st ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN 2022, 41–53. https://doi.org/10.1109/IPSN54338.2022.00011

Directions, F. (2023). Understanding of Machine Learning with Deep Learning:

Fawzy, S. S., & Jasem, N. H. (2020). Machine Learning Algorithms, models and Applications: A Review. International Journal of Engineering Research and Advanced Technology, 06(08), 29–37. https://doi.org/10.31695/ijerat.2020.3607

Gao, H., Zhu, L. T., Luo, Z. H., Fraga, M. A., & Hsing, I. M. (2022). Machine Learning and Data Science in Chemical Engineering. Industrial and Engineering Chemistry Research, 61(24), 8357–8358. https://doi.org/10.1021/acs.iecr.2c01788.

NEAR EAST UNIVERSITY

**SCIENTIFIC RESEARCH ETHICS COMMITTEE**

20.09.2024

Dear Abdullahi Said Abdullahi

Your project **"A Predictive Modeling of Dynamic Modulus in Asphalt Mixtures: A Machine Learning Approach"** has been evaluated. Since only secondary data will be used the project does not need to go through the ethics committee. You can start your research on the condition that you will use only secondary data.

Prof. Dr. Aşkın KİRAZ

The Coordinator of the Scientific Research Ethics Committee

# Appendix B Turnitin Similarity report


Dr. Shaban Ismael Albrka