NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF CIVIL ENGINEERING

A CRITICAL ASSESSMENT OF MACHINE LEARNING APPLICATIONS IN CIVIL ENGINEERING

M.Sc. THESIS

Abdullahi Mohamud ADAM

Nicosia

January, 2024

ABDULLAHI MOHAMUD ADAM
A CRITICAL ASSESSMENT OF MACHINE LEARNING APPLICATIONS IN CIVIL ENGINEERING
NEU 2024

NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF CIVIL ENGINEERING

A CRITICAL ASSESSMENT OF MACHINE LEARNING APPLICATIONS IN CIVIL ENGINEERING

M.Sc. THESIS

Abdullahi Mohamud ADAM

Supervisor

Assist. Prof. Dr. Ikenna D.Uwanuakwa

Nicosia
January, 2024

Approval

We certify that we have read the thesis submitted by **Abdullahi Mohamud ADAM** titled "**A CRITICAL ASSESSMENT OF MACHINE LEARNING APPLICATIONS IN CIVIL ENGINEERING**" 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
Head of the Committee:	Assoc.Prof. Dr. Shaban Albrka	....
Committee Member:	Asssist. Prof. Dr. Mustafa Alas	....
Supervisor:	Assist. Prof. Dr. Ikenna Uwanuakwa	....

Approved by the Head of the Department

.01../.04../2024

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 hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of the institute of graduate studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Abdullahi Mohamud ADAM

30/01/2024

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.

Assist. Prof. Dr. Assist. Prof. Dr. Ikenna D.Uwanuakwa**,** 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 CRITICAL ASSESSMENT OF MACHINE LEARNING APPLICATIONS IN CIVIL ENGINEERING

Abdullahi Mohamud ADAM, Assist. Prof. Dr. Ikenna D.Uwanuakwa
MSc, Department of Civil Engineering, Faculty of Civil and Environmental Engineering, Near East University, Nicosia.

January, 2024, 70 Pages

This research investigates on assessing values of dynamic modulus (E*) and LOG E* in asphalt mixtures conducted through Visual Studio Code using machine learning (ML). By implementing, running codes we use important ML algorithms – Gradient Boost, XGBoost, AdaBoost, LGBM Boost and Cat Boost - for doing predictions on these important parameters. In this study, the Shapley value, permutation importance and accumulated local effects are used for impact analyses of input variables to enhance model accuracy and interpretability. Our methodology gives evidence of the application of ML in civil engineering, provides more accurate efficiency to evaluate the mechanical property of asphalt mixtures. This study does not only represent the advancement that ML has brought in this regard but also is an instructive guide for future explorations in this filed.

Keywords: asphalt mixtures, machine learning, gradient boost, parameter interpretability.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

FCN: Fully Convolutional Network

SVM: Support Vector Machine

IoU: Intersection over Union

ANN: Artificial Neural Network

FC: Fully Connected

ReLU: Rectified Linear Unit

VOC: Visual Object Classes

CNN: Convolutional Neural Network

GPU: Graphics Processing Unit

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 becoming more prevalent in recent decades. This is mostly due to their consistently great performance in modeling complex and non-linear correlations. ML algorithms are being used in a wide range of industries nowadays, including science (Au et al., 2022) and civil engineering in particular (Kim & Jung, 2024). Since the inception of modern pavement engineering, data analysis has been a part of the process. To determine important variables and assess pavement performance, various machine learning models were used, and they turned out to be effective predictive tools (Lontra, n.d., 2022).

Asphalt mixtures play a vital role in pavement construction and maintenance (Zaumanis et al., 2018), The complex modulus has a real and imaginary part that defines the elastic and viscous behaviour of the linear viscoelastic material. The absolute value of the complex modulus |E*|, is defined as the dynamic modulus (Kaloush & Witczak, 2003). Dynamic modulus (E*) is an essential parameter of asphalt mixtures. It is the parameter that distinguishes the flexible pavement performance conditioning on multiple temperatures and loading circumstances in the Mechanistic-Empirical Pavement Design Guide (MEPDG) (Xu et al., 2022). Mathematically, the dynamic modulus is defined as the maximum (peak) dynamic stress (σo) divided by the recoverable axial strain (εo) (Kaloush & Witczak, 2003).

Traditionally, empirical methods were employed for predicting E* values, but it involves complicated, and time consuming. As therefore, advanced methods are required to develop more precise E* predictive models. Currently, machine learning (ML) techniques are widely employed to handle many kinds of civil engineering challenges because of their capabilities in data processing and optimization (Xu et al., 2022). This research employs a data-driven approach to create an approach for predicting the performance of asphalt mixtures. It is intents for the critical assessment of dynamic modulas (E*) and Log E* by applying to global and local interpretability methods such as Shapley value, Permutation importance, and Accumulated local effects(ALE) . furthermore, the study specifically evaluates and addresses the performance of five prominent models-Gradient boost, XGBOOST, Adaboost, LGBM boost, and Catboost- in predicting E* and Log E*.

1.2 Problem statement of the study

The accurate evaluation of asphalt mixtures' dynamic modulus (E*) remains a critical step in pavement engineering for predicting pavement performance. Machine learning research on the modelling of E* and Log E* presented in the literature lacks complete components of model interpretability; local and global effect and importance of model features. This study seeks out the critical assessment of E* and Log E* by applying to some global and local interpretability methods which include Shapley value, Permutation importance, Accumulated local effects to identifying the most important features. Furthermore , the study specifically evaluate and address the performance of five models-Gradient boost, XGBOOST, Adaboost, LGBM boost, and Catboost- in predicting E* and Log E*.

1.3 Objectives of the study

The main objectives of this study are as follows:

- To interpret the primary features of dynamic modules E* and Log E* in asphalt mixtures, by using global and local methods.
- To address and evaluate critically the performance of five models-Gradient boost, XGBOOST, Adaboost, LGBM boost, and Catboost- in predicting E* and its logarithmic value Log E*.
- To compare the efficiency, effectiveness, and accuracy of ML methods utilized by identifying the most precise one.

1.4 Scope of study

The study employs the use of the selected machine learning parameter interpretabilities such as Shapley value, Permutation importance, accumulated local effects (ALE).

And addresses the performance of five model predictions, which are Gradient boost, XGBOOST, Adaboost, LGBM boost, and Catboost to forecast the dynamic modulus E* and Log E* in asphalt mixtures. It involves using adequate data sets that will be handy in training and testing of the machine learning models to ensure accuracy in output results and predictions. More than an overall discussion of these modern techniques, the focus here lies in the application use of these advanced techniques within the sphere of transportation engineering with more special emphasis laid on their utility in assessing and improving pavement materials. Powerful machine learning methods will be used to upgrade the understanding and assessment of asphalt mixtures for more accurate and effective design and maintenance of pavements.

1.5 Limitations

The study acknowledges certain limitations. Firstly, complicated advanced models based on machine learning might need particular consideration to use and understand in order to determine whether or not they fully serve all predictions. Secondly, some parameter interpretability models require additional consideration in terms of comprehending the result, failure to understand them may lead to ineffective outcomes. Finally, the performance of the models could be affected by external factors like environmental conditions and material variability whereby they do not get fully controlled in the study.

CHAPTER II

Fundamental Concepts and Historical Overview

2.1 Basic Concepts of Machine Learning

Artificial intelligence (AI) serves as a highly effective modeling method extensively used in numerous scientific disciplines (Krzywanski, 2022).

Machine learning (ML) concentrates on developing systems that enable an algorithm to collect and apply information. This learning process does not necessitate consciousness; rather, it involves detecting statistical patterns or other data-related elements. As a result, numerous machine learning algorithms diverge significantly from the way that people learn objects. ML algorithms can offer helpful insights into the diverse levels of complexity found in various learning configurations (Oladipupo, 2010).

A basic concept for learning is the model, which contains the learned information and is used to make predictions. As a rule, models are only designed for a single task. For example, in a healthcare setting, ML can take patient medical history and diagnostic test results as input data and predict the likelihood of a certain disease or condition as the output. Most important concept is model training, in which the model is taught through information as input. Machine learning models are normally trained once and then used for predictions. (Dhandapani & Sivaramakrishnan, 2019). Traditional machine learning can be categorized following an algorithm's capacity to enhance its prediction precision through training. The five integral methods for learning are supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning (RL), and transfer learning (Xu et al., 2021).

2.1.1 Supervised Learning

These algorithms utilize a specific collection of input variables (training data) that have been pre-labeled, along with the target data. By employing input variables, it generates a mapping function to associate inputs with the required outputs. The parameters and adjustment methods in the algorithm remain in place until the system demonstrates a satisfactory level of accuracy in regard to the training data (Fawzy & Jasem, 2020).

Algorithms for supervised learning requires a substantial quantity of labelled training data to develop model that demonstrate superior prediction performance (Bianchini et al., 2013).

Supervised learning can be divided into two main subcategories: regression algorithms, which produce continuous output, and classification algorithms, which produce discontinuous output. Regression algorithms seek to identify the optimal function that best fits the points in the training dataset. The regression methods can be classified into three main categories: linear regression, multiple linear regression, and polynomial regression. By designating each input to its corresponding class, classification algorithms may accurately select the most suitable class for the given data. The output of the predicting function used in this instance is categorical, and its value is assigned to one of the available classes (Directions, 2023).

Regression is employed to address problems involving the prediction of continuous variables, whereas support vector machines (SVMs) are employed for algorithmic classification. The random forest algorithm is utilized to address classification and regression problems. Supervised learning is utilized when data is labelled and a classifier is applied for classifications or numerical prediction (Directions, 2023).

In civil engineering and other engineering fields, supervised learning has been widely applied. In civil engineering, supervised learning finds diverse applications. For instance, in Structural Health Monitoring, it is utilized to analyze sensor data from various infrastructures such as bridges and buildings. Its primary purpose is to determine the health condition of the structure by detecting damages and predicting structural failures. Additionally, supervised learning can be used to predict the seismic collapse of frame structures. Supervised learning algorithms are trained to identify patterns in sensor data by using vibration data, lamb waves and electromechanical impedance, acoustic emission, etc, enabling early warnings of potential issues (Amezquita-Sanchez et al., 2020).

Furthermore, supervised learning is involved in predicting the properties of construction materials, such as concrete strength(Mohtasham Moein et al., 2023).

Supervised learning plays a crucial role in project management within the construction. It has been widely adopted in various areas of construction, including safety, bridge inspection, and on-site operation monitoring (Xu et al., 2021).

Moreover, supervised learning models are used in geotechnical engineering to analyze soil composition data and historical performance data, enabling predictions about soil settlement, slope stability, shallow and pile foundations and other geotechnical parameters.(Baghbani et al., 2022).

2.1.2 Unsupervised Learning

Unlike supervised learning, this approach involves training the learning algorithm with an input dataset that does not contain any labelled outputs. In contrast to supervised learning, there is no definitive or erroneous answer associated with each input item. Unsupervised learning exhibits a higher degree of randomness compared to supervised learning. The main objective of unsupervised learning is to acquire a more profound understanding of the data by identifying its fundamental structure or distribution pattern. As the algorithm self-learns, it strives to accurately represent a particular identified input pattern while incorporating it into the broader framework of input patterns. Therefore, extracted properties from each input item are categorized relating many inputs to them. In solving these issues associated with association and clustering, unsupervised learning can be utilized (Directions, 2023).

For example, unsupervised learning could also be leveraged in civil engineering applications, such as topographic mapping or terrain analysis. For example, to deal with elevation data that is derived through the use of remote sensing instruments like light detection and ranging (LiDAR). This analysis helps in developing detailed topographic maps which are essential for diverse civil engineering applications such as site planning and great help in the detection of leveling applications (Gharineiat et al., 2022).

In addition, Unsupervised learning applies to computer engineering and this varies from Anomaly Detection for Intrusion Detection to both recognizing the abnormal activities or security breaches in the computing networks and systems, and even Fault Detection for hardware elements or systems to detect an abnormality in that system for instance CPU memory failure (Usama et al., 2019).

2.1.3 Reinforcement Learning

This method utilizes a capable computer to map activities to certain decisions, which will lead to the generation of feedback or reward signals. The system automatically trains itself to recognize important beneficial acts by using rewards and punishments grounded on previous experiences (Fawzy & Jasem, 2020).

Learning agents are having predefined purposes and capable at perceiving the condition of their environment to certain measure. Consequently, individuals are able to take action to modify the status of the environment and move closer to the objectives they have been assigned. Reinforcement learning and supervised learning are distinct from each other based on their respective learning methods. The supervised learning approach utilizes case studies provided by an external supervisor for instructional purposes. Conversely, reinforcement learning acquires information by directly engaging with the problem environment (Directions, 2023).

Reinforcement learning is widely employed in civil engineering. Reinforcement learning algorithms can control traffic signals and manage traffic flow based on real-time data, utilizing traffic cameras, loop detectors, and vehicle counters for data collection(Tan et al., 2022). Also This approach is pivotal in autonomous vehicles. It enables them to recognize streets, make turns, and make informed decisions about their path(Directions, 2023). Additionally, in geotechnical engineering, reinforcement learning can assist in optimizing drilling and excavation processes based on data from geotechnical instruments such as soil samples and geophysical instruments (Coelho & Smyrniou, 2023). Moreover, in irrigation control, RL has been applied in many applications, such as network planning, cellular data analytics, sensor energy management, mobile app prediction, and building energy optimization that can lead to several benefits like improved crop yield and quality, reduced energy costs, and minimized environmental impact (Ding & Du, 2022).

Reinforcement learning is generally used in other engineering fields, such as computer engineering, to train robots, providing them with the ability to learn, improve, adapt, and reproduce tasks using specific instruments like robotic arms, grippers, cameras, and torque sensors.(Kormushev et al., 2013).

2.1.4 Semi-supervised Learning

Semi-supervised learning is a subfield of machine learning that focuses on utilizing both labelled and unlabeled data to accomplish particular objectives of learning. Positioned between supervised and unsupervised learning in terms of conceptual framework (van Engelen & Hoos, 2020).

Machine learning, or ML, generally distinguishes between two primary tasks: supervised learning and unsupervised learning. Supervised learning involves being given a collection of data points that include an input x and its matching output value y. The objective is to develop a classifier or regressor that is capable of predicting the output value for inputs that have not been encountered previously In the context of unsupervised learning, there is no provision of a specific output value. Instead, individuals attempt to derive an underlying framework from the given inputs. In unsupervised aggregation, the objective is to derive a mapping from the provided inputs (e.g. real number vectors) to groups, ensuring that comparable inputs are assigned to the same category. Semi-supervised learning is an a subfield of machine learning that seeks to integrate these two missions. Semi-supervised learning algorithms aim to enhance performance in any of these two tasks by leveraging knowledge commonly associated with the other. When dealing with a classification challenge, one can utilize additional data points whose labels are unknown to assist in the process of categorization. Conversely, clustering algorithms can improve the learning process by utilizing the information that specific data points are part of the same class (van Engelen & Hoos, 2020).

Several commonly employed methods for semi-supervised learning include expectation-maximization (EM) with generative mixture models, self-training, co-training, transductive support vector machines, and graph-based methods (Zha et al., 2009). Semi-supervised learning deals with the Pavement Condition Assessment like Pavement roughness, paved layer strength, and visual surface stress. Building the prediction model involves merging of labeled data from manual surveys to unlabeled from for instance sensors lasers surface testers, accelerometers or road images. The model can be trained with the help of labeled data, and by means of this dataset, unlabeled data can as well be used so that it is augmented (Liu et al., 2021). For the detection and monitoring of water treatment and supply, but also for checking the potentials for pollution and identification of quality water, semi-supervised learning could be adapted (Yuan & Jia, 2016).

2.1.5 Transfer Learning

In this view, transfer learning can be considered as a reusing strategy of machine learning models that get previously trained on the developing new models. Generalized knowledge can get transferred between the models only when the models are designed to carry out identical tasks. This machine-learning method minimalizes required resources and tagged data for training of new models (Weiss et al., 2016).

Transfer learning is an a priori training focused approach for the improvement of learning in a subject in one domain by building on the learned experiences from a certain different but closely related domain. The scarcity of target training data begets the necessity for transfer learning. Such a scenario would emanate due to varied factors ranging from actual nonexistence of data, costs involved, or high inaccessibility expense. Transfer learning solution becomes an attractive option for the usage of existing datasets related but non-exactly to a domain of interest as big data repositories become more prevalent. For instance, transfer learning has been applied successfully in different machine learning cases like text sentiment classification, and in other cases like image classification, human activity categorization, software fault classification, and multi-language text classification among others (Weiss et al., 2016).

Transfer learning is where instead of developing a scratch model, the already pre-trained models are used. The pre-trained object detection or video analysis models can be further finetuned to improve construction site safety monitoring. This adaptation includes identifying potential hazards, ensuring compliance with safety regulations. Additionally, there are recognition models that determine whether workers are appropriately wearing safety helmets and vests, subsequently alerting workers to unsafe conditions (Lee & Lee, 2023)

Transfer learning can be applied to traffic management and optimization, such as adapting pre-trained models for traffic flow prediction, congestion detection, or intelligent traffic signal control. (Krishnakumari et al., 2018). By reusing models or knowledge from previous BIM projects, transfer learning can help improve the speed at which digital representations are made for new construction or renovation projects.(Zabin et al., 2022). In computer engineering, Transfer learning will fasten the development and verification of integrated circuits as well as hardware components by reusing trained models before their design rule check or fault analysis. (Pan et al., 2021).

2.2 Evolution of Machine Learning

2.2.1 Early Days

Artificial intelligence dates back to the fundamental work of Warren McCulloch and Walter Pitts in 1943. They developed the prototype of artificial neurons based on various theories regarding brain function, logical analysis, and computation. This is an emulator of neurons that are labeled with" on" and "off", responding to input from neighboring neurons.. McCulloch and Pitts demonstrated that any computable function could be computed by networks of these artificial neurons. They also showed how logical operations like AND, OR, and NOT could be implemented with these structures. Importantly, they suggested that such networks could learn. Donald Hebb, in 1949, introduced Hebbian learning, a rule for adjusting the connections between neurons, which remains influential in neural network research. The SNARC, the first neural network computer, was created by Marvin Minsky and Dean Edmonds at Harvard in 1950. This apparatus employed vacuum tubes and components obtained from a B-24 bomber to replicate a network comprising 40 neurons. Despite facing initial doubt from his Ph.D. committee, Minsky proceeded to study universal computation in neural networks. Alan Turing, commencing in 1947, exerted a crucial influence on molding AI. Turing's 1950 article "Computing Machinery and Intelligence" established essential notions such as the Turing Test, machine learning, genetic algorithms, and reinforcement learning. He put out the concept of the "Child Programme," advocating for a shift in AI research from replicating adult intellect to replicating the learning process of a child (Huang, 2010).

2.2.2 Rise of Deep Learning

Deep learning is based on artificial neurons that resemble the neurons found in the brain of human. In the 1950s, the foundation was laid with the invention of the Turing Test, the development of the first computer game of checkers, and the birth of Artificial Intelligence, setting the stage for future advancements. The 1960s introduced us to the first industrial robot, Perceptron, Decision Trees, and the Chain Rule Method, pioneering concepts essential for machine learning. The 1970s brought about crucial algorithms like backpropagation, Support Vector Machines, k-nearest neighbors, and the Neocognitron, advancing neural networks and pattern recognition. The 1980s witnessed the birth of Artificial Neural Networks (ANNs), the restricted Boltzmann machine (RBM), Explanation-Based Learning, and the refinement of the Backpropagation Algorithm, paving the way for deeper network architectures. In the 1990s, Long short-term memory (LSTM) networks and Behavior-based robotics (BBR) emerged, while Deep Blue demonstrated the potential of AI in defeating human champions in complex games. The 2000s marked a significant turning point with innovations like the Deep belief network, Deep Boltzmann machine, Deep Neural Networks (DNNs), and the creation of large-scale datasets like ImageNet, fuelling the deep learning revolution. The 2010s witnessed the rise of game-changing models such as AlexNet, Generative Adversarial Networks (GANs), U-Net for medical image segmentation, and the iconic AlphaGo's victory over human Go champions. In 2020, deep learning continued to evolve with the introduction of Denoising Auto Encoders (DAE), Nash_Qlearning, Reinforcement learning advancements, and Deep inverse techniques, shaping the future of AI (Directions, 2023).

2.2.3 Recent Advances

The academic, industrial, and service applications have faced considerable problems due to the rapid improvements in digital technology for cyber-physical systems in recent years. The widespread use of the Internet of Things (IoT) has resulted in the presence of complex data with several dimensions, noise interference, incomplete and inconsistent information, and large data quantities. Machine learning (ML)-based artificial intelligence models have emerged as powerful tools for data analytics and process optimization across various research areas. Since the last decade from 2012, ML technologies have evolved, proving their practical value in solving complex industrial problems. Applications encompass a wide range of uses, including but not limited to, predictive maintenance, process optimization, work planning, enhancement of quality, supply and demand forecasting, identifying defects, and vibration signal recognition. Machine learning is a leading-edge technology that is prominently utilized in production, service, medicine, and general science to detect and address inefficiencies in different operations (Wang, 2022).

2.3 Traditional Methods in Civil Engineering

2.3.1 Empirical Methods

Empirical methods in civil engineering rely on observations, experiments, and historical data to make decisions and design structures. These methods are often based on practical experience and are particularly useful when there is limited theoretical understanding or when dealing with complex, non-linear, or variable conditions. Some examples of empirical methods in civil engineering include: Load testing of materials and structures to determine their strength and stability, Field testing and monitoring to assess the performance of structures, Empirical equations and formulas for estimating factors like soil bearing capacity, concrete strength, and structural stability. (Jafari et al., 2015), (Mostoufi & Constantinides, 2022).

2.3.2 Analytical Methods

Analytical methods in civil engineering encompass the use of mathematical equations and theories to analyze and design structures. Basically, these methods are based on principles of physics, mechanics, and mathematics. Engineers use analytical methods to determine the behavior of structures under various loads and conditions. Some common analytical methods include: Structural analysis using principles of statics and dynamics to calculate forces, stresses, and deformations in structures. Hydraulics and fluid dynamics calculations for designing water and wastewater systems, including pipes, pumps, and channels, and Geotechnical analysis to study soil properties and predict settlement, bearing capacity, and slope stability. (Jafari et al., 2015), (Mostoufi & Constantinides, 2022).

2.3.3 Numerical Methods

Numerical methods in civil engineering encompass the use of computers and numerical techniques to solve complex engineering problems. These methods are particularly valuable when dealing with intricate geometries or non-linear behaviors. Numerical methods include: Structural design software that uses numerical methods to determine the dimensions and specifications of structural components, Numerical modeling of traffic flow and transportation systems for optimizing road networks and traffic management, and Computational fluid dynamics (CFD) for modeling fluid flow in hydraulic structures or environmental assessments. (Jafari et al., 2015), (Mostoufi & Constantinides, 2022).

2.4 Relevance in General Engineering

Machine Learning is a branch of Artificial Intelligence that is becoming increasingly popular in the fields of computers and data analysis. It involves developing algorithms and models that enable applications to exhibit intelligent behavior. Machine learning, a sophisticated technology, is employed for exploratory data processing. Machine learning algorithms play a crucial role in creating intelligent real-time engineering apps that can effectively analyze data to solve actual-world issues (Jhaveri et al., 2022).

2.4.1 Machine Learning in other engineering fields

Machine learning (ML) is imperative in computer engineering since it facilitates self-governing robots to see and analyze their atmosphere, therefore, making sound judgments. In addition, machine learning techniques have been used to develop robots able to perceive human motion and language and then respond to the same (Martinez-Martin et al., 2020), (Mosavi & Varkonyi-Koczy, 2017). Machine learning is utilized in the support of predicting precise timings for machinery and equipment maintenance, within mechanical engineering, to minimize downtimes and ensure efficiency. Machine learning makes it possible to use machine learning algorithms to enhance the effectiveness of mechanical components and system design, considering numerous limiting conditions as well as performance criteria (Guo et al., 2021). Machine learning (ML) is applicable to electrical engineering, cutting across demand prediction, optimal distribution network development, and grid abnormalities detection. 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 et al., 2021).Chemical engineering applies machine learning (ML) methodologies for better performance as well as control of chemical processes to improve qualities of products as well as low energy consumption. Machine learning is used to predict the properties and relationships of chemical compounds to facilitate the processes of drug discovery (Gao et al., 2022). Data collected using various sensors, and used in monitoring and assessing air quality, water quality, and other environmental parameters are possible subjects of analysis by machine learning models in environmental engineering. It is also employed to evaluate the condition of infrastructures for water and wastewater, enhance methods of treatment, identify and characterize sources of contamination, as well as undertake life cycle analysis. Machine learning models offer hope in the ability to predict natural disasters, including earthquakes and hurricanes, which would go a long way in advancing disaster preparedness and response (Zhong et al., 2021). This is because material engineering is updated in the new materials that would need specific characteristics that could be well predicted by the use of machine learning (ML). As the formation of graphs had elucidated on then the application of ML in the development of highly sophisticated material is to the advantage of this field. The quality control check of industrial operations to get the materials as per specification can also be useful work done by machine learning. In this area, machine learning helps in predicting material tasks, design of materials, characterization of materials, and mining activities (Stergiou et al., 2023). Furthermore, optimal flight control systems of Aerospace Engineering may have machine learning applied to them. Machine learning models can analyze data from fitted sensors on planes and make use of the results to detect maintenance needs as well as predictions (Le Clainche et al., 2023).

CHAPTER III

Machine learning applications in civil engineering

Algorithm development and implementation of machine learning represent multiple aspects of statistics, including structural health monitoring geotechnical and integrity structure seismic engineering among others (Barkhordari et al., 2023).
In recent times, the field of civil and structural engineering has seen many applications for machine learning techniques developed by the scientific community. There are many types of ML applications developed based on seismic engineering, structural feature identification procedures, and the major elements in their scopes. Such applications attempt to develop mathematical tools that can be used for solving intricate input-output problems (Ruggieri et al., 2022).

3.1 Structural Health Monitoring

Structural health monitoring (SHM) refers to the scientific field that deals with determining and observing whether a specific structure is healthy. Structural health monitoring systems are based on sensing systems and structural models for assessment of the condition structures and machines (Gomez-Cabrera & Escamilla-Ambrosio, 2022).
Such as the large bridges, dams and high rise buildings are more prone to functional weakening due corrosion or stress. This unavoidable circle demands great preservation. While there are numerous obstacles, onsite investigations can bring the temporary closure of bridges or building in order to examine and based upon ethical considerations. Over this approach, several researchers have suggested SHM procedures. Structural Health Monitoring is one of the latest technologies which has grown rapidly in recent years. One of the main uses for sensor development is Structural Health Monitoring (SHM). Early detection of damage reduces the cost and time for fixing minor damages. The main objective of Structural Health Monitoring (SHM) is to predict or detect the emergence Such structural failures, as well as earthquakes, waves vehicles or environmental vibrations can cause significant physical damage to the infrastructures (Singh et al., 2020).

3.1.1 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 Structural Health Monitoring (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 et al., 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 et al., 2005). Based on 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 MLPand RNN and CNN. The MLP is a neural network with input layers, one or more hidden layers and an output layer. 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 et al., 2021).

3.1.2 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 (Zaparoli Cunha et al., 2023). However, data-driven ML provides alternative approaches for structural reaction prediction. ML approaches have direct input-output modeling using certain function approximators such as random forests and NN based on the given training data. For dynamic loading, 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 materials's properties. The results are the field variables that we want to identify, which include displacements, stresses, and strains. For instance, Keshtegar et al. used support vector regression in predicting the shear strength of steel fiber-reinforced concrete beams among others. In the research by Alwanas et al., the extreme learning machine was used for modeling the load-carrying capacity and failure mode of a beam-column joint connection. Li et al. applied NN and various conventional ML approaches to predict the consequences of gas blows (Li et al., 2023).

3.1.3 Optimization of Monitoring Systems

Machine learning can be integrated with sensor networks and SHM systems to provide real-time monitoring and early warning systems for structural health.

The optimization process typically includes the following steps: definition of the optimization problem, specification of the objectives to be maximized or minimized, selection of decision variables, consideration of constraints, and formulation of final models. The intended goal is specified as the objective function, which consists of variables and restrictions, which are functional relations of the inequality and equality of the variables. Optimization algorithms have been utilized in various areas of SHM systems, including sensor system design and structural damage detection. As a result, optimization algorithms can assist in determining the ideal number of sensors and the best sensor positions. Based on the variables of the problem, a set of objectives is defined. The error function is divided into single- and multi-objective functions. An important aspect to consider is the choice of a suitable optimization algorithm, which should be selected based on objective function types. In one-step or multi-stage damage detection methods, objective functions may be used to identify the location and level of damage. In the one-step approaches, application methods are used to determine how far and where the damage occurred whereas in multi-step methods it detects how much damage has been done and where did this happen (Hassani & Dackermann, 2023).

3.2 Traffic and Transportation Systems

The purposes of machine learning in transportation are numerous and include data analysis, identification of road problems, congestion forecasting, predictive maintenance illustrating introduction about a description or fact However, the significance of ML is further augments in this field since it supports efficient transportation rerouting during emergency (Silva et al., 2020).

3.2.1 Prediction of Traffic Congestion

For the growing need for traffic prediction technologies, there are different approaches proposed to predict congestion scenarios. Several machine learning algorithms including Neural Networks (NN), Support Vector Machines (SVM), and Regression Analysis have been applied for predicting traffic congestion because they can effectively deal with large datasets while learning patterns from data. For example, Yisheng et al developed a new deep-learning framework to forecast the flow of vehicles under different urban road networks using modern developments in dimensional However, this approach knows how traffic flows and shifts over 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).

Lee et. al. 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 fairly accurate, hitting a 75.5% accuracy rate. But they missed including one important detail: the time of day, which can greatly influence traffic conditions (Yasir et al., 2022)..

Akbar et al. used a mix of two techniques, Complex Event Processing and Machine Learning, to try to predict when and where traffic jams might happen. They created a special formula called Adaptive Moving Window Regression (Yasir et al., 2022).

3.2.2 Optimization of Traffic Flow

Due to the swift expansion of urbanization and economic progress, the issue of traffic congestion has escalated significantly. Therefore, accurate measurements of traffic congestion are necessary for monitoring the status of road transportation and improving transport operations. To address urban traffic congestion, an intelligent transportation system (ITS) must have real-time data and a self-driving mechanism for evaluating traffic conditions. The ITS has the main challenge of conducting traffic control and orders. For effective management and control, short-term real-time flow traffic estimates are always required (Bharti et al., 2023).

Predicting models can generally be classified into two main groups: Traditional forecasting models and neural network-based ones. Generally, the conventional traffic flow prediction models provide moderate accuracy and poor resistance to external perturbations essentially because of the non-linear/non-stationary nature of traffic flow. On the other hand, neural network models have gained much popularity over recent years and are highly efficient in processing massive amounts of data especially due to their remarkable learning and adaptive abilities (Bharti et al., 2023).

One of the prominent deep learning technologies is The Recurrent Neural Network (RNN) which provides a variety of benefits including server storage, parameter sharing, fine-tuning accuracy, and nonlinear extraction features in predicting traffic flow. There are several types of models presented as RNN, with long-term short-term memory Neural Networks (LSTM) and Bi-directional Long Short Term Memory Neural Networks (Bi-LSTM) being particularly high in performance efficiency. Through this neural network, the LSTM method was proposed to predict short-term traffic flow because of long-term dependency on the series data for the movement. A Bi-LSTM model was then created and it seemed that the Bi-LSTM approach performed better in terms of prediction accuracy (Bharti et al., 2023).

4.2.3 Improvement of Traffic Safety

To improve the safety of traffic, it is essential to identify the possible risks of roads by conducting a profound analysis, Machine learning (ML) techniques have been widely used in various domains with the aims of safety prevention and risk detection. This includes activities such as the prevention of fraudulent traffic, preventing data channelizing of Internet of Things networks ensuring information safety, and even improving transport 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).

3.3 Water Resources and Environmental Engineering

Machine learning has been commonly used as a potent tool for addressing issues in different domains of water treatment and management systems. These domains cover real-time monitoring, anticipation, tracking the source of contaminants, estimating pollutant concentrations, allocating water resources, and optimizing water treatment technologies (Zhu et al., 2022).

Machine learning is increasingly becoming recognized as a very effective and versatile technology across various scientific disciplines. In addition to the information technology sector, machine learning is increasingly being applied in the fields of medical research, agriculture, and the legal industry. In contrast, it has been over thirty years since the introduction of artificial intelligence (AI) applications, such as genetic algorithms, fuzzy logic, and neural networks, in the field of hydrology. In 1992, French et al. attempted to predict short-term rainfall using a neural network consisting of three layers (Rozos, 2019).

3.3.1 Flood Prediction

Floods are some of the most devastating tragedies, resulting in extensive harm on the lives of people, infrastructure, farming, and social and economic structures (Mosavi et al., 2018). Based to the Organization for Economic Cooperation and Development, floods result in damages of over $40 billion annually on a global scale. The majority of nations lack effective flood warning systems. India accounts for 20% of flood-related deaths, as reported by the Central Water Commission. Bihar is the most noticeably horrible impacted state, with nearly 73% of its full surface territory getting swamped every year. In 2018, the reported cost of damage to infrastructure, crops, and public utilities across India amounted to almost 3% of India's gross domestic product (Kunverji et al., 2021). Hence, governments face the imperative to create dependable and precise maps of flood-prone regions and subsequently devise strategies for sustainable flood risk management, with a particular emphasis on prevention, protection, and preparedness. Over the past twenty years, machine learning (ML) techniques have significantly contributed to the improvement of prediction systems, offering highly accurate as well as affordable solutions (Mosavi et al., 2018).

In the past, natural disasters such as storms, rainfall/runoff, shallow water conditions, and flow hydraulics were predicted using numerical and physical models. These models, however, could not correctly predict flooding. Hence the use of advanced data-driven models like machine learning (ML) was recommended instead. Factually indeed, flood modeling has a long heritage of using data-driven models that have gained recent popularity as well. The data-driven approaches to prediction are then used, which use observed climate indices and hydro-meteorological data to enhance accuracy and understanding. For the flood frequency analysis (FFA), the common approaches for modeling flood prediction are statistical models such as autoregressive moving averages (ARMA), multiple linear regression (MLR), and autoregressive integrated moving averages (ARIMA). FFA is one of the first statistical techniques ever used for flood prediction. The modern development of machine learning (ML) techniques in the last twenty years has seen their inundate effectiveness in flood prediction well above traditional approaches. ML models were employed for prediction with higher precision compared to conventional statistical models. Many ML techniques like artificial neural networks (ANNs), neuro-fuzzy, support vector machine (SVM), and support vector regression (SVR) have been reported successful both in short-term and long-term flood prediction (Mosavi et al., 2018).

3.3.2 Water Resources Management

Water resources management (WRM) has received high attention in several recent developments in machine learning (ML) applications. The challenge of big data development provides significant improvement in the ability of hydrologists to tackle contemporary challenges and novel applications of machine learning have been encouraged (Ghobadi & Kang, 2023).

Raman and Chandramouli at the time used a feedforward neural network (FFN) in water resources management, in developing an operational guideline for a single reservoir used for irrigation purposes. Since then, FFN applications have advanced with increased complexity and frequency. Chandramouli et al. trained a Feedforward Neural Network (FFN) for predicting the best releases from a system of three reservoirs. They utilized data that was acquired from the dynamic programming method where they simulated a system run for 36 years with two week time step in the simulation. Cancelliere et al. used the model of soil-water balance, a model of dynamic programming as well as a neural network to make operational rules of the reservoirs for irrigation purposes. Other researchers on the other hand applied much more artificial means such as Chang et al., who adopted an adaptive network-based fuzzy inference system for computing the most suitable water discharge from a solitary reservoir in Taiwan. The operating rules of the best reservoir operating histogram produced up on the implementation of the genetic algorithm were combined with the knowledge of the currently applicable operating rules to train the FFN. According to their investigation, this strategy gave much better results as compared to the conventional (Rozos, 2019).

3.3.3 Water Quality Improvement

The growing and often more challenging problem is the contamination of rivers by as many point sources as by diffuse causes. It further affects aquatic systems and freshwater supply for drinking or irrigation (Bui et al., 2020). To evaluate the water quality and anticipate trends, it is crucial to conduct a specific study by using sophisticated technology (Khan & See, 2016).

Previously, numerous techniques have been devised worldwide to observe and evaluate the quality of water. These encompass multidimensional statistical techniques, fuzzy inference, and the water quality index (WQI) (Zhu et al., 2022). The geo-statistical methods that have been utilized include kriging, transitional probability, multivariate interpolation, and regression analyses, etc. (Khan & See, 2016).

For instance, Statistical models are designed to extract general rules from experimental data by utilizing information obtained from field data. The process of statistical modeling and assessment entails the careful selection of analysis techniques and the validation of assumptions and data. Many of these models are complex and require a significant amount of field data for comprehensive analysis. However, a challenge with numerous statistical water quality models is their assumption of a normal distribution and a linear relationship between prediction and response variables. Given that water quality is influenced by multiple factors, conventional data processing techniques are often insufficient to handle the complex non-linear relationships between these parameters and water quality predictions. Consequently, the use of statistical techniques can result in less accurate outcomes (Najah Ahmed et al., 2019).

Artificial intelligence (AI) has developed techniques that optimize operations, select equipment, and solve problems involving large data sets that can hardly be more processable by a computer than by a human for making decisions. AI methods are preferable to such data simply because they excel at reproducing and compensating for this lack. Through efficient parallel computing technology development and increased computational power, researchers are now able to employ AI approaches such as Artificial Neural Networks (ANNs) and the Adaptive NeuroFuzzy Inference System (ANFIS) in deriving solutions for modeling field data. The application of the neurofuzzy technique seems to increasingly be in demand for various other related fields of water resource engineering, such as the application to rainfall-runoff models and basin operations. Researchers appreciate ANFIS because it can improve evaporation estimation from day to day, reservoir water level prediction, and river flow forecast (Najah Ahmed et al., 2019).

3.4 Construction Management and Planning

Construction management (CM) is crucial on all construction sites, and machine learning provides an innovative way to deal with this issue. In recent years, however, machine learning has been shaping the construction industry by being used in a lot of construction applications. The technology is revolutionizing some of the aspects of construction project management, including risk evaluation and mitigation, working as a supervisor for worksite safety protocols, estimation, and prediction of costs, a schedule manager for the project, as well as predictions on building energy use. (Nguyen Van & Nguyen Quoc, 2021).

3.4.1 Construction Schedules Optimization

Construction scheduling in infrastructure and building projects consists of the establishment of project policies and processes, which are later further broken down into particular construction tasks. A total construction plan again considering constraints like sequence and resources minimizes budget overruns, delays, and safety and quality issues. Typically, there are some steps in any scheduling process. They include the definition of the project scope, identification of individual activities, and establishment of their dependencies (e.g., finish to start, start to start, finish to finish). (Wu et al., 2023).

In the past, construction planning was done with tools such as pencils and paper which were manual, time-consuming, and with probability of errors therein. To improve on the same, graphical representations such as Gantt charts, AOA network diagrams, and AON network diagrams have been developed in the quest to enhance planning and identification of delays. While start and end times for activities are defined as AOA diagrams, AON diagrams are characterized by an arrow of activities and nodes of dependency. (Wu et al., 2023).

One of the other significant functional uses of machine learning is in optimizing construction schedules. Machine learning helps to predict, model, and optimize the sequence and planning of tasks, especially in construction projects using data-driven techniques. It uses the application of historical and real-time data using machine learning to provide intelligence usable in making predictions on the time duration for tasks, resource allocations required, and also possible delays. For several years now, there has been increased research in trying to study machine learning methods used for optimizing construction. Artificial Neural Networks (ANN) have been in the limelight, especially. Kog et al. have developed an ANN model for determining the percentage effect taking into account variables like project manager experience and monetary incentives extended to designers on the varied schedules. Similarly, Attal trained a series of ANNs on highway project duration and cost data to check out key project features impacting duration predictions. Hola et al. Adopt a much more specific approach which was predicting earthworks durations with ANN, Bhokha et al. explored building construction durations with similar neural network models (Fitzsimmonsa et al., 2020).

3.4.2 Construction Safety Improvement

The construction industry is one of the most dangerous sectors in many countries. About the risk management process, top managerial officials should offer more than adequate safety information, since they take care of portfolios dealing with construction projects. This understanding would facilitate intervention in a more proactive form. Techniques, predictive analytics, and machine learning have been leveraged by data scientists to enhance decision-making at all levels in construction safety substantially. Machine learning deploys algorithms to unlock insights and predictive analytics previously out of reach with large amounts of data from sources (Poh et al., 2018).

Machine learning (ML) is a revolutionary way to improve safety at the construction site. In use with drones, ML can survey sites from the air, and algorithms processed using the imagery captured will flag out potential safety risks like weak scaffolding or open pits. These provide them with an understanding of site conditions so that they can then define their corrective actions. In addition, these ML algorithms can recognize through camera and lens capabilities if the workers are wearing their safety equipment like helmets, vests, and harnesses properly or not. The site supervisors are immediately informed in case of any violations. The huge construction sites are notorious for almost always having a chaotic movement pattern of people and vehicles. In such scenarios, ML algorithms can help by predicting traffic patterns and suggesting the best route for a given vehicle at any point in time. This ensures that the operations are smoothed out as well as risks are reduced over accidents or collisions over time. The supervised unsupervised combination of machine learning techniques is used for undertaking these tasks. Techniques such as linear regression, logistic regression, decision trees, as well as support vector machines (SVM) are put to use for such purposes. With the help of these techniques, it is possible to predict numerical values, for example - to predict the probability of safety incidents in certain conditions, and to classify types of safety incidents, reporting the conditions under which those of this type occur. Convolutional neural networks (CNNs) find use in elaborate pattern recognition because they possess the ability to study complex patterns, especially so in image analysis research, where they can be used for identifying fine differences like the usage of safety gear by workers. For sequential data, however, that involves monitoring via a pendulum event that may lead to occurrences of incidents the corresponding trains of recurrent neural networks (RNN) and long short-term memory (LSTM) is vital (Choi et al., 2023), (Alkaissy et al., 2023).

3.4.3 Cost Reduction

Managers need accurate cost estimating while deciding on minimizing the time-related risks in the evaluation process (Tayefeh Hashemi et al., 2020). Objective cost prediction is highly desired. Machine learning techniques have won great success, so the task of building cost estimation is much simpler to automatize and the biases, which were caused by the human factor, are suppressed. Hegazy and Ayed have developed a model with an artificial neural network (ANN) for the estimation of expenses concerning highway development (Hegazy, T., & Ayed, A. 1998). Generally, using various machine learning algorithms, such as Linear Regression, Decision Trees, Neural Networks, and Support Vector Machines, ML can analyze past projects to forecast the cost of upcoming ones. This aids in budgetary planning and avoids possible overrun costs. Moreover, ML can improve the prediction of the ordering schedules of materials and quantities based on predictive models that could result in volume discounts as well as reduced storage costs. Through the evaluation of historical data, ML can also predict the optimum number of resources, both human and material among other strategic predictions essentials for a project. This prerogative prediction averts both over-allocation and inadequacy utilization of resources which are human beings or materials hence no unnecessary expenses. Furthermore, ML can be used to fine-tune decision-making processes such as ascertain the optimal allocation of resources in construction projects (Kovacevic et al., 2021).

Materials Engineering

Materials engineering can be greatly accelerated through the use of machine learning. This process not only enhances and simplifies discovery, development, and deployment procedures but also critically analyzes highly complex data to predict properties and performance levels However, this technological revolution brings around optimal production processes that promote quality control and enhance innovation thereby evidencing significant improvement. Innovative applications of structural materials do not only advance improvements in load-bearing capacities such as strength, hardness, and toughness but also set up to address new restrictions and needs essential for contemporary technology (Sparks et al., 2020).

3.5.1 Material Properties Prediction

Recently, machine learning techniques have become popular as cost-effective and efficient tools for predicting the influence of material characteristics on actual quality, costs and schedules related to proposed mixtures (Shanmugasundaram et al., 2022). ML models have become a commonly used tool that can be considered as an effective mechanism to predict the mechanical properties of concrete. Common ML techniques employed for evaluating concrete strength can be categorized into four primary types: artificial neural networks (ANN), support vector machines (SVM), decision trees, and evolutionary algorithms (EA). These models are typically used together with a large dataset that is often divided into training (TR), validation (VAL), and testing (TS), where the subsets TS and VAL consist of elements that were unseen during TR. For example, ANN was used which is a machine learning approach to determine the 28-day compressive strength for engineered cementitious concepts (ECC) when pozzolan including fly ash and ground granulated blast slag (GGBS) added as industrial waste materials in their structure by Shanmugasundaram et al. The experiment was conducted using the pile amounts and physical properties of Polyvinyl Alcohol fibers (Shanmugasundaram et al., 2022).

3.5.2 Material Life Cycle Assessment

Life cycle assessment (LCA) is a method through which a systematic process involves the collection and evaluation of the physical resources as well as waste due to a system or product, and the environmental effects associated with it during the whole lifespan. Traditional Life Cycle Assessment (LCA) can be time-consuming and prone to falsehoods. The LCA provides a framework for the scope and purpose of assessment as well as analysis of inventory. ML algorithms are capable of handling complex datasets while incorporating issues such as geographical location; and climate conditions to set examples that can also local production methods and more precise environmental impact assessments. With its ability to analyze historical and real-time LCA data, ML can aid architects, engineers, or project managers in choosing building materials that fulfill certain environmental criteria such as building research establishment environmental assessment method (BREEAM) or leadership in energy and environmental design (LEED) while optimizing the sustainability performance tradeoff. (Ghoroghi et al., 2022), (Schwartz & Raslan, 2013).

3.5.3 Material Cost Analysis

Historical information analysis using ML algorithms allows for estimating future prices for building materials. They evaluate historical patterns and trends of material costs based on inflation, market demand, supply chain disruptions, or even possibly geopolitical events impacting prices. This innovative technique is not something theoretical and has real applications in the field of civil engineering. For example, Yeh (1998) applied a hybrid model which was the combination of a back-propagation neural network and statistical regression to reveal preliminary amounts needed for primary materials in civil buildings especially the amount of steel used on beams and columns as well as concrete quantity along with formwork. As in a practical case of the same application, Bakhoun et al. (1998) used an artificial neural network to calculate how much concrete was needed for a bridge project implemented within Egypt's territory Such cases demonstrate the pragmatic applicability and validity of machine learning to complex predictive cost calculations as well for material supplies estimates in construction engineering (Pham & Nguyen, 2023). ML also serves an instrumental role in the prediction of construction waste, helping substantially reduce such wastes at source by forecasting generation during each project phase. This process is centered on the creation of predictive models based on ML techniques, which largely rely upon Decision Trees and K-Nearest neighbor (KNN) Algorithms. All these models are based on data mining and statistical analysis to detect trends in collected information, crucial for the right waste estimation at different construction stages. Specifically, Decision Trees are used mainly for multi-variable analysis and present data in branch form to facilitate a detailed understanding. (Gulghane et al., 2023).

Geotechnical Engineering

As a part of geotechnical design, the determination requires an appropriate calculation between cost and engineering safety regarding the physical and mechanical qualities of soil. This entails the assessment of bearing capacity and long-term deformation in soil. The field of geotechnical engineering is embracing more and more the integration of machine learning (ML) to solve areas that are particularly complicated or difficult, some aspects about which better performance can be achieved. It holds a strong potential for predicting geotechnical design, stability, and settlement as well other soil properties-related decisions such as site selection and foundation solution optimization. The machine learning process provides an advanced method for predicting risk assessment that applies not only to current but also future site conditions. In particular, ML can be applied in zones that are susceptible to water running and volcanic activity from which valid knowledge regarding possible impaction of landslides is obtained (*Machine Learning‐Based Modelling of Soil Properties for Geotechnical.Pdf*, n.d.).

3.6.1 Soil Mechanics Analysis

Previously, the classical approaches to soil classification were highly based on experiments done in the lab. For example, experimentation involving the Standard Penetration Test (SPT), Cone Penetration Test (CPT), and other standard modern soil mechanic tests. However, in recent years, research has been showing that tools like Support Vector Machines (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Decision Trees (DT), and Naïve Bayes algorithms prove to be useful to understand soil types. Soil classification based solely on soil type can be difficult since many factors affect soil characteristics and do not always correspond to obvious types of soil. But with machine learning, including the probabilistic predictions of the Naïve Bayes algorithm, we can better understand and classify soil. Describing the behavior of soil in various scenarios, the modern techniques are much better than the ancient methods (Samadi & Samadi, 2022). In another research work, methods like the use of Support Vector Classification (SVC), Multilayer Perceptron(MLP), and Random Forest (RF) were adopted for classifying soil. Among these, the best performing was reported to be Support Vector Classification (SVC). Furthermore, other investigations targeting soil permeabilities in their assessment have used machine learning algorithms, specifically Gradient Boosting (GB). Even more enhanced performance metrics were tools such as SHapley Additive exPlanations (SHAP) as well as Partial Dependence Plot 1D (PDP 1D), this being integrated into the analysis (Aydın et al., 2023).

3.6.2 Foundation Design Optimization

In this regard, the prediction of soil behavior, estimation of load-bearing capacity, simulation of structural responses, and optimization of design-leading parameters concerning cost, safety, and sustainability can be executed by machine learning (ML) algorithms in foundation design optimization. For example, Pile foundations are very common where the construction is under heavy superstructures, and their design deals with the handling of complexities as well as large uncertainties. Machine learning technology has gotten tremendous development and application in various fields including Civil Engineering. Of all the machine learning approaches, the Bayesian network is an oddity in that it can draw up data and reason out solutions from different problems that involve complex and uncertain matters. In many pursuits, researchers have tried to adopt the Bayesian network in civil engineering. However, Zhang et al. propose a new approach that includes a machine learning method based on the Bayesian network specifically aiming to learn the cross-site variability as well as the site-specific statistics of model bias factor. This is an innovative approach not only the purpose of the accomplishment of cost-effective pile foundation designs but also provides a valuable tool for implementing reliability-based design of pile foundations specifically which is a remarkable advancement in this area (Zhang et al., 2020).

3.6.3 Landslide Prediction

Landslides are one of the most uncontrollable and damaging kinds of natural hazards, whereby only sophisticated prediction systems could enhance their mitigation. Coincidentally, though a few early warning models exist using the Internet of Things to monitor the environment in general, there is a recent tendency towards the utilization of machine learning techniques (MLT) (Kuradusenge et al., 2020). According to Kuradusenge et al., most of these MLTs particularly consider rainfall data in combination with internal parameters offering an optimum chance for improved landslide hazard prediction. In the landslide prediction, two Machine Learning Techniques (MLTs) of paramount significance include Random Forest (RF) and Logistic Regression (LR). The choice behind these two MLTs over others includes primarily their extensive application in the respective field. Specifically, RF succeeds in highlighting the importance of some parameters and its quick training makes it effective for application. On the other hand, LR has impressive skills in estimating regression coefficients and proves quite instrumental in clarifying how some factors influence each specific landslide occurrence (Kuradusenge et al., 2020). according to Korup & Stolle, Data mining and machine learning are emerging as common tools for modeling landslide susceptibility with a surge in related studies observed. Most of these studies are centered on Europe and Asia especially countries like Italy, China, and India. For that, the most popular methods are Logistic Regression and Artificial Neural Networks (ANN). The key parameters for predictions include hillslope gradient, hillslope aspect, and lithology. Other elements that influence predictions of landslide susceptibility are topography, the position of the site with concern to drainage networks as well and distance from tectonic faults (Korup & Stolle, 2014).

3.7 Structural Engineering

Essentially, Machine Learning (ML) is a set of techniques, which allow for the automatic discovery of underlying patterns in data. Once such patterns have been detected, they can also be used to induce models of the future that may help make decisions under uncertainty conditions. Methods based on machine learning (ML) have been applied to problems of building structural design and performance assessment (SDPA) since the late 1980-s. Adeli and Yeh (1989) were among pioneers in developing and using ML-based approach to a beam design problem in this field. After the pioneering, several series of studies of construction of SDPA problems using artificial neural networks (ANN) were subsequently done in the 1990s. These smart systems have proved useful during prediction and evaluation of structural performance, identification of structural conditions and providing useful information towards proactive and remedial decision making using data from a variety of sources and media collected. (Sun et al., 2021).

3.7.1 Structural Design Optimization

machine learning (ML) and structural design optimization collaborate to enhance the efficiency and effectiveness of engineering design as they act to assist in identification of very best structures. This could enable affordable safer cost-effective as well as innovative building solutions when used with advanced ML algorithms for quick analysis of big data, identifying ideal design parameters and predict structural performance. For instance, on this three benchmark models for the best structural designs regarding the achieving least volume and weight as well as cost reduction with two optimization strategies selected along with an Artificial Neural Network (ANN) method has been adopted by BEKDAŞ et al. for a stronger study and analysis. Explore different Harmony Search (HS) parameters such as Fret Width (FW) and Harmony Memory Consideration Rate (HMCR), modifications within iterations, and population size to find the best configuration for accuracy. The use of ANN proved to be pivotal to tailoring models that are robust and drawn from optimal variables and top objectives of the presented structures towards enhancing efficiency and efficacy in structural optimization (Bekdaş et al., 2021).

3.7.2 Structural Load Analysis

Machine learning (ML) methods are engaged in the predictions and analyses of the number of loads for engineered structures. Nowadays, much more frequently researchers and practical designers cover this topic popularizing ML solutions (Truong et al., 2022). Yang et al. (2016) from Peking University employed a BP neural network to detect the shifting load of a bridge. Sensitivity from input and output variables were analyzed, whereby the effect from various combinations of activation functions and methods on load accuracy are also discussed (Yang et al., 2016). In addition, the investigation of the durability of concrete-filled steel tubular (CFST) columns was also carried out in terms of the machine learning algorithms according to the forms of gradient tree boosting (GTB), deep learning (DL), categorical gradient boosting (Catboost), and support vector machine (SVM). For instance, better expressions of response have been realized through the application of the extreme gradient boosting (XGBoost) algorithm in analyzing stiffened plate girders landing to partial load resistance than was previously put in the standard design codes, specifically EN-1993-1-5 and BS 5400. Some other applications of machine learning in structural design include intelligent recognition of fire-vulnerable bridges, steel truss safety assessment, classification of ultimate load-carrying capacity of steel frames, shear strength prediction in steel fiber reinforced concrete beams, load capacity in cold-formed stainless steel tubular columns, and damage characterization of composite laminates under compression (Truong et al., 2022).

3.7.3 Structural Sustainability Assessment

Both the construction industry and its different sources such as pavement engineering, geotechnical engineering, concrete mix design, and structural engineering have officially recognized sustainable development as its important mission. Machine learning and neural networks (NN) have been recently applied in structural engineering primarily for predicting and modeling elastic properties of materials, compressive and bond strength of concrete, buckling load, development of cementitious composites, and improving finite element models. In this context

Machine learning (ML) stands as a transformative approach in enhancing structural sustainability assessment, fostering a more environmentally responsible and resource-efficient approach in the construction sector. ML algorithms efficiently process complex, multi-dimensional data, providing robust predictive models and insightful data analysis crucial for assessing and improving a structure's lifecycle sustainability and using sustainable structural materials like green concrete. These innovative techniques facilitate comprehensive evaluations of material durability, energy efficiency, and overall environmental impact, significantly influencing decision-making processes in structural design and construction. Furthermore, ML's predictive analytics play a pivotal role in forecasting potential structural vulnerabilities, optimizing maintenance schedules, and extending the lifespan of infrastructure, thereby promoting sustainability. The integration of ML in these assessments is instrumental in not only meeting but exceeding the increasingly stringent global sustainability standards and regulations (Naseri et al., 2020),(D'Amico et al., 2019).

CHAPTER IV

Methodology

4.1 overview

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

4.2 Interpretability of Parameters and Feature Selection

4.2.1 Importance of Parameters Interpretability in Machine Learning

Interpretability of parameters forms an essential aspect while conducting data analysis and developing machine learning approaches. Through interpretability, it would be possible for a person to judge the effect as well as the importance that is attached to various parameters in any given system(S. Li et al., 2023). It can be defined which parameters of the model determine the outputs by methods of interpretability. Such awareness accompanies decision making and contributes to the choice of the analysis approach (C. Li et al., 2022). Besides the above, interpretability also assists in uncertainty reduction, reducing training time, and overfitting mainly associated with diagnostic models. Furthermore, the interpretation of models enables to identify parameters which have sense of some physical or physiological meaning, making them more describable for conditions modeling and therefore applicable for real-life applications. Although as a general guideline, the interpretability of the parameters is important in enhancing the comprehensibility, trustworthiness, and efficacy of models of data analysis and machine learning (Lema, 2018).

4.2.2 Different types of interpretability Parameters

Within interpretability in machine learning, explanations of model behavior can be divided onto local and global methods. Local one are those that concern an individual prediction, global ones have to be associated with the whole model. Additionally, a distinction can be made between feature effect and feature importance methods (as illustrated in Fig. 1). A feature effect method indicates the direction and magnitude of a change in the predicted outcome due to changes in feature values. In contrast, feature importance methods quantify the contribution of a feature to the model's performance (e.g., via a loss function) or the variance of the prediction function.

	Local	Global
Feature Effects	ICE LIME Counterfactuals Shapley Values SHAP	PDP ALE
Feature Importance	ICI	PI PFI SAGE

Figure 4.1

A set of global or local model-agnostic interpretation methods based on affect or importance.

4.2.3 Global Methods

Although there are many various types of global methods as illustrated in (fig.1), two of them was used in the thesis and are accumulated local effects (ALE) and permutation importance (PI).

4.2.3.1 Accumulated Local Effects (ALE)

Accumulated Local Effects (ALE) is a technique used for interpreting the predictions of machine learning models, especially in the context of feature effect analysis. ALE focuses on understanding how changes in a feature affect the prediction of a model, locally, over small regions of the feature's distribution. Unlike some other methods, ALE accounts for the interactions between features and is less influenced by correlated features. In ALE, the feature space is divided into small intervals, and for each interval, the average change in the model's prediction is calculated as the feature values vary within that interval. These local effects are then accumulated over the range of the feature to provide a global picture of how the feature influences the model’s prediction. This approach helps in understanding the relationship between a feature and the target variable in a more nuanced way, especially in cases where this relationship is complex or nonlinear.

4.2.3.2 Permutation importance (PI)

Permutation Importance is a technique used to estimate the importance of individual features to the predictions of a machine learning model. It assesses how much the accuracy of a model degrades when taking out the information brought by a feature. It does so by randomly shuffling the values of each feature in turn in the data set and observing the change in model performance. The rational behind this technique is that, permuting the values of an important feature should highly degrade the model's performance since the model depends on that feature to make accurate predictions. Conversely, permuting a less important feature should have a minimal impact on model performance. For this reason, the permutation importance method is considered global since it approximates the importance of features to the whole dataset, giving an overall view of how useful each feature was in predicting the target.

4.2.4 Local Methods

4.2.4.1 SHapley value

Shapley value is a technique in machine learning for the interpretation of individual feature contributions into a predictive model. In this case, each feature stands as 'player' within a game where accurate predictions are made after all. Using the Shapley value method which considers all possible combinations of features, it determines how addition of a particular feature affects the prediction and the contribution by each feature. The calculation involves taking the average of the contribution of that feature across these combinations and therefore accounts for both its stand-alone effect and dependence to other features. This approach gives detailed insights into why certain predictions of a model were made by attributing each part of the output to a feature for understanding on what influences by the features. Although the Shapley value provides a complete and fair way to assign contributions, it can get computationally costly in models that consist of many features, hence significant practical implications of its application in complex settings.

4.5 Machine learning models

4.5.1 Artificial neural network (ANN)

An artificial neural network ANN is the computational model that imitates the structures as well as functions of biological neural networks in involving our human brain. It therefore is an arrangement of connected processing units, which are known as neurons or nodes and put in layers. Indeed, artificial neural networks ANNs could be called the nonlinear models with their ability to spark complicated relations between both the output and input data long back. One type of machine learning is deep learning, and artificial neural networks ANNs are also sub-components. ANNs have found wide applications in different fields from image and speech recognitions, natural language processing to recommendation systems and even intelligent systems like self-driving cars.

4.5.2 Support vector machine learning (SVM)

Support vector machine learning is a powerful algorithm in ML. Its classification capability allows the use of SVM method to regression type issues. Support vectors refer to specific points of data or coordinates on a 2-dimensional plane where a hyperplane like a line or circle can be precisely identified. SVMs try to find the hyperplane that maximizes class margin or minimizes regression error. Margin is the difference between the hyperplane and closest support vectors of each class. SVMs allow linear and non-linear classification. Linear classification uses a linear hyperplane to classify. Kernel functions, for instance, polynomial or radial basis functions can convert data into spaces of higher dimensions such that SVMs can determine non-linear classification decision boundaries. SVMs can deal with nonlinearity data by getting kernels, processing multi-dimensionalities of the data, and stopping overfitting while regularized. Both types of classification and regression are possible. SVMs are applied in many fields such as image, text, and face recognition (Prakash et al., 2023).

4.5.3 The Extreme Learning Machine (ELM)

The Extreme Learning Machine ELM is an extremely powerful training algorithm that is specifically developed for the single-hidden layer feed forward neural networks SLFNs. This algorithm aims to enhance the operation of SLFNs as it deals with some restrictions present in typical neural network learning algorithms like BP. Unlike traditional learning algorithms, ELM does not require users to set parameters like the rate and epochs of learning. It operates automatically without cyclic tuning and thus minimizes the need for user presence. ELM is known as the fastest learning algorithm in comparison to traditional neural network methods. This gives a significant advantage in situations where quick model learning is essential. ELM's ability to approximate complex functions has been identified, especially using additive or RBF activation functions.

4.5.4 Boosting models

4.5.4.1 Gradient boost

Gradient boosting is a machine-learning approach used for regression and classification. It describes booster, which means to take poor learners and combine them to produce a strong one. Unlike Random Forest, Gradient Boosting builds trees incrementally. Every new tree makes up for the shortcomings of its ancestors. Gradient Boosting can be applied to many regression and classification tasks across various loss functions. The model's hyperparameters including the number of trees, depth, and learning rate provide fine-tuning control. Improper calibration increases the risk of overfitting. Because of its sequential nature and adaptability, Gradient Boosting is often superior to other algorithms in terms of prediction accuracy especially when handling complicated relationships between features-target variables. XGBoost, LightGBM, and Catboost are popular frameworks that use Gradient Boosting leading to various optimizations as well as capabilities.

4.5.4.2 XGBoost

XGBoost is a machine learning technique that combines multiple models using the gradient-boosting decision tree approach. 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 improves accuracy by utilizing a differentiable loss function called mean squared error and an optimization process called gradient descent. The procedure follows the first and second-order terms of the Taylor's series expansions. To control the complexity of the enlarged term, a regularization term is included. The regularization term regulates the over-fitting phenomenon and contributes to the equalization of the final learning weight w. XGBoost utilizes fitting techniques to identify and forecast the regular term to mitigate over-fitting by employing sampling methods for rows and columns. The approach is capable of simultaneously utilizing parallel processing (dividing nodes) and distributed computing (multi-threading), hence demonstrating the feasibility of the quickest model (Prakash et al., 2023).

4.5.4.3 LightGBM Boost

LightGBM, a machine learning model, has been used in different applications such as the algorithm to build construction steel bar size detection system which improves the accuracy of steel bar quality testing (Tian, M. 2022), and also used air quality predicting and personalized head-related transfer function (HRTF) predicting. LightGBM achieved excellent prediction performance, high stability as well as good generalization in these various domains across all the test fold groups. LightGBM showed better performance in testing accuracy, good general correlation between predicted ground-level particulate matter than other machine learning algorithms in air quality prediction. With regard to sailboat pricing used, it is notable that mathematical model establishment using the LightGBM regression model is applied to provide the accurate prediction of prices. For the use of LightGBM in HRTF prediction in this paper, it explicitly avoids the overfitting phenomenon and achieves a perfect predictive performance, as the fitting errors are seemingly considerably lower than those provided by all other methods. As far as the health estimation for lithium-ion batteries is concerned, the overall performance of the LightGBM model is excellent and it can strongly generalize (Qiu et al., 2023).

4.5.4.4 CatBoost

CatBoost is a gradient tree boosting method. CatBoost algorithm was developed by Yandex engineers and researchers, which is the largest Russian firm of search engines in 2017. Its source code has been published for public access in April. At present, this algorithm is regarded as the most advanced in the world by the indicators of efficiency and speed of the XGBoost and LightGBM algorithms in global open source. The name "CatBoost" combines the words "Category" and "Boosting," and it belongs to the Boosting algorithms family. The CatBoost algorithm represents an enhancement within the context of the GBDT algorithm. It efficiently addresses the problem of gradient bias and prediction shift, avoids the chance of overfitting, and enhances the precision of calculations and the ability to generalize (Yang et al., 2023).

4.5.4.5 Adaboost

Adaboost is a machine learning algorithm that combines multiple weak classifiers to create a strong classifier. AdaBoost is an ensemble machine learning methodology that constructs a strong learning algorithm by combining weak learning methods. A decision tree with only one degree of split is commonly referred to as a stump. The inherent model assigns equal weights to each of the data points. Enhancing the fundamental principles of algorithms allows for the fine-tuning of training weights to align with the dataset. Meanwhile, the training will continue until the smallest inaccuracy is observed. During the training phase, AdaBoost constructs weak classifiers with low accuracy that gradually improve on their predecessors. To get higher accuracy, it is necessary to accurately tweak the hyperparameters of AdaBoost. AdaBoost is less prone to overfitting since it requires less pre-processing in the dataset estimation process (Prakash et al., 2023).

4.4 Random forest

Random 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 et al., 2023).

4.5 Optimization of machine learning models

Optimization of machine learning models is important, dynamic for the performance and the flexibility of the models in complicated environments. Optimization needs a deep knowledge of data and algorithms. Proper data preparation is crucial. Optimization requires missing data, feature normal ionization categorical variable encoding, and feature engineering. Feature quality and relevance help determine model performance. Choice of an approach based on the problem type – classification, regression, etc., and data attributes. Different models are strong or weak in different contexts. Random Forest and Gradient Boosting can outperform individual models by combining them. Optimizing the model is common for several repetitions. Testing, validation, and improvement cannot stop.

CHAPTER V

Results and discussions

5.1 Results

5.1.1 Overview

This part summarizes the findings to ascertain the efficacy and comprehensibility of five sophisticated machine learning models: Gradient Boost, XGBOOST, Adaboost, LGBM Boost, and Catboost in forecasting the characteristics of asphalt mixtures. It also depends on three fundamental interpretative parameters: the Shapley Value, Permutation Importance, and Accumulated Local Effects (ALE), among others to understand the impact of various aspects on model predictions. The model outcomes are situated on two characteristics of asphalt mixtures: Elastic Modulus (E*) and Dynamic Modulus (Log E*). The results are elaborated in detail through graphs and tables that lead to the perceptiveness of the predicted accuracy and the influence of the features of the model. This offers a full examination of how these models might be applied to predict the behaviour of asphalt mixtures.

5.1.2 The performance comparison for five models

Table 5.1

Estimation performance comparison for five models: test and train data on asphalt mixtures (E)*

Models	Train				Test			
	R^2	RMSE	MAE	MSE	R^2	RMSE	MAE	MSE
GB	0.925	4.30E+05	2.84E+05	1.85E+11	0.91	4.58E+05	2.99E+05	2.10E+11
Adaboost	0.841	7.17E+05	5.91E+05	5.13E+11	0.83	7.25E+05	5.95E+05	5.26E+11
xboost	0.998	6.55E+04	4.50E+04	4.29E+09	0.98	2.20E+05	1.37E+05	4.83E+10
LGBM boost	0.987	1.82E+05	1.18E+05	3.30E+10	0.97	2.78E+05	1.68E+05	7.71E+10
catboost	0.993	1.35E+05	9.07E+04	1.81E+10	0.98	2.12E+05	1.33E+05	4.50E+10

Table 5.2

Estimation performance comparison of five models: test and train data on dynamic modulus of asphalt mixtures (LOG E)*

Models	Train				Test			
	R^2	RMSE	MAE	MSE	R^2	RMSE	MAE	MSE
GB	0.971	1.17E-01	9.14E-02	1.40E-02	0.968	1.23E-01	9.52E-02	1.50E-02
Adaboost	0.926	1.91E-01	1.55E-01	3.66E-02	0.926	1.92E-01	1.54E-01	3.69E-02
xboost	0.998	2.72E-02	1.98E-02	1.00E-03	0.989	7.28E-02	5.31E-02	5.00E-03
LGBM boost	0.993	5.78E-02	4.46E-02	3.34E-03	0.987	7.82E-02	5.92E-02	6.11E-03
catboost	0.996	4.29E-02	3.23E-02	2.00E-03	0.992	6.12E-02	4.54E-02	4.00E-03

Table 5.1, as well as Table 5.2, presents five models' performance comparisons on train and test data for the elastic modulus E* and dynamic modulus of asphalt mixtures Log E*. The models used here include Gradient Boosting (GB), Adaboost, XGBoost, Light Gradient Boosting Machine (LGBM boost), and CatBoost.

Performance metrics to be used for comparing the models would consist of:

R^2 (R-squared): It shows how much of the variation is explained by the dependent variable when there is a prediction that is made from the independent variable. A value close to 1 or 100% is better for the output by the model. RMSE (Root Mean Square Error): It averages the squares of the errors. MAE (Mean Absolute Error): This will calculate an average of the absolute errors. MSE (Mean Squared Error): Measures the average of the squares of the errors. For all of the RMSE, MAE, AND MASE Lower values are better.

5.1.2.1 ADABOOST

Figure 5.1

Comparing measured and predicted E(psi) using AdaBoost for (a) test and (b) train data.*

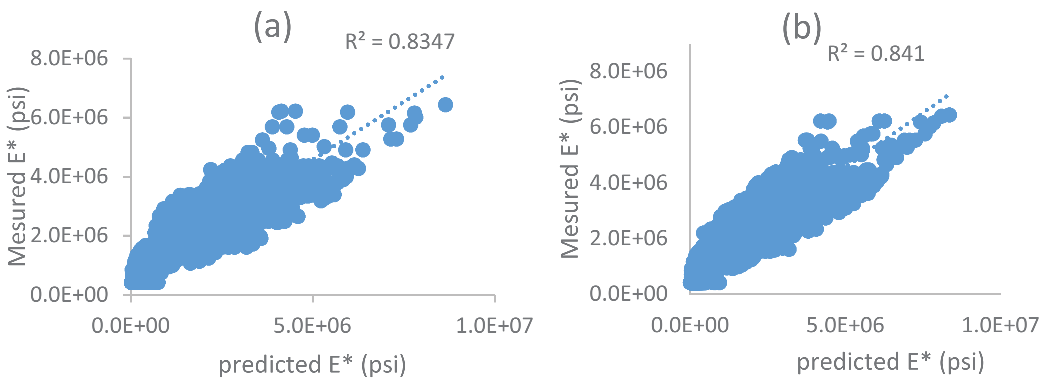
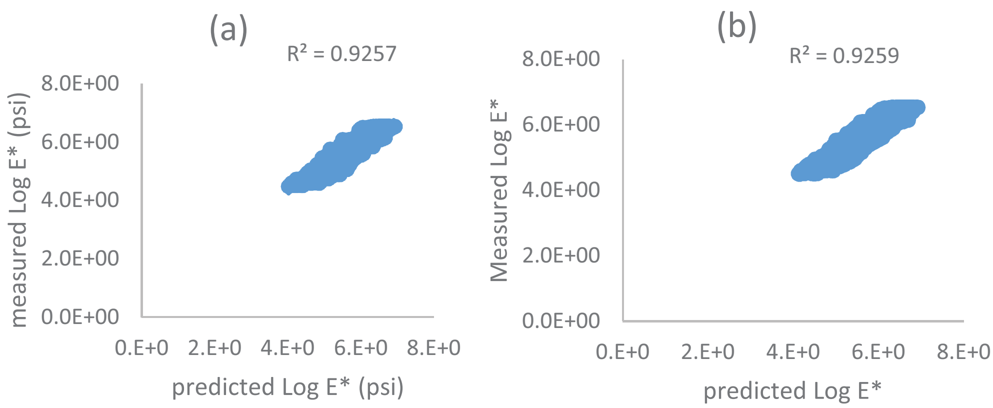


Figure 5.2

Comparing measured and predicted Log E(psi) using AdaBoost for (a) test and (b) train data.*



5.1.2.2 Gradient Boost

Figure 5.3

Comparing measured and predicted E(psi) using Gradient boost for (a) test and (b) train data.*

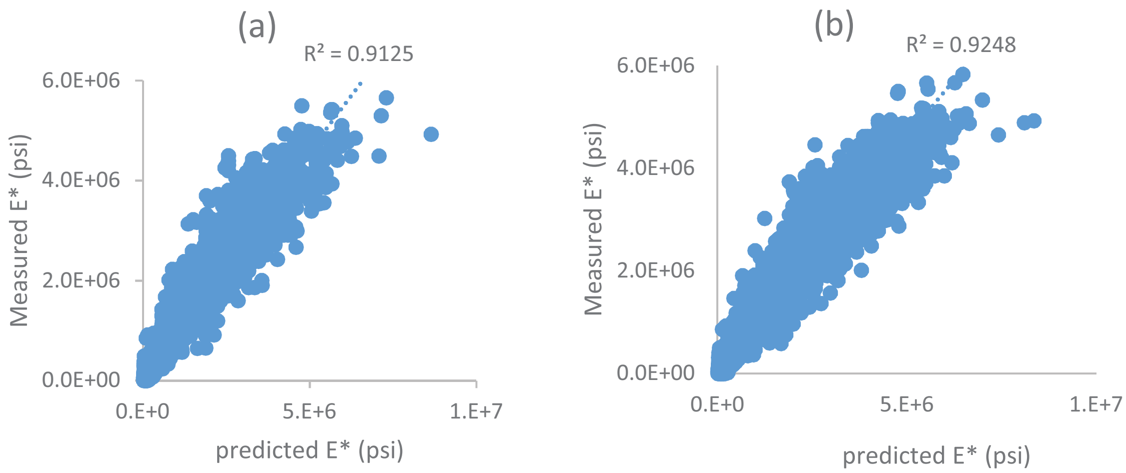
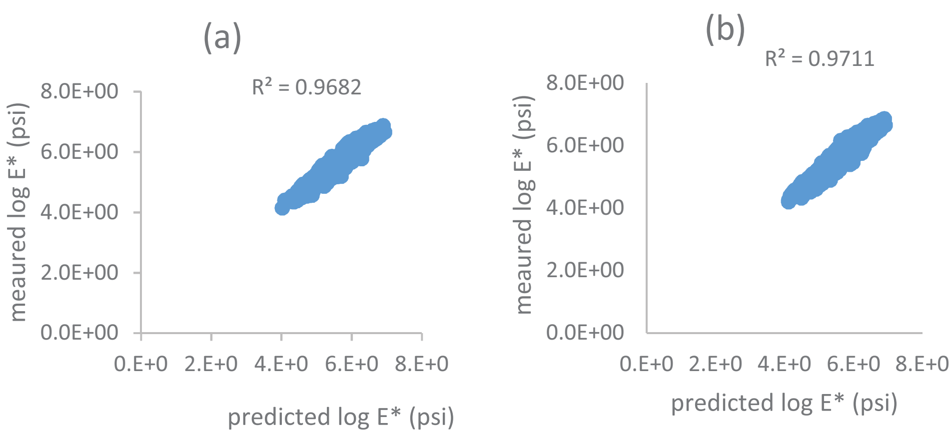


Figure 5.4

Comparing measured and predicted Log E(psi) using Gradient boost for (a) test and (b) train data.*



5.1.2.3 LightGBM Boost

Figure 5.5

Comparing measured and predicted E(psi) using LightGBM Boost for (a) test and (b) train data.*

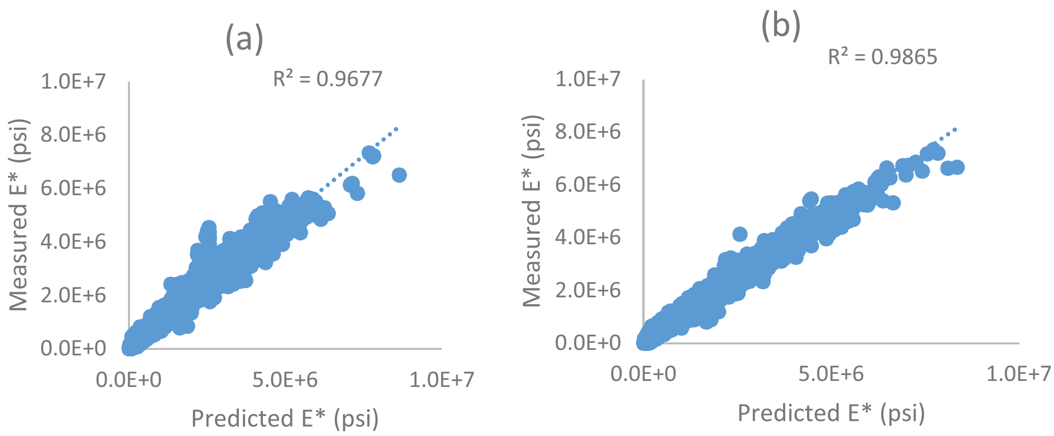
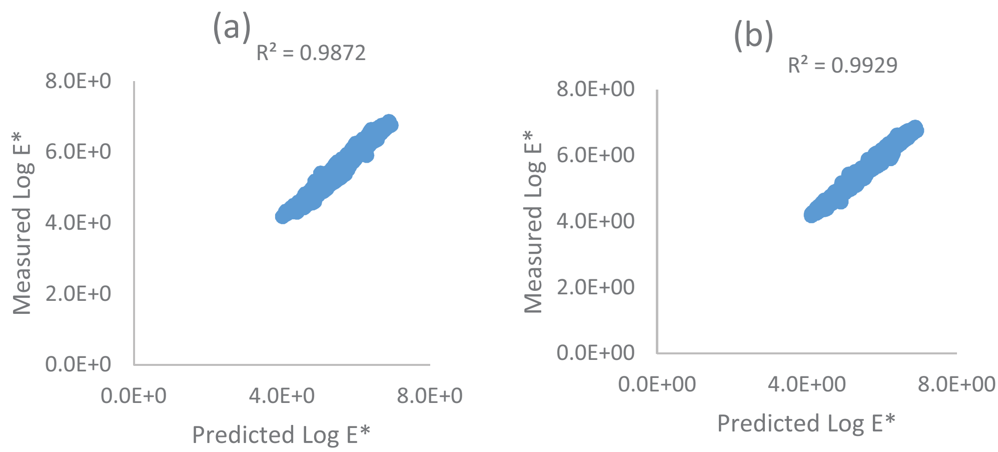


Figure 5.6

Comparing measured and predicted Log E(psi) using LightGBM Boost for (a) test and (b) train data.*



5.1.2.4 CatBoost

Figure 5.7

Comparing measured and predicted E(psi) using CatBoost for (a) test and (b) train data.*

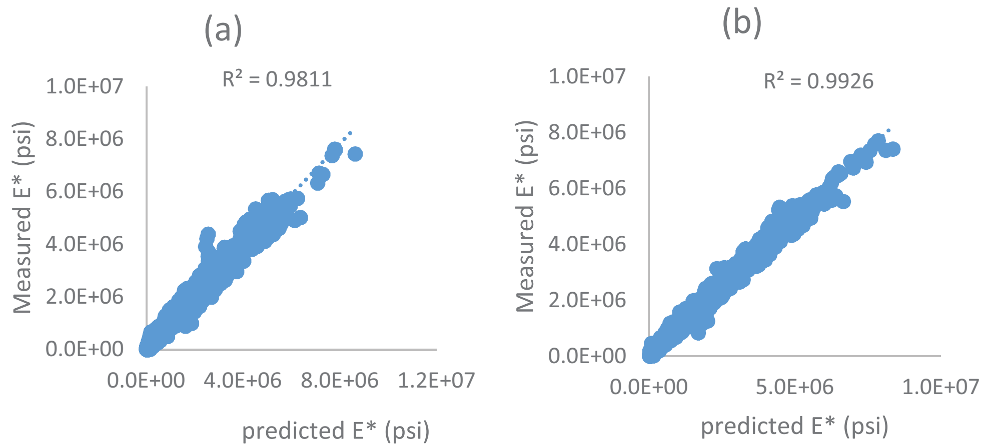
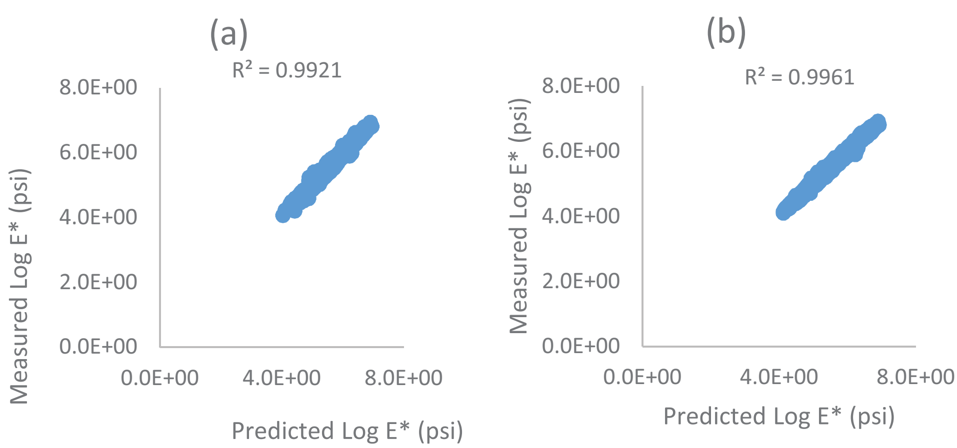


Figure 5.8

Comparing measured and predicted Log E(psi) using CatBoost for (a) test and (b) train data*



5.1.2.5 XGBoost

Figure 5.9

Comparing measured and predicted E(psi) using XGBoost for (a) test and (b) train data*

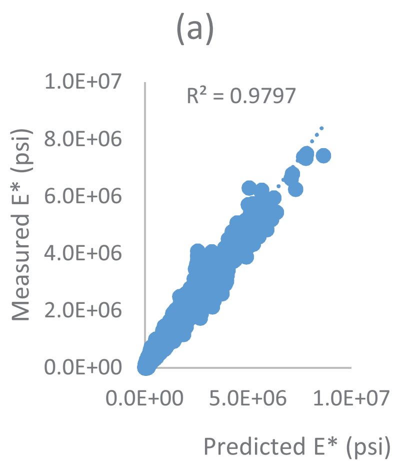
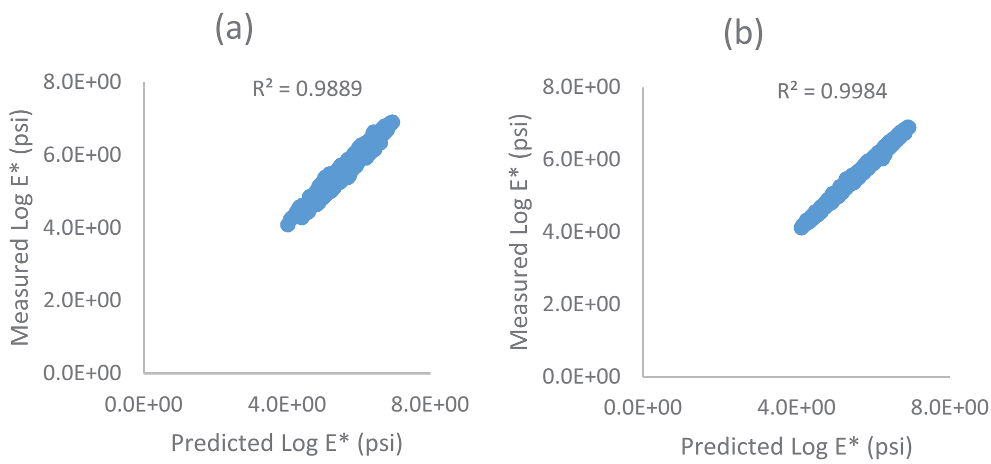


Figure 5.10

Comparing measured and predicted Log E(psi) using xgBoost for (a) test and (b) train data*



5.1.3 Using parameter interpretability (shapley value, permutation importance, and accumulated local effects) finding E* and Log E*.

Table 5.3

*Ranking of three different parameters on data E**

Feature	**SHapley value**	**Permutation importance**	**Accumulated local effects(ALE)**
|G*|	1	1	1
T	2	3	4
δb	3	2	2
Va	4	4	3
r34	5	5	10
r38	6	7	7
Vbeff	7	6	5
fc	8	12	13
Ac	9	8	6
r200	10	11	9
r4	11	9	11
VTS	12	13	8
A	13	10	12

Table 5.4

*Ranking of three different parameters on data Log E**

Feature	**SHapley value**	**Permutation importance**	**Accumulated local effects(ALE)**
|G*|	1	1	1
T	2	2	4
Va	3	3	3
δb	4	4	2
r34	5	8	6
Vbeff	6	5	5
r38	7	6	7
Ac	8	7	8
r200	9	10	9
VTS	10	13	10
fc	11	11	12
r4	12	12	13
A	13	9	11

The feature rankings for the prediction of asphalt mixture properties of E* and Log E* from the interpretability methods: Shapley value, permutation importance, and accumulated local effects (ALE) can be seen in Tables 5.3 and 5.4. Each feature (e.g., |G*|, T, db) is ranked across those methods; this comparison of significance amongst the features is imperative for model refinement, understanding of prediction drivers, and practical applications in asphalt mixture design.

Figure 5.11

Revealing the feature values of SHapley value of E in asphalt mixture*

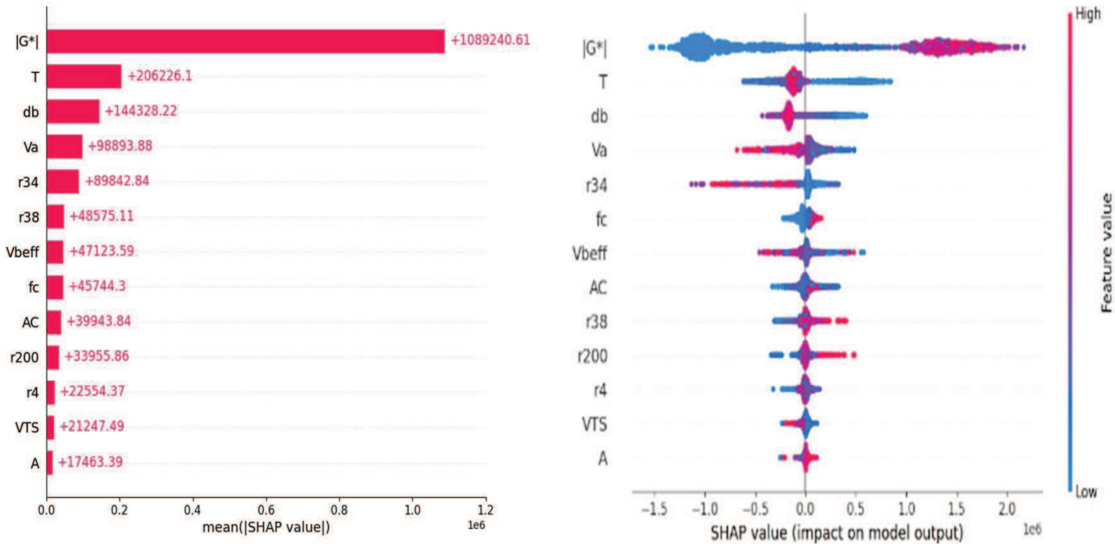


Figure 5.12

Revealing the feature values of permutation importance of E in asphalt mixture*

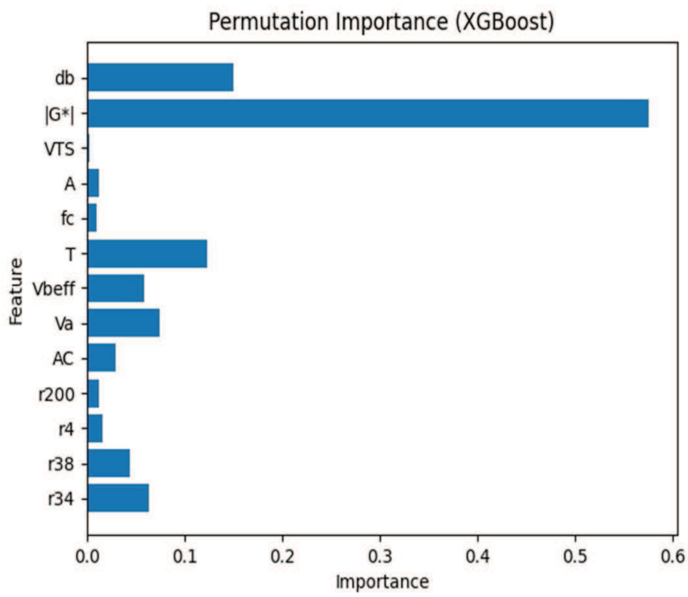
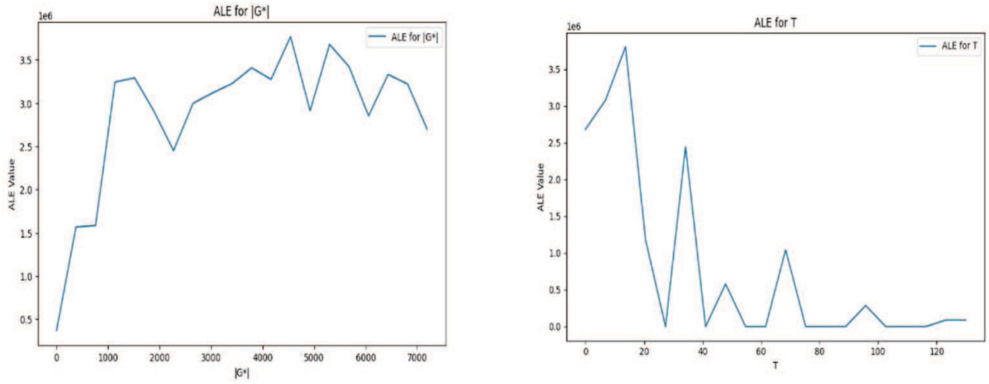


Figure 5.13

Revealing the most important two feature values of Accumulated local Effects (ALE) of E in asphalt mixture.*



(a) (b)

Figure 5.14

Revealing the feature values of SHapley value of Log E in asphalt mixture*

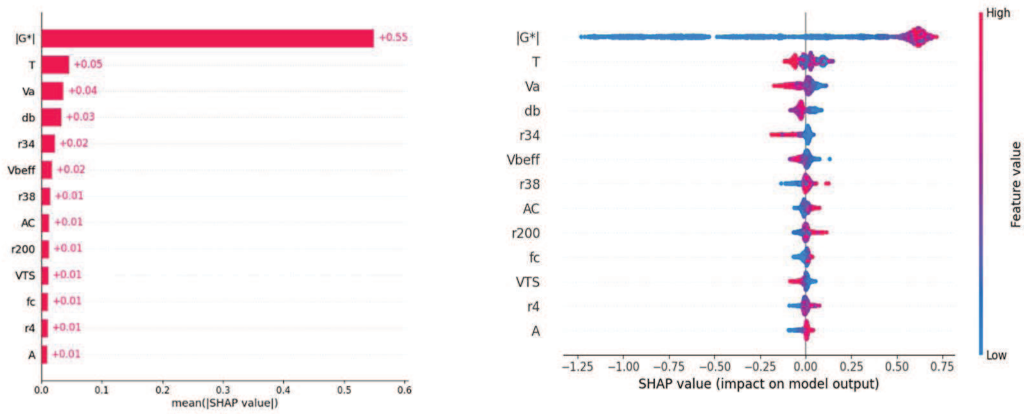


Figure 5.15

Revealing the feature values of permutation importance of Log E in asphalt mixture*

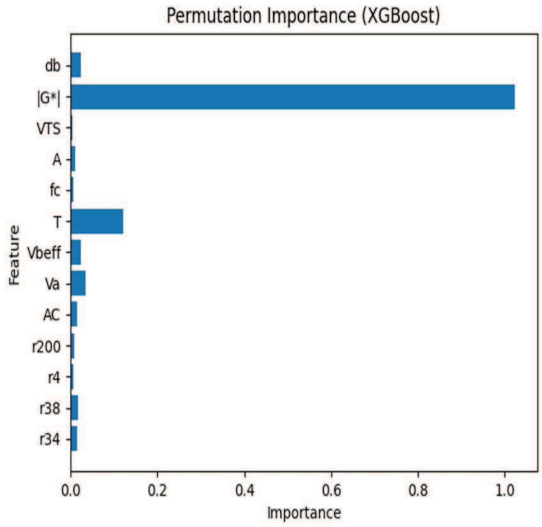
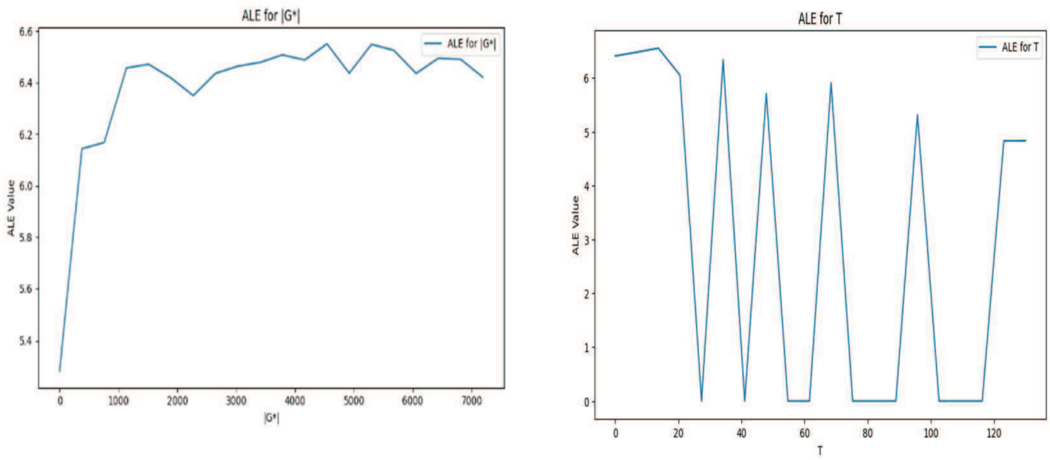


Figure 5.16

*Revealing the most important two feature values of Log E*Accumulated local Effects (ALE) of E* in asphalt mixture*



5.2 Discussion

5.2.1 Discussing the performance Comparison for five models

Table 5.1 and Table 5.2 show the comparative results of five machine learning models which are Gradient Boosting (GB), Adaboost, XGBoost, Light Gradient Boosting Machine (LGBM boost) and CatBoost respectively on the estimation of asphalt mixtures properties (E*) and the pavement material dynamic modulus (Log E*). The models are evaluated on four key metrics, both for train dataset and test dataset: R2 (R squared), RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MSE (Mean Squared Error).

In Table 5.1, Performance on Asphalt Mixtures (E*), XGBoost performs better than other models with the highest R^2 values (0.998 train, 0.98 test) indicating that this model has an exceptional accuracy and good prediction capability. It also boasts the least error metrics (RMSE, MAE, MSE) in both train and test data, pointing to its accuracy of estimating E*. The Adaboost contrasts this by recording the lowest R2 values (0.841 train, 0.83 test). It indicates lesser accuracy in explaining the variance of E*. Its error metrics are also the highest among the models, indicating less precise predictions.

Whereas on the contrary, XGBoost again comes out to be the best-performing model while predicting Log E*, shown as listing of Table 5.2. Performance on Dynamic Modulus of Asphalt Mixtures (Log E*), which comprises the highest R2 values (0.998 Train, 0.989 Test) and thus the least metrics of error, thus establishing its robustness and accuracy in predicting Log E*. Adaboost has the lowest R2 values (0.926 both train and test) along with higher errors metrics as compared to XGBoost but does better in estimating Log E* than E* proportionately.

The RMSE, MAE, and MSE values across all models are consistently smaller in Log E* than in E*, thereby suggesting that the models in general predict a more accurate result of logarithm dynamic modulus. This could be contributed to the fact that logarithmic transformation is known to stabilize variance, hence enhancing pattern recognition ability of machine learning algorithms.

In a nutshell, XGBoost performs better in predicting E* and also for Log E* in asphalt mixes than other models because of its highest accuracy and reliability of predicting. Hence, it has proved always to provide high R2 value and low error metrics thereby showing its capability to grapple with the entire complexity found within the data. On the other hand, though useful, Adaboost shows comparatively lower performance especially in predicting E*. The insights from these evaluations are paramount in selecting the most appropriate models for accurate predictions useful in the analysis and design for asphalt mixture.

5.2.2 Discussing the parameter interpretability (shapley value, permutation importance, and accumulated local effects) finding E* and log E*.

The rank of the importance of different features in predicting dynamic modulus E* and its logarithm Log E* of asphalt mixtures using three interpretability methods: Shapley value, permutation importance, and accumulated local effects (ALE) is provided in Tables 5.3 and 5.4. In each method, every feature is ranked in which 1 represents the most significant influence towards the predictions.

From Table 5.3 Ranking for E*, across the three methods, the most influential feature is |G*| which all the methods gave it a rank of 1. That is, |G*| has the highest driver in predicting dynamic modulus (E*), with the highest effect on model output. Looking at the average ranks, it is observed that 'A' seems to be the least impactful, with ranks 13, 10, and 12 across the three methods.

In Table 5.4, As with E*, |G*|was again considered the first feature for predictive ability on Log E*, where it held a top rank across all interpretability methods. The Log E* Ranking Clearly, |G*| has strong and consistent predictive power for both the dynamic modulus and its logarithm. The feature 'r4' appeared to be of the least influence for Log E* with ranks 12, 12, and 13 respectively across methods indicating low information contents of this feature and inclusion in the model provides little additional information.

CHAPTER VI

Conclusion and recommendation

6.1 Conclusion

Through the machine learning interpretability parameters Shapley value, Permutation importance, and Accumulated local effects (ALE), and machine learning predict models Gradient boost, XGBOOST, Adaboost, LGBM boost and Catboost approaches with advanced machine learning methods, the assessment of Elastic modulus (E*) and LOG E* dynamic modulus of asphalt mixtures has been considered excellent acumen to enhancing the accuracy and effectiveness of the pavement engineering assessments. These approaches have used different datasets and presented varied views regarding the influencing aspects of the behavior of asphalt mixture.

Common findings established include that ensemble methods such as Adaboost e gradient boost, XGBOOST, LGBM boost, and Catboost were effective when handling complex non-linear relations within the parameters of a dataset. The use of Shapley value and Permutation importance brought more light on the contributions made by features as well as feature importance on the model predictions adding up to the interpretations. Accumulated local effects came in handy due to their ability to be able to capture the impacts borne out of potential interaction between features at different scales or ranges.

However, computational intensity and the requirement of substantially pre-processed data were also the challenges witnessed. The reasons for variation in performance across different datasets and methods indicate that model selection and tuning are context-dependent.

6.2 Recommendations

Model Selection and Tuning: Sensibly choosing and tuning one or more machine learning models according to inherent characteristics of the dataset at hand as well as practical requirements of pavement engineering tasks, is highly needful.

Interpretability Focus: Owing to the involved nature of asphalt mixes, in engineering judgments, it can be hypothesized that some models that offer a proper interpretation as shown by Shapley value and Permutation importance, for instance, will be favored.

Data Quality and Diversity: Ensuring diverse, high quality datasets to train and validate models covering the full variety of asphalt mixture types and environmental conditions for more generalized outcomes.

Resource Allocation: Take note of computational demands the said techniques, in particular for ensemble methods and those pertaining to Shapley value calculations. Allocate necessary resources for such or if possible provide appropriate adjustments to go along with efficient model training and analysis.

Integration with Traditional Methods: Integrating machine learning evaluations along with traditional evaluative approaches to get better insight into asphalt mixture.

Continued Research and Development: Efforts should be towards triggering continuous research in further improving these machine learning models in terms of efficiency and accuracy, and also on how techniques newly matured in the area can be put to use.

References

Au, Q., Herbinger, J., Stachl, C., Bischl, B., & Casalicchio, G. (2022). Grouped feature importance and combined features effect plot. *Data Mining and Knowledge Discovery*, *36*(4), 1401–1450. https://doi.org/10.1007/s10618-022-00840-5

Kaloush, K., & Witczak, M. (2003). Laboratory evaluation of asphalt rubber mixtures using the dynamic modulus (E*) test. *Annual Transportation Research Board*. http://www.ltrc.lsu.edu/TRB_82/TRB2003-000927.pdf

Kim, J., & Jung, D. (2024). *State-of-the-Art Review of Machine Learning Models in Civil Engineering : Based on DAMIE Classification Tree*. *3*.

Lontra, B. M. (n.d.). *Machine learning study on performance prediction of asphalt mixtures under Dutch conditions*.

Xu, W., Huang, X., Yang, Z., Zhou, M., & Huang, J. (2022). Developing Hybrid

Machine Learning Models to Determine the Dynamic Modulus (E*) of Asphalt Mixtures Using Parameters in Witczak 1-40D Model: A Comparative Study. *Materials*, *15*(5). https://doi.org/10.3390/ma15051791

Zaumanis, M., Poulikakos, L. D., & Partl, M. N. (2018). Performance-based design of asphalt mixtures and review of key parameters. *Materials and Design*, *141*, 185–201. https://doi.org/10.1016/j.matdes.2017.12.035

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

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

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.

Lema, L. (2018). *Parameters interpretability in phenomenological-based semiphysical models. A human glucose homeostasis model*. 174. http://bdigital.unal.edu.co/73186/

Li, C., Li, M., & Qiu, Z. (2022). Post-Hoc Interpretability Based Parameter Selection for Data Oriented Nuclear Reactor Accident Diagnosis System. *International Conference on Nuclear Engineering, Proceedings, ICONE*, *7-B*. https://doi.org/10.1115/ICONE29-92505

Li, S., Sun, Y. Y., & Chen, C. H. (2023). An Interpretable Machine Learning Procedure Which Unravels Hidden Interplanetary Drivers of the Low Latitude Dayside Magnetopause. *Space Weather*, *21*(3), 1–12. https://doi.org/10.1029/2022SW003391

Prakash, A., Thangaraj, J., Roy, S., Srivastav, S., & Mishra, J. K. (2023). Model-Aware XGBoost Method Towards Optimum Performance of Flexible Distributed Raman Amplifier. *IEEE Photonics Journal*, *15*(4). https://doi.org/10.1109/JPHOT.2023.3286272

Qiu, Y., Wang, J., & Li, Z. (2023). Personalized HRTF prediction based on LightGBM using anthropometric data. *China Communications*, *20*(6), 166–177. https://doi.org/10.23919/JCC.2023.00.025

Yang, H., Liu, Z., Li, Y., Wei, H., & Huang, N. (2023). CatBoost–Bayesian Hybrid Model Adaptively Coupled with Modified Theoretical Equations for Estimating the Undrained Shear Strength of Clay. *Applied Sciences (Switzerland)*, *13*(9). https://doi.org/10.3390/app13095418

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.

Appendix

Appendix A Ethics certificate

NEAR EAST UNIVERSITY

SCIENTIFIC RESEARCH ETHICS COMMITTEE

29.01.2024

Dear Abdullahı_Mohamud Adam

Your project **"A Critical Assessment Of Machine Learning Applications In Civil Engineering"** 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