

**AN INTEGRATED SMART PLATFORM FOR
ESTIMATING ONSITE PRODUCTIVITY FOR
HEAVY CONSTRUCTION PROJECTS
IN SAUDI ARABIA**

BY

HABIBULLAH RAHNAWARD

A Thesis Presented to the
**COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

CONSTRUCTION ENGINEERING AND MANAGEMENT

December 2024

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

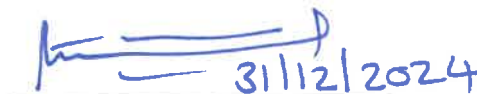
DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by HABIBULLAH RAHNAWARD under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN CONSTRUCTION ENGINEERING AND MANAGEMENT.



Dr. Laith Al Hadidi
Department Chairman


31/12/2024

Dr. Awsan Mohammed
(Advisor)



Dr. Sulaiman Saleh Al-Homidan
Dean of Graduate Studies



Dr. Adel Alshibani
(Member)

16/ 12/ 2024

Date



Dr. Ahmed Al Ghaithan
(Member)



© HABIBULLAH RAHNAWARD

2024

Dedication

I thank Allah deeply for the many blessings in my life. To my father, mother, brothers, sisters, friends, and colleagues, I am so grateful for your love, support, and encouragement. Your belief in me made this journey possible.

ACKNOWLEDGMENTS

With sincere gratitude, I acknowledge the support and encouragement I've received while working on my thesis, "An Integrated Smart Platform for Estimating Onsite Productivity for Heavy Construction Projects in Saudi Arabia." This journey has been one of significant growth and learning, made possible by many.

Foremost, I express my profound appreciation to my advisor, Doctor Awsan Mohammed, whose guidance was invaluable. Your expertise, patience, and encouragement were crucial in navigating the complexities of this project. Thank you for believing in my potential and pushing me to excel.

To the committee members Dr. Adel Alshibani and Dr. Ahmed Ghaithan and our kind chairman Dr. Laith Hadidi and staff of AECM at KFUPM, thank you for the inspiring environment you have provided and for the opportunities to expand my views. Your dedication to fostering learning and innovation has left a lasting impact on me.

I also extend my gratitude to the professionals and participants from the construction industry whose insights and contributions were pivotal to my research. Your willingness to share your experiences has greatly enriched my work. This thesis reflects not just my efforts, but the collective support, encouragement, and belief of everyone who has been part of this journey. To all of you, I offer my sincerest thanks. Your roles in this journey have been invaluable, and I am deeply grateful.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	V
TABLE OF CONTENTS	VI
LIST OF TABLES	XI
LIST OF FIGURES	XII
LIST OF ABBREVIATIONS	XIV
ABSTRACT.....	XV
ملخص الرسالة.....	XVII
CHAPTER 1	1
INTRODUCTION.....	1
1.1 Study Background	1
1.2 Artificial Neural Network ANN.....	2
1.2.1 An Artificial Neural Network's Main Components (ANN).....	4
1.2.2 Artificial neural networks' most significant subtypes (ANNs).....	7
1.3 Research Gaps.....	10
1.4 Statement of the Problem	11
1.5 Objective of the study	13
1.6 Significant of the study	13
1.7 Scope and Limitations	15
1.8 Thesis Organization	16
1.9 Expected Outcomes:	18
1.10 Chapter Summary:	19

CHAPTER 2	22
LITERATURE REVIEW	22
2.1 Review of Existing Studies	22
2.2 Bibliometric Analysis.....	23
2.3 Concept of Onsite Productivity:	24
2.4 Construction Industry in Saudi Arabia:	25
2.5 Sensor Data Integration and Analysis.....	27
2.6 Productivity Estimation Models	27
2.7 Key Factors Influencing Construction Equipment Productivity	31
2.7.1 Factors Details	32
2.8 Integration of Technology in Construction:	36
2.9 Challenges and Gaps in the Literature:	38
2.10 Theoretical Framework:	40
2.11 Chapter Summary:	41
CHAPTER 3	43
RESEARCH METHODOLOGY	43
3.1 Introduction.....	43
3.2 Research Design	43
3.3 Identifying Key Productivity Factors from the Literature	46
3.4 Data Collection	49
3.4.1 Tracking Bulldozer Movement with GPS.....	49
3.4.2 Site Condition and Operational Data Recording	51
3.4.3 Integration of GPS and Site Condition Data.....	54

3.5 Data Extraction and Preparation	55
3.5.1 GPS Data Transformation and Path Extraction	55
3.5.2 Data Cleaning and Outlier Removal	59
3.5.3 Data Analysis	60
3.5.4 Feature Engineering and Final Dataset Preparation.....	62
3.6 Model Development and Validation.....	63
3.6.1 Artificial Neural Network (ANN) Model	64
3.6.2 Support Vector Machine (SVM) Model	66
3.6.3 Random Forest (RF) Model	67
3.6.4 Model Evaluation Metrics.....	68
3.6.5 Web-Based Platform Development	69
3.7 Chapter Summary	71
CHAPTER 4	72
RESULTS AND DISCUSSION	72
4.1 Introduction.....	72
4.2 Survey of Industry Professionals.....	73
4.3 Validation of Productivity Factors with Project Experts.....	75
4.4 Data Cleaning and Outlier Removal	78
4.5 Data Analysis.....	80
4.5.1 Normalization Boxplots:.....	80
4.5.2 Pairplot of Features:.....	81
4.5.3 Correlation Heatmap:.....	82

4.6 Models Performance	84
4.6.1 The ANN Model	85
4.6.2 The ANN model Parameters:	86
4.6.3 The ANN Architecture	89
4.6.4 The ANN Model Performance Evaluation	91
4.6.5 The SVM Model	93
4.6.6 The SVM model Parameters	93
4.6.7 The SVM Model Representation	94
4.6.8 The SVM Model Performance Evaluation	95
4.4.9 The RF Model	97
4.6.10 The RF model Parameters	98
4.6.11 The RF Model Representation	99
4.6.12 The RF Model Performance Evaluation	100
4.7 Models Comparison Summary	102
4.8 Validation of the ANN Model	105
4.8.1 K-Fold Cross-Validation Results	106
4.8.2 Validation Using Random Sample Data	107
4.9 Development of the Web-Based Productivity Prediction Platform	109
4.10 Discussion	113
4.10.1 Comparative Analysis of Model Performance	114
4.10.2 Practical Implications for Construction Management	115
4.10.3 Alignment with Existing Research and Industry Trends	116

4.11 Chapter Summary	118
CHAPTER 5.....	121
CONCLUSION	121
5.1 CONCLUSION	121
REFERENCES:	125
VITAE.....	136

LIST OF TABLES

Table 2.1: Factors Affecting On The Productivity Of Bulldozers.....	36
Table 3.1: Key Productivity Factors Identified From The Literature	48
Table 3.2: Site Condition And Operational Data Recording For Bulldozer Operations ..	53
Table 4.1: Survey Results RII For Each Factor	75
Table 4.2: Description And Labels Of Dataset Features	84
Table 4.3: The ANN Model Prameters	88
Table 4.4: Summary Of The Prameters	88
Table 4.5: Hyperparameters And Descriptions For The Random Forest Model	99
Table 4.6: Performance Comparison Of SVM, RF, And ANN Models	103
Table 4.7: Cross-Validation Result For The ANN Model.....	107

LIST OF FIGURES

Figure 1.1: Architecture Of The Artificial Neural Network (ANN) Model	5
Figure1.2: Overview Of Major Types Of Artificial Neural Networks (ANNs)	8
Figure 2.1: Co-Occurrence Network.....	23
Figure 3.1: Framework For Developing A Predictive Model.....	45
Figure 3.2: Publication Trend By Year (1950-2024).....	46
Figure 3.3: Aerial View Of Construction Project Locations In Saudi Arabia	49
Figure 3.4: Gps Tracking Screenshots, Bulldozer Movement Paths With Color-Coded.	50
Figure 3.5: Gps Traked Points	55
Figure 3.6: Gps Paths Segmented By Turning Points	58
Figure 4.1: Bulldozer,Qatif,Saudi Arabia.A.Blade Interaction,B.Sandy-Clay Mix.	77
Figure 4.2: D155a Bulldozer, (A) Blade Interaction, (B) Sandy Soil Texture.	77
Figure 4.3: Box Plots of Before And After Outlier Removal	79
Figure 4.4: Normalized And Unnormalized Boxplots.....	80
Figure 4.5: Pairplot of Key Variables	82
Figure 4.6: Heatmap of Feature Correlations	83
Figure 4.7: Architecture of The ANN Model	91
Figure 4.8: Training Set Results of The ANN	93
Figure 4.9: Testing Set Results of The ANN	93
Figure 4.10: Conceptual Representation of SVM.....	95
Figure 4.11: Training Set Results of The SVM	97

Figure 4.12: Testing Set Results of The SVM.....	97
Figure 4.13: Random Forest Model Architecture	100
Figure 4.14: Training Set Results of The RF	102
Figure 4.15: Testing Set Results of The RF.....	102
Figure 4.16: Comparison of Observed And Predicted Productivity	105
Figure 4.17: Validation Set Results for The ANN Model	108
Figure 4.18: A Visual Representation of The Platform Workflow.....	111
Figure 4.19: The Platform’s Upload Interface	112
Figure 4.20: The Platform’s Result Interface	113

LIST OF ABBREVIATIONS

ANN:	Artificial Neural Network
AI:	Artificial Intelligence
BIM:	Building Information Modeling
BUF:	Blade Utilization Factor
CE:	Construction Equipment
DPF:	Distance Penalty Factor
MAE:	Mean Absolute Error
MSE:	Mean Squared Error
RF:	Random Forest
RBF:	Radial Basis Function
RII:	Relative Importance Index
RMSE:	Root Mean Squared Error
ReLU:	Rectified Linear Unit
SVM:	Support Vector Machine
VRF:	Volume Reduction Factor

ABSTRACT

Full Name : Habibullah RAHNAWARD

Thesis Title : An Integrated Smart Platform for Estimating Onsite Productivity for
Heavy Construction Projects in Saudi Arabia

Major Field : Construction Engineering & Management

Date of Degree : 2024/12/20

The productivity of heavy construction is crucial in construction for managing and controlling all aspects of the earthwork operations. However, the traditional techniques are ill for accurate and automated estimating of the equipment productivity. Consequently, the primary objective of this study is to develop an integrated smart platform for accurately estimating productivity in heavy construction projects, particularly bulldozer operations. Field data collected from Saudi Arabian construction sites were analyzed, incorporating over 20 critical factors such as cycle time, operator skill, weather conditions, and soil type. A set of sensors and smart tools are used to collect data. The data are prepared, cleaned, and flittered to eliminate the noise and enhance the quality. The research proposes a machine learning algorithm based on the artificial neural network to estimate the productivity of heavy equipment. In addition, the performance of the proposed algorithm is compared with the performance of other two machine learning algorithms: support vector machine and random forest. Different evaluation measures are used to evaluate the performance of the proposed machine learning algorithms. These measures include R-square, mean absolute error, and root mean square error. Moreover, the proposed

algorithm parameters are tuned optimally using an exhaustive algorithm. A smart web-based platform that integrates the ANN model, providing real-time productivity estimates under diverse conditions is developed. The findings indicated that the proposed artificial neural network is capable of estimating the productivity of heavy equipment with the accuracy of an R^2 value of 0.98 and low error metrics (RMSE: 14.59, MAE: 10.37), outperforming support vector machines and random forest models. The results also revealed that the proposed platform has the ability to transform complex data into actionable insights. This smart platform will enable construction managers to reduce costs, accelerate timelines, and improve operational efficiency, setting a new benchmark for productivity management in the construction industry.

ملخص الرسالة

الاسم الكامل: حبيب الله رهنورد

عنوان الرسالة: منصة ذكية متكاملة لتقدير الإنتاجية في الموقع لمشاريع البناء الضخمة في المملكة العربية السعودية

التخصص: قسم الهندسة المعمارية وإدارة الإنشاءات

تاريخ الدرجة العلمية: 2024

إنتاجية مشاريع البناء الثقيلة تُعد عاملاً حاسماً في إدارة ومراقبة جميع جوانب عمليات الحفر. ومع ذلك، فإن الأساليب التقليدية غير كافية لتقدير إنتاجية المعدات بدقة وبشكل تلقائي. وبناءً على ذلك، يهدف هذا البحث إلى تطوير منصة ذكية متكاملة لتقدير الإنتاجية بدقة في مشاريع البناء الثقيلة، وخاصة عمليات الجرافات. تم تحليل البيانات الميدانية التي جُمعت من مواقع البناء في المملكة العربية السعودية، حيث شملت أكثر من 20 عاملاً حاسماً مثل وقت الدورة، مهارة المشغل، ظروف الطقس، ونوع التربة. تم استخدام مجموعة من أجهزة الاستشعار والأدوات الذكية لجمع البيانات، ثم تم إعدادها وتنظيفها وتصفية الضوضاء منها لتحسين جودتها. يقترح البحث خوارزمية تعلم آلي تعتمد على الشبكة العصبية الاصطناعية (ANN) لتقدير إنتاجية المعدات الثقيلة. بالإضافة إلى ذلك، تمت مقارنة أداء الخوارزمية المقترحة بأداء خوارزميتين أخريين للتعلم الآلي: آلات الدعم النقطية (SVM) وغابات القرارات العشوائية (Random Forest). وتم استخدام مجموعة من مقاييس التقييم المختلفة لتقييم أداء الخوارزميات المقترحة، وتشمل هذه المقاييس معامل التحديد (R^2)، متوسط الخطأ المطلق (MAE)، وجذر متوسط مربع الخطأ (RMSE). علاوة على ذلك، تم تحسين معايير الخوارزمية المقترحة بشكل مثالي باستخدام خوارزمية شاملة. تم تطوير منصة ويب ذكية تدمج نموذج الشبكة العصبية الاصطناعية، وتوفر تقديرات إنتاجية في الوقت الفعلي تحت ظروف متنوعة. أشارت النتائج إلى أن الشبكة العصبية الاصطناعية المقترحة قادرة على تقدير إنتاجية المعدات الثقيلة بدقة عالية، حيث حققت معامل تحديد R^2 بقيمة 0.98 ومقاييس خطأ منخفضة (MAE: 10.37 , RMSE: 14.59) متفوقةً على نماذج آلات الدعم النقطية وغابات القرارات العشوائية. كما كشفت النتائج أن المنصة المقترحة قادرة على تحويل البيانات المعقدة إلى رؤى عملية. وستُمكن هذه المنصة الذكية مديري البناء من تقليل التكاليف، وتسريع الجداول الزمنية، وتحسين كفاءة العمليات، مما يضع معياراً جديداً لإدارة الإنتاجية في صناعة البناء.

CHAPTER 1

INTRODUCTION

1.1 Study Background

Productivity is a crucial factor to be considered in every industry, and its definition may vary depending on the particular circumstances. Productivity is defined as the efficiency of productive effort, particularly in the industrial sector, and is calculated by the amount of output per unit of input [1]. The measurement of construction productivity commonly involves the use of unit rates, which represent the amount of time spent performing a specific amount of work [2]. All industries, including the construction sector, must prioritize productivity to reduce costs and generate profits. Depending on the construction jobs and the types of output being monitored, the units of measurement for productivity may change [3].

Heavy machinery, usually referred to as construction equipment (CE), is any large, strong self-propelled vehicle that is intended to perform a variety of construction activities. Since the usage of CE is essential for completing civil projects, the construction industry must make a significant financial commitment [4]. The term (CE) or Heavy Equipment, as used in this study, refers to machinery commonly employed in earth-moving operations, including bulldozers, excavators, dump trucks, loaders, compaction rollers, graders, scrapers, and similar types of equipment [5].

The productivity of large construction equipment, such as bulldozers, loaders, excavators, and hauling trucks, is crucial to the operation of earthworks [6]. Monitoring the productivity of heavy equipment is crucial for successfully managing and completing earthmoving projects. In this procedure, the operational effectiveness and performance of the equipment are measured, analyzed, and improved [7]. Real-time awareness of the status of the construction site is essential for ensuring efficient administration of the site and timely plan updates throughout the development process [8]. To accurately estimate productivity and calculate the time and cost required for earthworks operations, it is necessary to have an automated system that can capture job-site data and handle a large volume of data in near real-time. This system would enable correct productivity estimations and more accurate calculations of time and cost, improving overall efficiency and performance [9].

1.2 Artificial Neural Network ANN

In biological learning and memory processes, neural arrays act similarly to how ANNs do. The basic building blocks of ANNs are units or nodes, which are comparable to neurons, and weighted connections, which are comparable to the synapses seen in biological systems. These nodes act as the network's foundational information processing units. ANNs can have different node counts and connection patterns. The input and output layers of the network have several nodes that match the input and output variables in the dataset. The optimal number of nodes in the hidden layer is typically determined by trial and error. It is known that increasing the number of neurons enhances memory but may reduce an ANN's ability to reason. In order to reproduce the training data, an ANN should ideally have the fewest number of neurons possible [10 - 12].

The weight of each link between nodes corresponds to the knowledge that has been acquired through earlier learning cycles. The input-output relationship of the network can be simulated by changing these weights. In order to train the network, a number of processes are required, including (i) computing outputs based on the input data, (ii) comparing the results of the measurements and the calculations, and (iii) modifying each node's weight to reduce the discrepancy between measured and calculated values. The tool evaluates the present network using an independent calibration dataset at intervals that are specified by the user. The older network is replaced with the more recent one if the neural network's mean squared error predictions decrease as compared to the prior optimal network. If not, the prior network is kept in use until a new optimal network is found or the training process is finished [13].

A machine learning method called a neural network imitates the structure and function of the human brain. The network is comprised of linked nodes or "neurons," which analyze data and forecast results. Neural networks are commonly applied to image identification, natural language processing, and predictive analysis tasks [14]. Artificial neural networks are capable of enhancing the assessment of equipment productivity by providing a more precise and trustworthy approach to forecasting construction machinery production rates. Conventional techniques for productivity estimation, such as utilizing manufacturers' catalogs or relying on the expertise of site personnel, often yield substantial differences between expected and actual values. Linear regression models have also been employed for productivity estimation, but their precision may be limited. On the other hand, neural network models can be trained on historical data to recognize patterns and correlations between input variables and output values, enabling them to make more precise predictions

about future performance. Several experts have suggested that artificial neural network-based models are effective tools for estimating the productivity of construction equipment [15 - 23].

1.2.1 An Artificial Neural Network's Main Components (ANN)

The major components of an Artificial Neural Network (ANN), it is essential to understand the fundamental structure that underlies this powerful computational model. ANNs are made up of connected artificial neurons and are modeled after the biological neural networks in the human brain. Through weighted connections, these layered neurons analyze and transfer information. The flow of data through an ANN follows a forward direction, with each neuron performing computations and applying activation functions. Understanding the basic structure of an ANN lays the foundation for comprehending the key elements that contribute to its functionality and learning capabilities, as shown in [Figure 1.1](#) below [10, 13, 24].

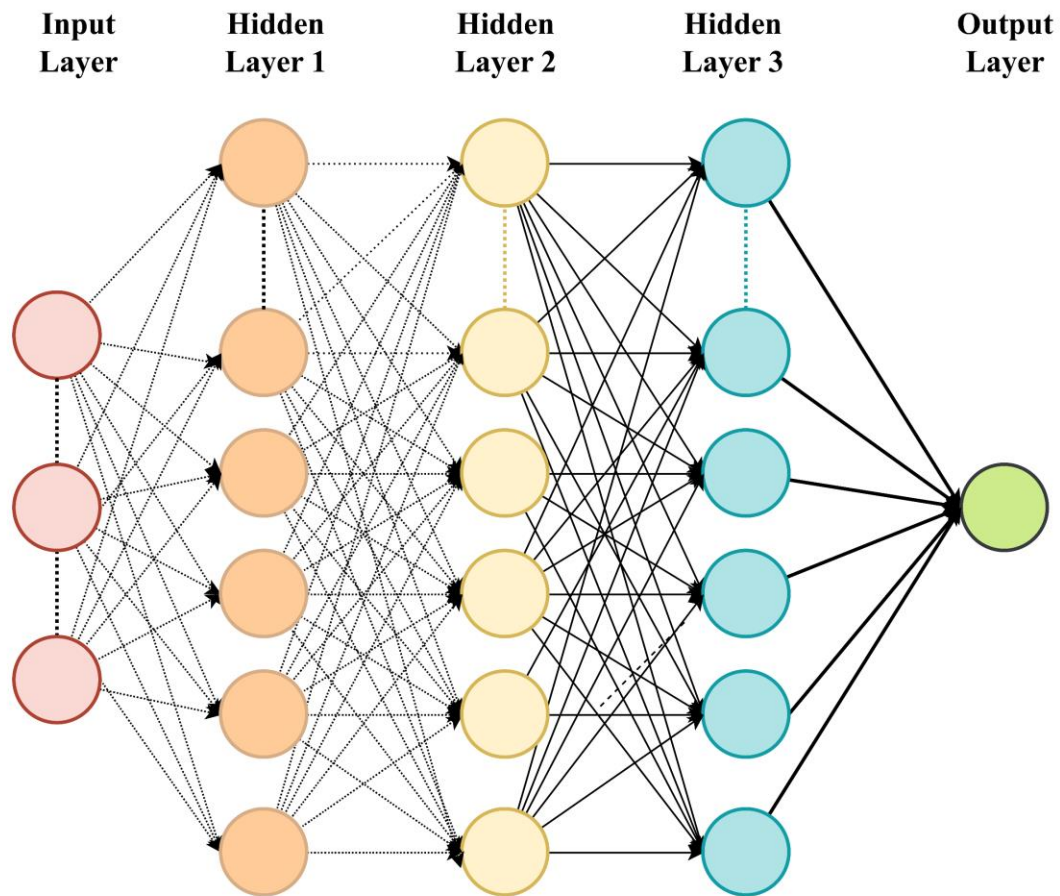


Figure 1.1: Architecture of the Artificial Neural Network (ANN) Model

1. **Input Layer:** An input layer is the first layer of the ANN and is where the network receives its input of data or features. A characteristic or property of the input data is represented by each neuron in the input layer. The input layer serves as the connection point between the network and external data.
2. **Hidden Layers:** The hidden layers of an ANN are the layers in-between the input and output layers. They perform the computations and transformations that enable the network to learn complex patterns and relationships within the data. The number of

hidden layers and the number of neurons within each layer can vary depending on the complexity of the problem being addressed.

3. **Output Layer:** It is the final layer of the ANN, which produces the desired output or prediction. The number of neurons in the output layer depends on the nature of the task—whether it is a classification problem with multiple classes or a regression problem with a continuous output. Each neuron in the output layer represents a specific class or value to be predicted.
4. **Connections (Weights):** Connections in an ANN represent the links between neurons in different layers. These connections are associated with weights that determine the strength of the influence between neurons. At the beginning of the training process, random values are assigned to the weights, and throughout training, they are modified to enhance the network's performance. The weights govern the flow of information through the network and play a crucial role in determining the output.
5. **Activation Functions:** The computations carried out by each neuron in an ANN become nonlinear due to activation functions. Based on the weighted sum of a neuron's inputs, they calculate its output. ReLU (Rectified Linear Unit), tanh, and sigmoid are often used activation functions. The network is made more adaptable by the activation function, enabling it to represent complex interactions and make nonlinear predictions.
6. **Training Algorithm:** The weights of the connections in the network are modified by the training algorithm to reduce the discrepancy between the expected and actual outputs. Backpropagation is a frequently employed training approach that determines the gradient of the mistake and propagates it backward through the network to update

the weights. Adaptive moment estimation (Adam) and stochastic gradient descent (SGD) are two other techniques that are frequently employed.

These major parts work together in an ANN to process input data, learn from examples, and generate predictions or outputs based on the learned patterns and relationships.

1.2.2 Artificial neural networks' most significant subtypes (ANNs)

The explanations of different types of Artificial Neural Networks (ANNs), as shown in [Figure 1.2](#), highlight the diverse architectures developed to cater to specific data characteristics and problem domains. ANNs have evolved to encompass a wide range of models, each designed with unique features and capabilities. From feedforward networks that excel in traditional classification and regression tasks to specialized architectures like Convolutional Neural Networks for image processing and Recurrent Neural Networks for sequential data analysis, the realm of ANNs offers a rich array of tools for various applications. Let us now explore some of the most important types of ANNs and their distinct advantages and applications.

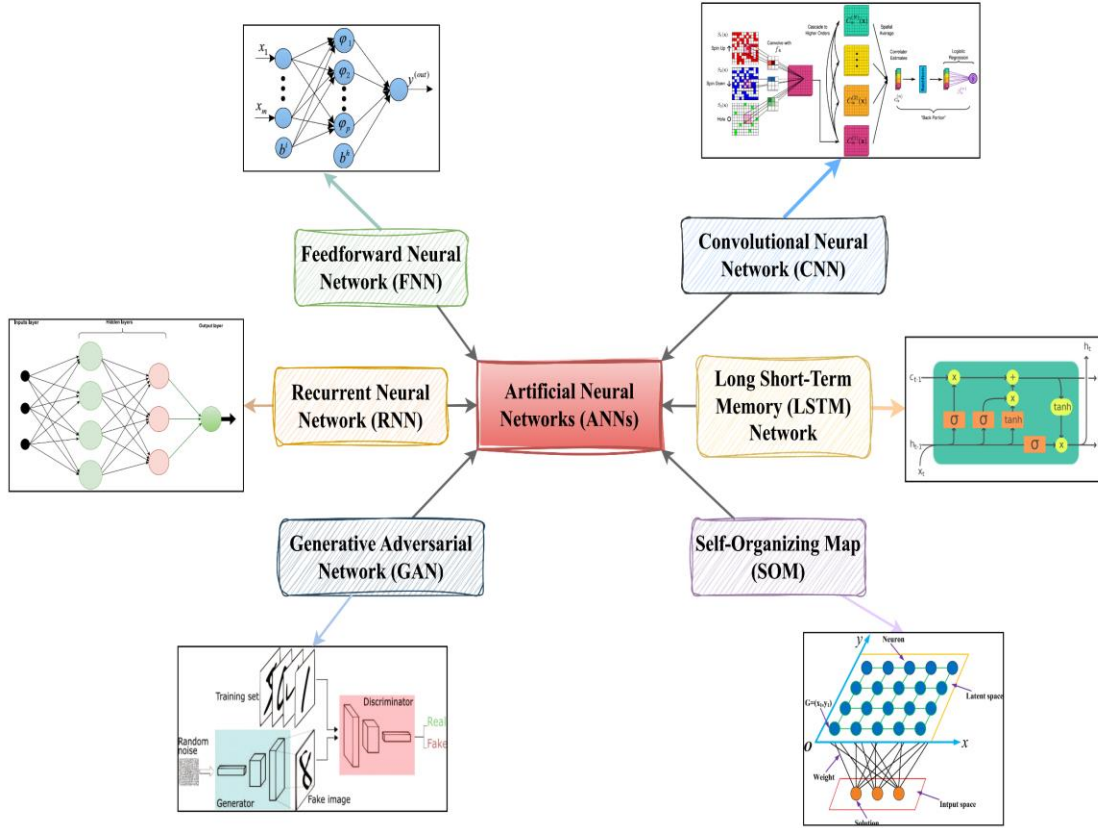


Figure1.2: Overview of Major Types of Artificial Neural Networks (ANNs)

1. **Feedforward Neural Network (FNN):** An FNN is the simplest and most common type of ANN. Information flows in only one direction, from the input layer through the hidden layers (if any) to the output layer, without any loops or feedback connections. FNNs are primarily used for tasks like classification and regression, where the output depends solely on the input [25].
2. **Convolutional Neural Network (CNN):** The processing of grid-like input, such as images or signals, is the sole purpose of a CNN. Convolutional, pooling, and fully linked layers, among others, are used in CNNs. Convolutional layers use convolution

- procedures to extract spatial information, whereas pooling layers shrink the spatial dimensions of the input. For computer vision, object identification, and picture classification applications, CNNs are frequently utilized [26].
3. **Recurrent Neural Network (RNN):** Recurrent connections between the neurons in an RNN are used to create a feedback loop and process sequential data. RNNs are effective for tasks like time series analysis, speech recognition, and natural language processing where context and temporal dependencies are crucial [26, 27].
 4. **Long Short-Term Memory (LSTM) Network:** An LSTM Network is a kind of RNN that solves the vanishing gradient issue and is better able to learn long-term dependencies. Gating mechanisms and memory cells are introduced, which regulate the network's information flow. In tasks requiring long sequences, such as language modeling, machine translation, and speech recognition, LSTM networks have excelled [27, 28].
 5. **Generative Adversarial Network (GAN):** A generator and a discriminator, which are trained together in a competitive environment, make up a GAN. The discriminator aims to discriminate between actual and fake samples, whereas the generator creates artificial data samples. GANs are being used to produce synthetic data, deep fakes, and realistic visuals for a variety of applications [29].
 6. **Self-Organizing Map (SOM):** An unsupervised learning algorithm called a SOM, also referred to as a Kohonen network, uses competitive learning to produce a low-dimensional representation of the input data. Clustering, visualization, and data exploration tasks frequently involve the use of SOMs. In high-dimensional data, they can make hidden links and patterns apparent [30].

These are just a few illustrations of the various ANN architecture designs that have been created to address various data and problem domains. Each type of ANN is built to handle particular data characteristics and excel in a variety of application domains.

1.3 Research Gaps

The methods that are currently used to determine the actual productivity of large construction equipment on job sites are inadequate and underutilized. This suggests that one of the main challenges is the lack of a reliable and widely accepted method for estimating actual productivity [31]. Contractors, especially those working on earthwork projects, face a sizable challenge when estimating the productivity of construction equipment. Equipment production rates have typically been estimated using data from manufacturer catalogs or by consulting site personnel's observations and evaluations. However, the actual production rates seen at various construction sites show how ineffective these methods are at producing precise results. This has led to a rise in the demand for more precise models that can accurately predict the productivity of construction equipment [14].

Despite the widespread adoption of ANN in various fields, its application specifically for estimating the productivity of heavy equipment in construction projects remains relatively underexplored. Previous research studies have predominantly focused on traditional regression models or statistical approaches for productivity estimation. The potential of ANN, with its ability to capture complex nonlinear relationships and adapt to changing input patterns, has not been extensively ruled in this context. Therefore, there is a notable research gap that calls for a deeper investigation into the feasibility and effectiveness of

utilizing ANN models for accurately predicting and estimating the productivity of heavy construction equipment. Furthermore, while ANN holds promise as a predictive modeling tool, there is a need for more comprehensive studies that address the specific challenges and refinements associated with using ANN for productivity estimation in the heavy equipment domain. Factors such as equipment types, project characteristics, operational parameters, and environmental conditions all contribute to overall productivity. Understanding how these variables interact and influence the accuracy of the ANN models is critical. Closing this research gap will offer insightful information about the benefits, drawbacks, and real-world implications of using ANN to estimate the productivity of heavy machinery, ultimately facilitating better resource allocation and decision-making in construction projects.

1.4 Statement of the Problem

Heavy construction equipment in Saudi Arabia faces significant challenges in accurately estimating and managing onsite productivity. Existing estimation methods often rely on manual calculations and subjective assessments, leading to inconsistencies, inefficiencies, and potential errors. The lack of an integrated smart platform specifically designed for estimating onsite productivity further exacerbates this problem. The absence of a centralized system that leverages advanced technologies, data analytics, and real-time monitoring hinders the ability to make informed decisions and optimize resource allocation. Consequently, there is a pressing need to develop an integrated smart platform tailored to the unique requirements of heavy construction projects in Saudi Arabia, capable of providing accurate and timely estimates of onsite productivity. Estimating the actual productivity achieved by construction machinery on a construction site can help avoid

incorrect, imprecise, and subjective estimates made during the planning phase of earthworks performance [31]. Vision-based methods allow direct visualization of equipment state from images and videos, facilitating easy identification of false recognition and analysis of the underlying causes of low productivity. Conversely, non-vision-based methods have difficulty in categorizing activities in detail and require the attachment of sensors to monitor the equipment [32]. The actual productivity levels achieved during earthworks operations are significantly distinct from the estimated productivity values assumed during the planning stage [33].

Another critical issue in the estimation of onsite productivity for heavy construction projects in Saudi Arabia is the scarcity of standardized and reliable data sources. Construction projects involve numerous variables, including equipment usage, labor productivity, weather conditions, and material availability. However, gathering comprehensive and up-to-date data related to these factors is often challenging, resulting in uncertainties and inaccuracies in productivity estimates. The lack of a reliable data infrastructure poses a significant hurdle in developing an effective smart platform for productivity estimation. Addressing this problem requires the establishment of robust data collection mechanisms, data integration protocols, and quality assurance processes to ensure the availability of accurate and consistent data. By addressing these issues, an integrated smart platform can offer a solution to enhance onsite productivity estimation for heavy construction projects in Saudi Arabia, leading to improved project planning, resource management, and overall project success. Hence, it is crucial to identify and manage the factors that influence productivity in construction projects to enhance their economic performance. Having a thorough and all-encompassing comprehension of these

factors would enable the allocation of research efforts toward improving productivity, ultimately leading to increased opportunities for productivity enhancement [3].

1.5 Objective of the study

In the dynamic region of construction projects, optimizing productivity and efficiency is of paramount importance. Heavy construction equipment plays a pivotal role in shaping the success of such projects, making it crucial to understand the factors that impact their productivity. This endeavor necessitates a comprehensive approach that combines existing research findings with expert knowledge. In this context, the objectives of this study revolve around identifying the key factors affecting heavy construction equipment productivity, developing an ANN model for productivity prediction, and designing a web-based system to estimate onsite productivity. By accomplishing these objectives, the aim is to enhance construction practices in Saudi Arabia. Specifically, the main objectives of this research are to:

- i. Identify factors that most affect the productivity of heavy construction equipment in construction projects in KSA based on previous studies and conducting expert meetings.
- ii. Develop an ANN to predict the productivity of heavy construction equipment.
- iii. Develop the required specifications for a web-based system for estimating onsite productivity for heavy construction projects in Saudi Arabia.

1.6 Significant of the study

Large amounts of data from numerous sources, including equipment sensors, weather conditions, past performance, and project parameters, can be analyzed by ANN algorithms. ANN models can produce more accurate estimates of equipment productivity by

processing this data. This precision aids in informed decision-making, resource allocation, and effective project planning that boosts output and lowers costs for project managers. ANN-based systems are capable of continuously monitoring the real-time operation of heavy equipment and providing predictive insights. ANN models can find patterns and anomalies by gathering and analyzing data on equipment performance, fuel consumption, maintenance records, and operator behavior. This makes it possible to perform preventative maintenance, identify equipment problems before they become serious, and identify potential productivity bottlenecks. With the help of these insights, project managers can act quickly to avoid delays, reduce downtime, and increase productivity.

Continuous Learning and Improvement: ANN models can develop over time. ANN algorithms can adjust and improve their predictions of equipment productivity by gathering feedback data and incorporating it into the training process. Future estimates are more accurate and reliable thanks to this iterative learning process. Construction companies can make data-driven decisions, implement process optimizations, and increase productivity on upcoming projects thanks to AI systems' ability to analyze historical data to spot trends, performance patterns, and potential improvement areas.

Finally, the significance of our research can be summarized in three key points.

- Automating data acquisition for estimating onsite actual productivity of outdoor operations of heavy construction projects.
- Automating the productivity estimation process for outdoor operations in heavy construction projects.
- Optimizing the available resources to contractors in heavy construction projects.

1.7 Scope and Limitations

By using the ANN methods and sensors to track the productivity of heavy equipment on construction sites, the research aims to develop a predictive model for estimating the productivity of heavy equipment. The study aims to investigate whether the construction industry can estimate productivity more accurately by combining sensor data with ANN algorithms. The research project will include data gathering from construction sites, sensor data preprocessing and analysis, ANN model training and validation, and evaluation of the developed predictive model's performance.

Research Limitations:

1. **Data Availability and Quality:** Access to thorough and high-quality sensor data for the productivity of heavy equipment may be difficult for the research. The availability of historical data, the variety of data collection techniques, and potential data biases may have an impact on the developed model's accuracy and dependability.
2. **Generalization to Diverse Construction Settings:** Due to differences in project types, equipment characteristics, and environmental conditions, the study's findings and the developed predictive model may not be fully generalizable to a variety of construction projects. The model's efficacy and accuracy may differ across various construction settings, and further validation may be necessary.
3. **Complex Relationship and Variables:** Several variables, such as operator expertise, project difficulty, site conditions, and maintenance activities, affect the productivity of heavy equipment. There may be difficulties in accurately capturing all these variables with sensors and adding them to the ANN model. If specific important

variables are not sufficiently captured or taken into account, the model's ability to predict outcomes may be limited.

4. **Implementation Obstacles:** Setting up a sensor-based data collection system and creating an ANN model for real-world construction applications can be difficult due to factors like cost, technical know-how, integration with existing equipment, and data management. The scalability and applicability of the research findings in actual construction projects may be constrained by these issues.

The proper interpretation and application of the study's findings will be ensured by addressing these limitations, communicating the research scope and potential limitations clearly, and advancing the use of sensor data and ANN techniques for productivity estimation in the construction industry.

1.8 Thesis Organization

This thesis is methodically arranged to explore the research objectives and delve into an in-depth analysis of applying the ANN for estimating the onsite productivity of heavy construction equipment in Saudi Arabia. The organization of the thesis unfolds as follows:

Chapter 1: Introduction This initial chapter sets the foundation for the thesis, delineating the study's backdrop with a special emphasis on productivity within the heavy construction sector and the innovative role of ANNs. It identifies the research gaps, crafts the problem statement, articulates the objectives, and underscores the significance of the research. The scope and limitations inherent to the study are also discussed.

Chapter 2: Literature Review An extensive review of the literature is presented, surveying the definitions of productivity, the progression of productivity estimation methodologies, the application of ANNs across diverse sectors, and previous investigations within the construction industry. This chapter will also discuss factors affecting the productivity of heavy equipment, drawing from both empirical studies and theoretical discourse, setting a solid framework for the study, and positioning this research within the broader academic dialogue.

Chapter 3: Methodology The methodology chapter will explicate the research design and outline the strategies employed for data collection, as well as the comprehensive process adopted for developing the ANN model. This will encompass the selection of productivity-affecting factors, neural network configuration, and the methods for data preprocessing, model training, testing, and validation.

Chapter 4: Results and Discussion Here, the outcomes of the ANN model will be elucidated and deliberated, with an emphasis on how the identified factors affect the productivity of heavy construction equipment. The performance of the ANN model will be benchmarked against conventional productivity estimation approaches, with an interpretive discussion on what the results signify for the management of heavy construction projects in Saudi Arabia.

Chapter 5: Conclusion and Recommendations The concluding chapter will synthesize the primary insights of the research, discussing both their practical applications and theoretical contributions. It will present recommendations informed by the research findings and propose avenues for future inquiry. The chapter will also reflect on the

research's contribution to the academic and industrial understanding of productivity in heavy construction and the prospective influence of an integrated smart platform on the construction landscape in Saudi Arabia.

1.9 Expected Outcomes:

This research is anticipated to produce several meaningful outcomes that will significantly contribute to the field of construction management and demonstrate the transformative potential of artificial intelligence, particularly ANNs, in productivity estimation.

A key outcome of the study is the identification and analysis of critical productivity factors affecting heavy construction equipment in Saudi Arabia. By compiling and analyzing a comprehensive list of these factors, the research will provide a nuanced understanding of how each variable interacts to influence overall productivity. This will equip construction managers with actionable insights to prioritize and address the most impactful factors for optimization.

The study will also develop an advanced ANN model capable of accurately predicting the productivity of heavy construction equipment. Designed to handle the complexities and dynamic nature of construction site operations, this model will set a benchmark for employing advanced computational tools in construction. Its performance will be systematically evaluated against traditional estimation methods, showcasing its superiority in managing diverse and complex data sets. This validation will further strengthen the case for adopting ANN-based approaches in the industry.

To bridge the gap between theory and practice, the research will propose a practical framework for implementing the ANN model in real-world construction projects. This framework will provide detailed guidelines for data collection, model training, and output interpretation, ensuring that construction teams can seamlessly integrate AI-driven productivity estimation into their workflows. Additionally, the study will deliver practical recommendations for construction managers and policymakers, outlining strategies for adopting ANN-based methods and formulating policies to support technology integration in the sector.

Farther than its practical applications, the research will make a substantial academic contribution by addressing gaps in the existing literature. It will provide empirical evidence of ANN's effectiveness in estimating construction productivity, enriching the theoretical foundation of the field and paving the way for future studies. By highlighting unexplored areas, such as the broader application of machine learning algorithms in construction management, the study will inspire further innovation and investigation.

In realizing these outcomes, the thesis will not only underscore the viability of employing advanced computational methods in construction but also provide actionable insights and strategies that can be translated into improved productivity and operational efficiency across the industry.

1.10 Chapter Summary:

In this introductory chapter, the foundation was established for exploring the application of ANNs to enhance onsite productivity estimation for heavy construction projects in Saudi Arabia. The background emphasized the critical role of productivity in the construction

industry, especially given the substantial financial investments associated with construction equipment. The necessity of developing an accurate and automated system to improve efficiency and performance in such projects was highlighted. The concept of ANNs was introduced, drawing comparisons to biological neural processes, and their potential to revolutionize productivity estimations was underscored through their advanced pattern recognition and predictive capabilities.

The research identifies a gap in utilizing ANN methodologies for estimating the productivity of heavy construction equipment, as traditional methods are less accurate. This led to the problem statement focusing on the need for an integrated smart platform in Saudi Arabia that can provide precise productivity estimations. Consequently, the study aims to identify key factors affecting productivity, develop an ANN model for accurate predictions, and outline the specifications for a web-based system to support onsite productivity estimation.

The significance of this study is its potential to transform the construction industry by offering a more reliable basis for project planning and management through the use of AI algorithms and continuous learning systems. However, the scope of the study is constrained by challenges such as data availability and quality, generalization issues, and the complexity of capturing relevant variables. The expected outcomes include a detailed ANN model and a framework for its practical implementation, contributing to both academic literature and real-world construction practices.

As the study progresses, the subsequent chapters will expand on the foundation established in this introduction. The next chapter will present a comprehensive literature review,

offering a theoretical context for the research. This will be followed by a detailed discussion of the methodology employed, an in-depth analysis of the results, and a final chapter summarizing the conclusions and recommendations derived from the findings.

CHAPTER 2

LITERATURE REVIEW

2.1 Review of Existing Studies

The two most important resources in the construction sector are laborers and equipment. By monitoring the performance of these resources, project managers can boost productivity at construction sites and identify any potential performance problems. Location tracking, activity recognition, activity tracking, and performance monitoring are the four main components of a typical construction workforce monitoring system [34]. One of the principal responsibilities of the overseer is to monitor and oversee operations to ensure that the entire project meets the anticipated production rates [34]. The factors that affect productivity in the construction industry have been the subject of extensive research [3]. The expected productivity, or the amount of work that is expected to be produced in each amount of time (such as an hour or a day), is a key factor in determining the price and time required for construction activities [35].

Therefore, efficiency and effectiveness come out as shared subthemes in all contextual interpretations of productivity, regardless of the viewpoint or terminology used [36]. The estimation approaches employed and the operational factors selected significantly impact productivity estimation [37] compared different productivity estimation techniques and evaluated how sensitive they were to changes in operational coefficients, specifically for excavation activities. A structured framework was introduced, and a generalized linear mixed model was developed to determine the productivity of a common bulldozer type

depicted as a central node, highlighting its pivotal role in the discourse. Connected closely to it are terms such as construction, construction industry, and construction equipment, emphasizing their interdependence in productivity-related studies. Green nodes represent themes like efficiency, decision-making, maintenance, and benchmarking, which indicate a focus on operational improvements and resource optimization. Blue nodes cluster around labor productivity, project management, and construction productivity, signifying the critical examination of human and managerial aspects. Red nodes, including sustainability, optimization, and excavation, reflect emerging concerns with environmental impact and task-specific challenges. This interconnected web signifies the multidisciplinary nature of productivity studies, combining technological, managerial, and sustainable approaches to drive progress in the construction sector. The graph underscores the broad spectrum of research directions, from factor analysis to cost-effectiveness and accident prevention, offering a wholistic view of the industry's evolving priorities.

2.3 Concept of Onsite Productivity:

Onsite productivity in construction fundamentally refers to the efficiency and effectiveness with which resources are utilized to complete construction activities within a set timeframe and budget. It is a measure of output typically in terms of volume, area, or units achieved per unit of input, such as labor hours, equipment usage, or material consumption. Gained from the content provided, onsite productivity is influenced by a constellation of factors ranging from the performance of labor and equipment to the deployment of modern technologies like ANNs for monitoring and prediction. The introduction of smart platforms is reshaping the understanding and management of onsite productivity, allowing for real-

time data capture and analysis, ultimately leading to enhanced decision-making and operational efficiency.

Existing studies, as highlighted in the literature, shed light on the various components and influences on onsite productivity, including location tracking, activity recognition, and performance monitoring. The superintendent's role in overseeing operations to meet production rates is critical, as they bridge the gap between predictive estimations and actual onsite output. Additionally, the extensive research into productivity factors highlights the industry's focus on optimizing the efficiency of both labor and equipment as critical resources. The integration of technology into construction operations has introduced a new dimension to onsite productivity. With tools for monitoring equipment efficiency and emissions, the industry can now better understand the environmental impact of productivity. The development of sensor technologies and AI applications also suggests a trend towards a more refinement and advanced approach to productivity estimation, incorporating real-time monitoring and predictive analytics. Therefore, the concept of onsite productivity, as it stands today, is an uniting of traditional practices, innovative technologies, and a deep understanding of the multitude of factors that contribute to efficient construction project delivery. This comprehensive view encapsulates not just the quantitative aspect of productivity but also the qualitative elements that ensure project success.

2.4 Construction Industry in Saudi Arabia:

The construction industry in Saudi Arabia holds a central position in the nation's economic development, driven by substantial government investments in infrastructure and strategic

projects. As part of Vision 2030 an ambitious plan to diversify Saudi Arabia's economy and reduce reliance on oil revenues the government has initiated transformative construction ventures, including megaprojects like NEOM, the Red Sea Project, and Qiddiya. These initiatives aim to redefine urban planning and economic diversification by fostering innovation, enhancing tourism, and integrating sustainable construction practices. For instance, NEOM is envisioned as a futuristic city that embodies smart technologies and renewable energy, reflecting the country's commitment to a sustainable and diversified economy. This dynamic growth, however, is underpinned by a clear vision to align the construction sector with global standards while addressing the country's unique regional needs [39].

Despite the rapid advancements, the sector faces notable challenges, such as a heavy reliance on an expatriate workforce, the need for skilled labor, and the incorporation of advanced construction technologies. The Saudi government has proactively addressed these issues through Saudization policies, aimed at increasing the participation of Saudi nationals in the construction workforce, and initiatives to streamline project approvals. Furthermore, technological advancements such as BIM, smart construction methods, and the use of sustainable materials have gained prominence in the sector. These technologies not only enhance operational efficiency and productivity but also support Vision 2030's environmental objectives by promoting green building practices. The construction industry's expansion significantly contributes to Saudi Arabia's economic diversification, creating jobs, stimulating local industries, and attracting foreign investments, thereby cementing its role as a cornerstone of the nation's development agenda [39].

2.5 Sensor Data Integration and Analysis

The positioning, communication, and mapping sensors work with a variety of platforms. The choice of platform is crucial, as different platforms offer varying benefits and limitations in terms of precision and effectiveness [40]. The Sensor Data Processing Module can recognize actions and occurrences that have been determined experimentally to be inconsistent or mistaken [41]. A variety of sensors are used on construction sites to track the productivity of large equipment [42]. The number of operational hours is frequently recorded using hour meters, which provide information on how effectively equipment is being used and how to plan maintenance. Assisting in fleet management and productivity analysis, GPS tracking systems use satellite technology to track the movements and locations of large machinery in real-time. Load cells are necessary for determining the weight or load being lifted or carried by machinery to ensure safe operations [43]. Critical component temperatures are monitored by temperature sensors, allowing for the early identification of overheating or potential failures [42]. By incorporating these sensors into heavy equipment, contractors can gather valuable data for productivity analysis, maintenance optimization, and informed decision-making on construction sites.

2.6 Productivity Estimation Models

Technologies such as video analysis and other non-vision-based sensing and tracking methods can be utilized to assess the productivity of heavy construction equipment on job sites. These technologies offer potential advantages, including more precise and efficient productivity estimates, as well as the ability to monitor and analyze data in real-time [31]. These techniques can assist in identifying and resolving problems more quickly, reducing

downtime, and increasing overall efficiency by enabling real-time monitoring and control of the production process [44].

In the construction industry, a wide range of models and methods are utilized to estimate productivity. These approaches span from traditional techniques, such as manual observations and historical data analysis, to advanced technologies like machine learning and computer vision. Below, some of the commonly applied methods and their relevance to productivity estimation are explored [16, 45 - 49]:

- i. **Traditional Methods:** Traditional productivity estimation methods rely on established industry practices and techniques, often involving manual data collection, analysis of historical records, and expert judgment. For instance, the work sampling method requires observing workers at regular intervals to assess their activities and productivity levels. Similarly, the production function method evaluates productivity by analyzing inputs such as labor, equipment, and materials. While these traditional approaches can be time-intensive, they offer valuable insights and serve as useful benchmarks for estimating construction productivity.
- ii. **Computer Vision and Image Analysis:** In recent years, computer vision and image analysis techniques have become increasingly important for estimating productivity in construction. These methods process images or video footage captured on-site to extract relevant information, using machine learning algorithms trained to identify and classify activities, equipment, and materials. This approach enables automated analysis of construction processes and provides accurate productivity metrics. By offering real-time monitoring, improved precision, and

reduced manual effort, computer vision-based methods present a transformative advancement in productivity estimation.

- iii. **Building Information Modeling (BIM):** Building Information Modeling (BIM) serves as a digital representation of the structural and functional aspects of a construction project. It integrates key components such as design, scheduling, and cost estimation. By simulating construction sequences, resource allocation, and task durations, BIM facilitates productivity estimation with a high level of detail and accuracy. This approach provides a holistic view of the project, enabling stakeholders to identify bottlenecks, streamline workflows, and make data-driven decisions to enhance productivity and overall project efficiency.
- iv. **Data-Driven Modeling and Artificial Intelligence:** Data-driven modeling and AI techniques have become increasingly popular for estimating productivity in the construction industry. These methods leverage large datasets to identify patterns, and correlations, and develop predictive models. Machine learning techniques such as ANNs and SVMs are commonly employed, enabling the creation of models trained on historical data to estimate productivity based on multiple input parameters. These models continuously adapt and improve with new data, offering more accurate and dynamic productivity estimations while reducing reliance on traditional, manual methods.
- v. **Mobile and Sensor Technologies:** Mobile and sensor technologies have revolutionized productivity estimation by providing real-time data collection and analysis capabilities. Mobile applications allow construction professionals to input data on-site, enabling instant productivity tracking and reporting. Sensor

technologies, such as wearable devices and IoT sensors, capture data on worker movements, equipment usage, and environmental conditions. This data can be processed to derive productivity indicators and identify areas for improvement.

By leveraging a combination of traditional methods and emerging technologies like computer vision, BIM, data-driven modeling, and mobile/sensor technologies, the construction industry can enhance its ability to estimate and improve productivity. These models and methods offer opportunities for more accurate, efficient, and data-driven decision-making, ultimately leading to optimized resource allocation and improved project outcomes. Moreover, the integration of AI and Sensor Technologies presents a promising approach for estimating and enhancing productivity in the construction industry. By combining the power of AI-driven data analysis and real-time data capture from sensors, construction professionals can gain valuable insights and make informed decisions to optimize productivity [50].

Artificial Intelligence, specifically machine learning algorithms, can analyze vast amounts of historical data to identify patterns, correlations, and predictive models related to productivity. These models can then be used to estimate productivity based on various input parameters, enabling proactive planning and resource allocation [13]. Sensor technologies, on the other hand, provide real-time data on worker activities, equipment usage, and environmental conditions. Wearable devices, IoT sensors, and other sensor technologies capture and transmit data, allowing for continuous monitoring of construction processes. This real-time data feeds into the AI models, enabling dynamic productivity estimation and identifying potential bottlenecks or areas for improvement [40]. The

synergy between AI and Sensor Technologies offers several advantages. It allows for data-driven decision-making, immediate identification of issues or inefficiencies, and proactive intervention to optimize productivity. Construction professionals can leverage this combination to monitor worker performance, track equipment utilization, identify safety risks, and implement targeted interventions to enhance overall productivity. By embracing this hybrid approach, researchers and practitioners can bridge the gap between data analysis and real-time monitoring, leading to improved productivity estimation, enhanced resource allocation, and more efficient project management in the construction industry. The integration of AI and Sensor Technologies holds significant potential to revolutionize productivity estimation and contribute to the overall advancement of the construction sector.

2.7 Key Factors Influencing Construction Equipment Productivity

Understanding the variables that affect heavy equipment performance on-site is essential for effective management. An overview of the critical elements that are critical in determining equipment productivity in construction projects is provided in [Table 2.1](#). These factors cover a wide range of variables, such as Soil type, resource allocation, site conditions, operator skill level, and equipment maintenance procedures. Project managers can implement strategies to optimize equipment utilization and raise overall productivity by thoroughly examining and addressing these factors [3]. In order to provide a more thorough understanding, these aspects will be further developed in the parts that follow.

2.7.1 Factors Details

Construction equipment productivity is influenced by a wide spectrum of interrelated factors, each contributing to the operational efficiency, economic feasibility, and overall success of construction projects. These factors can be categorized into six major domains: operational, environmental, human, economic, technological, and management. Together, they form a complex ecosystem where the optimization of one aspect often impacts the others.

Operational factors play a foundational role in determining productivity [51 - 53]. The age and condition of equipment directly influence its reliability and performance, with older machines often requiring more maintenance and experiencing frequent breakdowns [54, 55]. The equipment's utilization rate, blade size, load size, and hauling distance are critical in determining how effectively it can execute tasks [3, 54 - 57]. For instance, equipment operating below its capacity results in inefficiencies, while improper blade size or type can hinder excavation and grading activities. Idle time, downtime due to breakdowns, and maintenance frequency add layers of complexity, as each moment of inactivity translates to lost productivity. Additionally, factors like fuel efficiency and engine power impact operational costs, making it essential to select machines with optimal performance metrics. The type of terrain and soil, as well as site congestion, further dictate equipment mobility and working cycle times, affecting how quickly and efficiently tasks are completed [3, 58 - 62]. Methods such as work cycle optimization and proactive maintenance crew response can mitigate delays and ensure seamless operations.

Environmental factors also significantly influence equipment performance, as construction sites are subject to dynamic and often harsh conditions. Variables such as weather (rain, wind, snow) and temperature extremes (hot or cold) affect equipment functionality and operator efficiency. Soil type, ground moisture levels, and site grades add further challenges, necessitating specialized equipment configurations to handle adverse conditions effectively [3, 63 - 66]. Visibility during operations, often hindered by dust, wind, or poor lighting, can reduce precision and increase safety risks. Seasonal variations, such as prolonged rainy or dry periods, require adaptive planning to maintain consistency in output. Additionally, environmental constraints, such as noise and air pollution restrictions or vegetation clearance, necessitate compliance with local regulations. These factors, coupled with challenges like flood-prone areas and ground frost depth, underscore the importance of environmental adaptability in construction operations [58, 67 - 69].

The human element is equally critical in construction productivity. Operators' skills, training levels, and experience have a direct bearing on the efficiency and safety of operations. Fatigue, work shifts, and the availability of rest breaks impact decision-making and precision, which are crucial for heavy equipment operation [70 - 72]. A motivated workforce and a positive on-site work culture foster productivity, while collaboration between operators and supervisors ensures cohesive teamwork and task alignment. Additionally, communication between operators and site managers facilitates real-time issue resolution and adherence to safety training minimizes the risks of accidents. Operators' physical health and cognitive load also play a role, as mental and physical strain can lead to errors or reduced performance. Managing absenteeism, retaining skilled

operators, and ensuring readiness to adopt advanced technologies are essential strategies for maintaining workforce stability and capability [70 - 74].

Economic factors highlight the financial pressures associated with equipment productivity. Fuel costs, maintenance expenses, and equipment replacement costs are among the most significant contributors to project budgets. Frequent downtime, high repair costs, and investment in new technologies often create financial strain, particularly for projects operating within tight margins. Costs related to transportation, equipment upgrades, lubricants, and engine repairs must also be carefully planned and managed. Budget constraints can result in the use of suboptimal equipment, leading to inefficiencies and delays. Additionally, rental versus ownership decisions require careful consideration of project duration, equipment requirements, and financial capacity, as these choices directly impact long-term economic feasibility [3, 75, 76].

Technological advancements are redefining the construction industry's approach to productivity. GPS tracking systems, real-time monitoring, and machine learning-based predictive maintenance are among the cutting-edge tools enabling contractors to improve efficiency and precision [6, 60, 70, 77]. For instance, GPS systems allow real-time tracking of equipment location and movement, which aids in fleet management and scheduling. Predictive maintenance technologies reduce downtime by identifying potential failures before they occur, ensuring seamless operations. Onboard sensors, cameras, and automatic blade positioning technologies enhance operator accuracy and reduce manual intervention. IoT-enabled maintenance and digital twin technologies are revolutionizing data integration, allowing for more informed decision-making and streamlined operations.

Additionally, the development of autonomous bulldozers and advanced navigation systems points to a future where automation plays a central role in construction productivity [40, 78 - 80].

Effective project planning and supervision ensure that all operational and human resources are aligned with project goals. Resource allocation, particularly in the context of equipment, labor, and finances, must be optimized to address potential bottlenecks. Clear communication channels, regular progress assessments, and proactive risk management contribute to smoother project execution. As construction projects grow in complexity, strong management practices become the backbone of operational efficiency and productivity [81 - 84].

The top 20 factors as shown in [Table 2.1](#), identified as the most significant influences on productivity, draw attention to specific critical areas. These include operational aspects like equipment age, utilization rate, blade size, and load size, which are pivotal in achieving optimal efficiency. Environmental factors such as soil type, weather conditions, and temperature extremes highlight the need for adaptability to external conditions. Human factors like operator skill, motivation, and collaboration underscore the value of a trained and cohesive workforce. Economic considerations, including fuel costs and maintenance expenses, require strategic resource management to balance operational efficiency and financial sustainability. Lastly, technological tools like GPS tracking and real-time monitoring enable data-driven decision-making, setting the stage for advanced construction practices.

Table 2.1: Factors Affecting the Productivity of Bulldozers

Category	Factors	References
Operational	Equipment Age and Condition	[81, 85 - 88]
	Blade Size and Type	[3, 58, 73, 89]
	Load Size	[3, 58, 73, 89]
	Hauling Distance	[3, 60 - 62]
	Equipment Utilization Rate	[3, 58, 73, 89]
Economic	Equipment Rental vs. Ownership Costs	[6, 58, 90]
	Maintenance costs	[6, 58, 90]
	Fuel Costs	[6, 58, 90]
Environmental	Weather Conditions (Rain, Snow, Wind)	[3, 63 - 66]
	Temperature Extremes (Hot, Cold)	[58, 67 - 69]
	Soil Type (Clay, Sand, Silt, etc.)	[3, 63 - 66]
	Site Grade and Slope	[3, 63 - 66]
Human	Operator Skill and Experience	[70 - 74]
	Operator Fatigue	[70 - 74]
	Work Shifts and Rest Breaks	[70 - 74]
Technology	GPS Tracking System	[6, 60, 70, 77]
	Real-Time Monitoring System	[40, 78 - 80]
Management	Project Planning	[81 - 84]
	Supervision Quality	[81 - 84]
	Resource Allocation	[81 - 84]

2.8 Integration of Technology in Construction:

The construction industry is undergoing a profound digital transformation, reshaping traditional practices and integrating advanced technologies to revolutionize project execution and management. BIM, drones, augmented and virtual reality (AR/VR), and 3D

printing have emerged as transformative tools. BIM facilitates the creation of detailed 3D models that integrate all project information, enabling enhanced collaboration among stakeholders and fostering better decision-making. By minimizing errors and providing real-time insights, BIM improves project efficiency and ensures streamlined workflows. Drones, on the other hand, provide high-resolution aerial imaging for site surveys, inspections, and progress monitoring, offering valuable real-time data that enhances site safety and optimizes timelines. Furthermore, AR/VR technologies introduce immersive visualization techniques, aiding in project planning, training, and stakeholder communication. Similarly, 3D printing is gaining traction as a method to manufacture complex building components with remarkable efficiency, reducing both material waste and costs [91 - 94].

Smart construction practices further amplify this transformation by leveraging the Internet of Things (IoT), AI, and machine learning. IoT enables the connection of various sensors on construction sites to monitor equipment performance, worker productivity, and environmental conditions [93, 94]. When paired with AI, this data is analyzed to predict potential equipment failures, project delays, and safety hazards. Such predictive insights allow managers to adopt proactive measures, minimizing downtime and enhancing safety. Advanced algorithms optimize resource allocation and site operations, transforming construction sites into data-driven ecosystems where efficiency is maximized [94 - 97].

While these advancements present immense opportunities, they are not without challenges. High implementation costs, a shortage of skilled professionals capable of managing advanced technologies, and concerns regarding data security and privacy pose significant

hurdles. Despite these obstacles, the benefits far outweigh the challenges [93, 94]. The adoption of digital tools and automation enhances project outcomes by reducing inefficiencies, streamlining processes, and ensuring safer working environments. Globally, regions such as North America, Europe, and parts of Asia have led the way in adopting these technologies. In the Middle East, countries like Saudi Arabia, spurred by initiatives such as Vision 2030, are increasingly adopting smart construction methods to modernize their industries and strengthen their economies [39, 93 - 95, 97].

The future of technology in construction is promising, with innovations such as artificial intelligence, blockchain, and robotics poised to redefine the industry. Blockchain has the potential to enhance transparency in contract management and supply chains, while robotics can automate labor-intensive tasks with precision. These advancements not only promise to revolutionize operational efficiency but also align with global sustainability and urbanization challenges. As the construction industry continues to evolve, the strategic integration of these technologies will play a vital role in creating smarter, more sustainable, and resilient built environments. These innovations represent a transformative era for construction, where technology drives both efficiency and sustainability.

2.9 Challenges and Gaps in the Literature:

While technology integration and the application of advanced construction practices have demonstrated significant potential, several challenges and gaps in the literature warrant further investigation. One of the most prominent challenges is the inconsistent adoption and integration of new technologies across the construction sector. High costs, resistance to change, and a lack of technical expertise are critical barriers preventing widespread use.

Furthermore, limited studies address the long-term implications of technological adoption on productivity and sustainability. These gaps underline the need for comprehensive analyses that explore how technologies such as AI, BIM, and IoT can be seamlessly incorporated into construction practices. Similarly, the shortage of a skilled workforce remains a significant barrier. There is an urgent need to research effective strategies for training and education, particularly in emerging technologies. Establishing robust, industry-aligned educational programs could significantly enhance the readiness of the workforce for advanced construction demands.

Another critical concern in the literature is the growing importance of data security and privacy as construction projects increasingly depend on digital tools. Despite the critical nature of this issue, there is limited research on frameworks and best practices for protecting sensitive data while enabling collaboration. Addressing this gap is vital as construction companies adopt digital platforms. Additionally, sustainability and environmental impact are often discussed in the literature, but comprehensive analyses of the lifecycle impacts of new technologies are lacking. Further research is needed to assess the environmental benefits and drawbacks of emerging practices to create sustainable guidelines that reduce the ecological footprint of construction activities.

The regulatory and policy frameworks governing the construction industry also struggle to keep pace with technological advancements. There is a clear gap in governance and standards that can adapt to the rapidly evolving landscape of innovative construction practices. Collaborative efforts between industry stakeholders and policymakers are needed to develop flexible frameworks that prioritize safety, quality, and sustainability. At

the organizational level, cultural and structural challenges remain underexplored. Research into managing organizational change, fostering innovation, and addressing resistance to new practices could facilitate smoother transitions for construction firms adopting advanced technologies.

In the context of Saudi Arabia, these challenges take on a unique dimension. The literature lacks sufficient focus on the nation's specific economic, environmental, and cultural circumstances. More research is needed to explore how global best practices can be tailored to Saudi Arabia's construction sector, especially in alignment with Vision 2030. This initiative offers a significant opportunity for Saudi Arabia to position itself as a leader in modern construction technologies, but achieving this requires addressing the gaps in adopting innovations, workforce readiness, and policy adaptation. Comprehensive studies in these areas would not only benefit the Saudi construction industry but also serve as a valuable blueprint for other nations facing similar challenges.

2.10 Theoretical Framework:

This research is grounded in a framework that examines how technology is integrated into the construction industry in Saudi Arabia, focusing on its impact on productivity and efficiency. The framework highlights the challenges and gaps in current practices while providing a structured approach to analyzing the adoption of innovative tools like sensors and ANNs. By understanding how these technologies are evaluated and implemented, the study explores the factors influencing their acceptance and integration. It emphasizes the behaviors and decisions of construction professionals, offering insights into how these

tools can be effectively incorporated to enhance workflows and improve overall productivity in the construction sector.

These approaches address the challenges of driving change in a traditionally structured industry, offering practical strategies to facilitate the adoption of new tools and processes. When applied to the construction sector in Saudi Arabia, they provide a framework for understanding the unique challenges and opportunities aligned with the country's Vision 2030 initiative. This vision seeks to transform the construction industry by embracing advanced technologies, enhancing efficiency, and promoting sustainability. Analyzing technology adoption, efficient management principles, and organizational change within the local context supports the development of strategies that improve productivity and ensure the construction sector effectively contributes to the nation's growth and sustainability goals.

2.11 Chapter Summary:

This chapter explored how technological advancements are reshaping the construction industry, focusing on productivity factors in Saudi Arabia. Key productivity determinants include operational, environmental, economic, human, technological, and managerial factors, with tools like BIM, drones, AR/VR, 3D printing, and IoT playing a transformative role. These technologies enhance efficiency, safety, and decision-making while optimizing resource management. However, challenges such as implementation costs, limited expertise, resistance to change, and data security persist.

In the Saudi context, Vision 2030 is driving the adoption of advanced technologies and sustainable practices to boost productivity. Factors such as equipment utilization rates, operator skills, and project planning are critical, but addressing workforce readiness and sustainability constraints is essential. Green building practices and smart technologies offer the potential for balancing productivity and environmental goals, though localized strategies and regulatory frameworks are needed. To ensure impactful adoption, the construction sector must invest in training, research, and the alignment of global innovations with Saudi Arabia's unique needs, positioning the industry as a leader in sustainable development.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes the process of developing a smart platform to predict bulldozer productivity on construction sites using a dataset recorded from actual construction projects in the Eastern Province of Saudi Arabia. Insights from existing research, survey responses, GPS tracking data, and expert interviews were combined to establish a comprehensive understanding of the factors influencing bulldozer productivity. Initially, potential productivity factors were identified through a literature review, followed by surveys and interviews with industry professionals to validate these factors. Real-time data was collected from GPS trackers to capture bulldozer movement and operational details. This data underwent cleaning, segmentation, and preparation for training in an ANN model. The model was evaluated against SVM and RF models to ensure its accuracy and reliability. The validated model was then integrated into a web-based platform to enable productivity prediction for bulldozers in construction projects.

3.2 Research Design

This study adopts a structured research approach, illustrated in [Figure 3.1](#), to build a predictive model by integrating data from multiple sources. The process begins with a survey of 87 industry professionals in Saudi Arabia to identify key productivity factors. Interviews with project managers, site engineers, and bulldozer operators were conducted to gain a deeper understanding of how these factors affect operations in practice. GPS data

was then collected from three construction sites in Dammam, Qatif, and East Dhahran, capturing real-time bulldozer operations. This data included details such as movement patterns, fuel consumption, operator skill, and weather conditions. This comprehensive and structured approach resulted in a robust dataset, which was subsequently used to develop and test the ANN model for predicting bulldozer productivity.

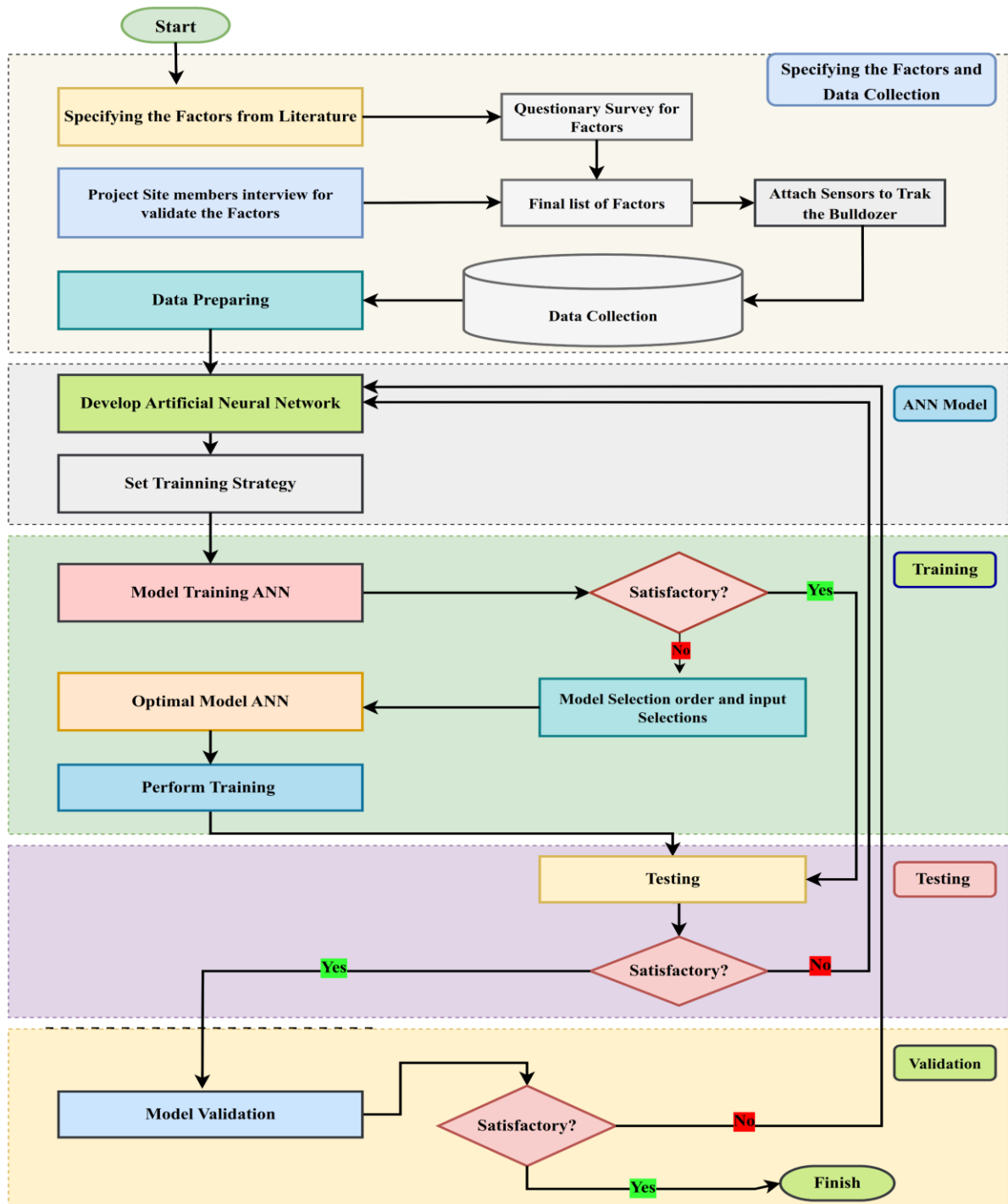


Figure 3.1: Framework for Developing a Predictive Model

3.3 Identifying Key Productivity Factors from the Literature

The initial step in developing a predictive model for bulldozer productivity involved identifying the critical factors influencing productivity on construction sites. This began with an extensive review of existing literature on heavy equipment productivity, operational efficiency, and construction site management. By analyzing over 200 sources, as shown in [Figure 3.2: Publication Trend by Year \(1950-2024\)](#), insights were drawn from peer-reviewed journal articles, industry reports, and case studies. This thorough examination enabled the identification and compilation of a robust set of factors recognized by researchers and industry experts as key to productivity, forming a solid foundation for the model's development.

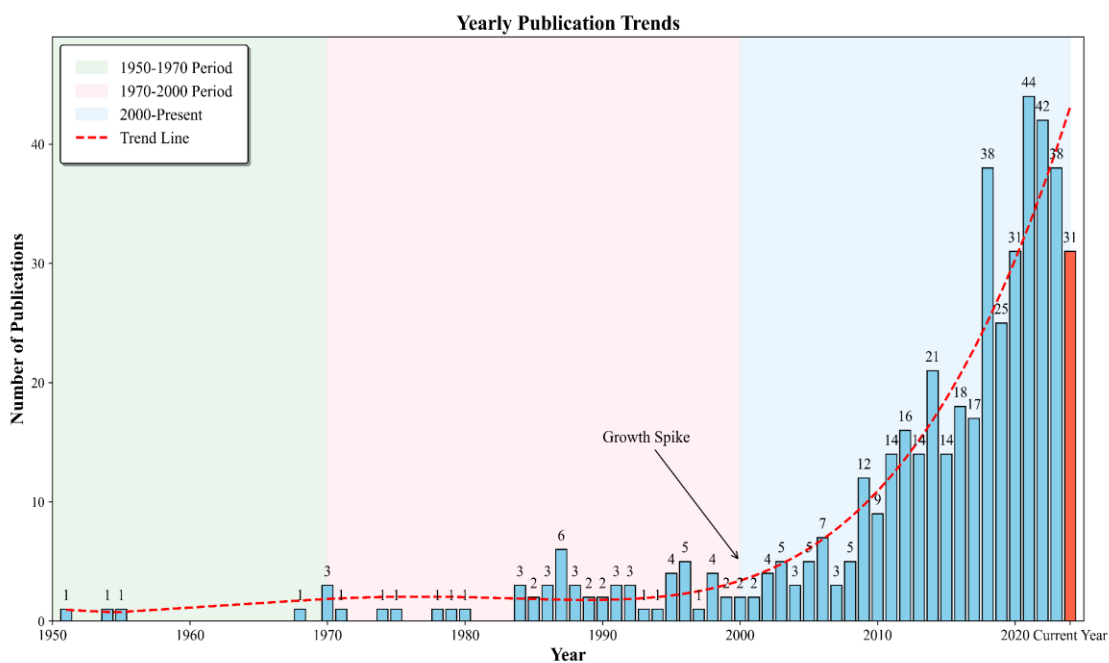


Figure 3.2: Publication Trend by Year (1950-2024)

From this review, several categories of factors were identified as significant for construction productivity. Equipment-related factors, such as machine age, utilization rate, and maintenance history, emerged as critical due to their influence on performance and reliability. Operator-related factors, including skill level, training, and fatigue, were also highlighted for their direct impact on equipment efficiency. Environmental conditions, such as temperature, soil type, and weather patterns, frequently appeared in studies focusing on variable climates, emphasizing their role in machine effectiveness and safety. To ensure relevance to Saudi Arabia's specific conditions, particular attention was given to studies conducted in similar hot, arid climates, which underscored factors like fuel consumption and the effects of heat on equipment. This comprehensive list of factors provided a robust foundation for further investigation and informed the design of a survey aimed at gathering practical insights from industry professionals.

Table 3.1: Key Productivity Factors Identified from the Literature

Category	Factors	Frequency
Operational	Equipment Age and Condition	12
	Blade Size and Type	11
	Load Size	9
	Hauling Distance	13
	Equipment Utilization Rate	11
Economic	Equipment Rental vs. Ownership Costs	12
	Maintenance costs	14
	Fuel Costs	13
Environmental	Weather Conditions (Rain, Snow, Wind)	16
	Temperature Extremes (Hot, Cold)	13
	Soil Type (Clay, Sand, Silt, etc.)	18
	Site Grade and Slope	15
Human	Operator Skill and Experience	17
	Operator Fatigue	8
	Work Shifts and Rest Breaks	9
Technology	GPS Tracking System	13
	Real-Time Monitoring System	12
Management	Project Planning	14
	Supervision Quality	11
	Resource Allocation	10

3.4 Data Collection

Data collection was conducted on-site at three active construction projects in the Eastern Region of Saudi Arabia: Dammam, Qatif, and East Dhahran as shown in Figure 3.3. Each project site offered distinct environmental and operational characteristics, which helped create a comprehensive dataset for analyzing bulldozer productivity under varied conditions.



Figure 3.3: Aerial View of Construction Project Locations in Saudi Arabia

3.4.1 Tracking Bulldozer Movement with GPS

To accurately monitor bulldozer operations at each site, GPS tracking devices were installed on selected bulldozers to capture real-time movement data. The GPS trackers were configured to log positional data at frequent intervals, providing precise information on each bulldozer's route, distance traveled, operational speed, and time spent on different activities. This tracking method allowed for detailed insight into each bulldozer's daily

productivity patterns, capturing both idle times and active work periods. The screenshots displayed in Figure 3.4 show GPS tracking data, with clear visualizations of movement paths and operational statistics. The GPS devices were securely mounted and calibrated before data collection to ensure accuracy. As part of the setup process, GPS devices logged coordinates at frequent intervals, enabling continuous and high-resolution tracking of bulldozer movements across the project site. These GPS visualizations also facilitated path segmentation and activity detection. Significant turning points were extracted from the recorded tracks, marking transitions between distinct activities such as loading, transporting, and bulldozing. The GPS tracking data, illustrated in Figure 3.4, highlights segmented paths with color-coded routes representing variations in speed and activity durations, enabling a deeper analysis of operational cycles and productivity phases.

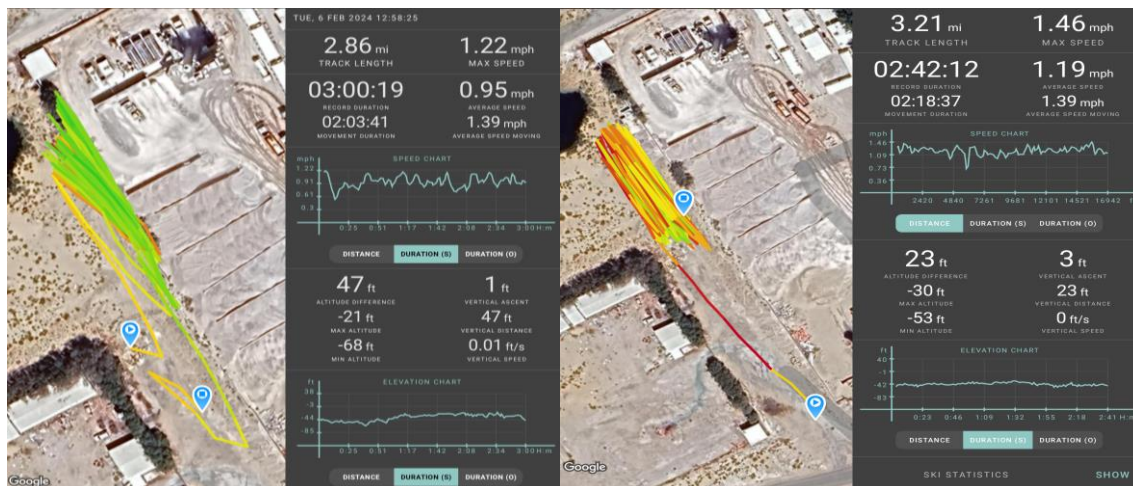


Figure 3.4: GPS tracking screenshots, bulldozer movement paths with color-coded.

3.4.2 Site Condition and Operational Data Recording

Alongside GPS tracking, a structured record sheet, [Table 3.2](#) was used to document a variety of site-specific and operational factors that could influence bulldozer productivity. These included:

- i. **Fuel Consumption:** The table indicates daily fuel usage, ranging from 19 to 26 liters per hour, reflecting variations in operational intensity and site conditions. Higher fuel consumption is often associated with more demanding tasks or challenging terrains, such as operating on mixed soil compositions or during extended working hours. For instance, Day 2, which recorded the highest fuel consumption (26 L/hour), involved operations under sunny conditions with an experienced operator in the afternoon, suggesting optimal conditions for high-intensity work.
- ii. **Weather Conditions:** Weather patterns such as sunny, rainy, cloudy, and windy days were recorded. These environmental conditions play a significant role in shaping bulldozer productivity. For example, operations on Day 3 under rainy conditions likely posed challenges such as reduced traction and increased fuel consumption (22 L/hour). Conversely, sunny days like Day 1 and Day 2 facilitated higher productivity, as evidenced by higher fuel usage and smoother operations.
- iii. **Operator Skill and Experience:** The table categorizes operators' skill levels into Poor, Average, and Very Good, highlighting the human factor's role in productivity. Experienced operators (e.g., on Days 2 and 7) demonstrated higher efficiency, as reflected in their ability to handle equipment under challenging conditions with

minimal productivity losses. In contrast, days with Poor operator skill levels (e.g., Day 5) corresponded with lower fuel consumption and likely suboptimal performance.

- iv. **Soil and Site Conditions:** The soil type (e.g., sand or a mix of clay and sand) and site characteristics significantly influenced productivity. Bulldozers working on mixed soils, such as on Days 1 and 6, often required more power and fuel to navigate, whereas operations on sandy terrains (e.g., Day 4) were associated with slightly lower productivity due to challenges like reduced traction.
- v. **Blade Types and Equipment Configuration:** The table documents the blade types used (U-Blade and S-Blade) and their impact on operations. U-Blades, used on Days 1, 2, and 6, are designed for moving larger volumes of material, making them suitable for intensive tasks. Conversely, S-Blades, employed on Days 4 and 5, are optimized for precision tasks such as grading, explaining their use in scenarios requiring finer control.

Table 3.2: Site Condition and Operational Data Recording for Bulldozer Operations

Day	Site Condition and Operational Data Recording						
	Weather Conditions	Fuel Consumption (Liters/Hour s)	Operator Skill and Experience	Soil and Site Conditions	Time of Operation	Dozer Type	Blade Type
Day 1	Sunny	23.5	Average	Mix(Clay and Sand)	Morning	D155 A	U- Blade
Day 2	Sunny	26	Very Good	Mix(Clay and Sand)	Afternoon	D155 A	U- Blade
Day 3	Rainy	22	Average	Mix(Clay and Sand)	Morning	D155 A	U- Blade
Day 4	Cloudy	25	Average	Sand	Afternoon	D155 A	S- Blade
Day 5	Sunny	20	Poor	Sand	Evening	D155 A	S- Blade
Day 6	Cloudy	23	Average	Mix(Clay and Sand)	Morning	D155 A	U- Blade
Day 7	Windy	25	Very Good	Sand	Afternoon	D155 A	S- Blade
Day 8	Sunny	19	Average	Mix(Clay and Sand)	Afternoon	D155 A	U- Blade
Day 9	Cloudy	24	Poor	Mix(Clay and Sand)	Evening	D155 A	U- Blade
Day 10	Sunny	22	Average	Sand	Morning	D155 A	S- Blade

3.4.3 Integration of GPS and Site Condition Data

GPS data was synchronized with manually recorded site information to develop a comprehensive dataset capturing the full scope of bulldozer operations. By aligning GPS time stamps with fuel consumption records, weather conditions, and site-specific observations, each data point was connected to the bulldozer's operational state at that moment. This integrated approach revealed movement patterns and demonstrated how various factors influenced performance. For example, higher fuel consumption was linked to operations on inclines or compacted ground, while slower speeds were associated with adverse weather conditions. This holistic perspective provided valuable insights into the context of each work cycle, offering opportunities to optimize operations and enhance efficiency on construction sites.

The detailed tracking of bulldozer operations, as depicted in [Figure 3.5](#), captures each moment in time and location as a dot, creating a comprehensive visualization of daily activities. This representation goes beyond mapping movements, integrating GPS coordinates with real-world site activities to provide a dynamic understanding of operations. By blending spatial data with other collected insights, the analysis identifies the key factors driving productivity. The focus is not just on the paths taken but on understanding the reasons behind these movements and evaluating how effectively the equipment performed at each point.

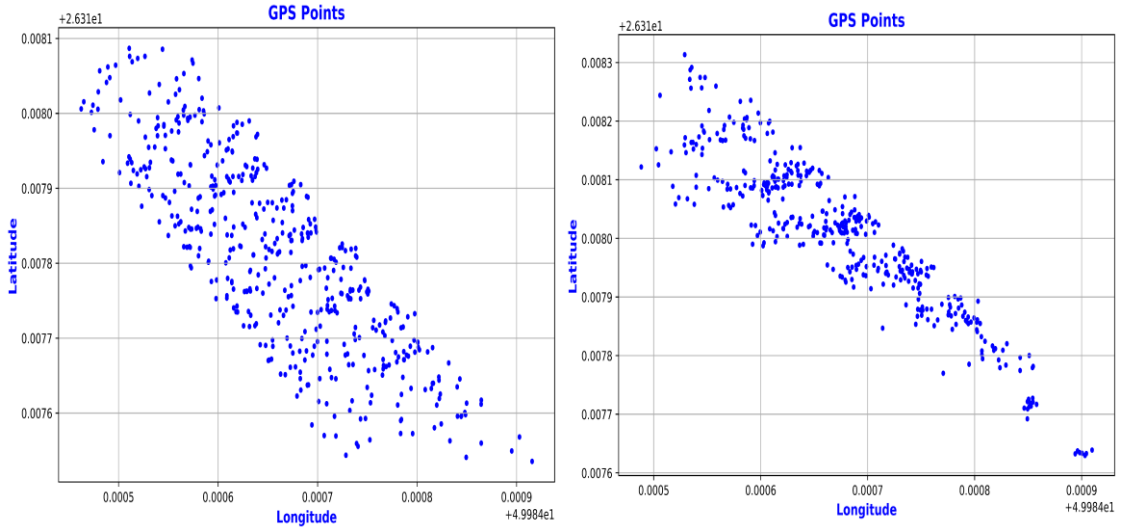


Figure 3.5: GPS Tracked points

3.5 Data Extraction and Preparation

After collecting raw GPS tracking data from bulldozer operations on-site, the next step involved processing this data to make it usable for productivity analysis and modeling. This phase included transforming raw GPS coordinates into meaningful paths and metrics, segmenting operational cycles, and cleaning the dataset to remove any noise or outliers. The goal was to create a refined dataset that accurately represented each phase of bulldozer operations, enabling precise analysis and model training.

3.5.1 GPS Data Transformation and Path Extraction

The process began with raw GPS data, a dense collection of latitude and longitude points tracing the movements of bulldozers across construction sites. These seemingly random coordinates were transformed into a structured narrative of operational efficiency. Using advanced segmentation techniques, the continuous GPS trails were divided into distinct operational paths, each representing a specific phase of the bulldozer's workday. Turning

points were identified to signify task transitions, while critical metrics such as path distance and speed were calculated. This analysis differentiated productive activities, like earthmoving, from less efficient periods, such as idling. The result was a rich dataset offering actionable insights, revealing not just where the machinery went, but how effectively it performed. By converting raw geographic data into a detailed performance roadmap, this approach provided construction managers with a nuanced understanding of equipment operations. It moves beyond documentation to deliver a strategic framework for optimizing workflows, ensuring every minute of bulldozer operation contributes meaningfully to project productivity.

Turning Points and Path Segmentation: The segmentation process focused on identifying critical turning points in the GPS data moments when the bulldozer changed direction or initiated a new task, marking a transition in its workflow. By isolating these moments, the continuous GPS trail was divided into distinct paths, each representing a specific phase of the bulldozer's operations. This method offered a structured view of its activities, revealing patterns and performance across the project site. As shown in [Figure 3.6](#), the segmented paths are visually distinguished by different colors, enabling a clear depiction of movement cycles. This visualization not only underscores the variety of tasks performed but also provides valuable insights into the productivity of the bulldozer during each operational phase.

Path Metrics Calculation: Following the segmentation of paths, key metrics such as total path distance, average speed, and operational duration were calculated for each segment. These metrics offered critical insights into the bulldozer's performance, shedding light on how it operated under varying site conditions. By correlating these metrics with the

segmented paths, it became possible to differentiate between productive activities like bulldozing or hauling, and less productive ones, such as idling or transitioning between tasks. This integration of segmentation and metric analysis established a robust framework for evaluating productivity, delivering actionable insights for enhancing equipment management, and optimizing construction operations.

By employing these techniques, the raw GPS data was transformed into a treasure trove of actionable insights. What initially appeared as a sequence of geographic coordinates evolved into a robust tool for evaluating productivity and identifying opportunities for operational improvement. This systematic approach to dissecting and analyzing data not only proved invaluable for this project but also established a replicable framework for extracting meaningful insights from GPS-tracked activities across the construction sector. In effect, this process converted streams of raw data into a strategic blueprint for driving smarter, more efficient construction practices.

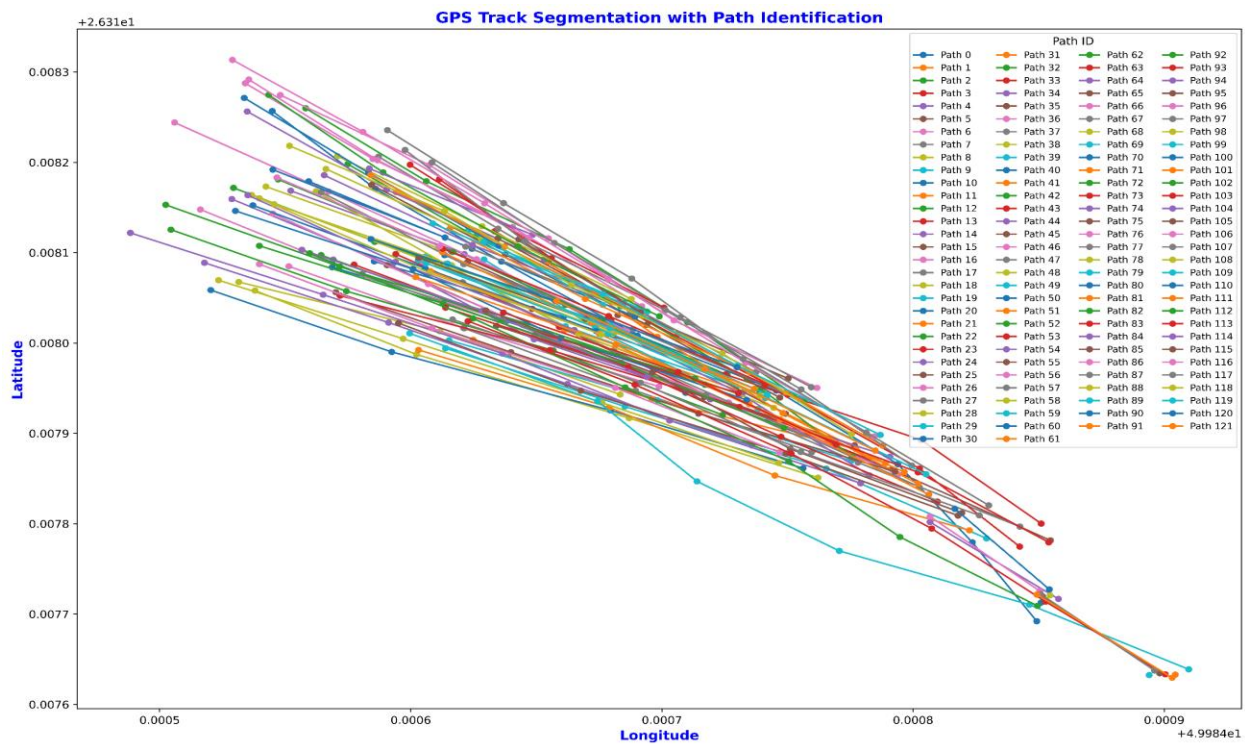
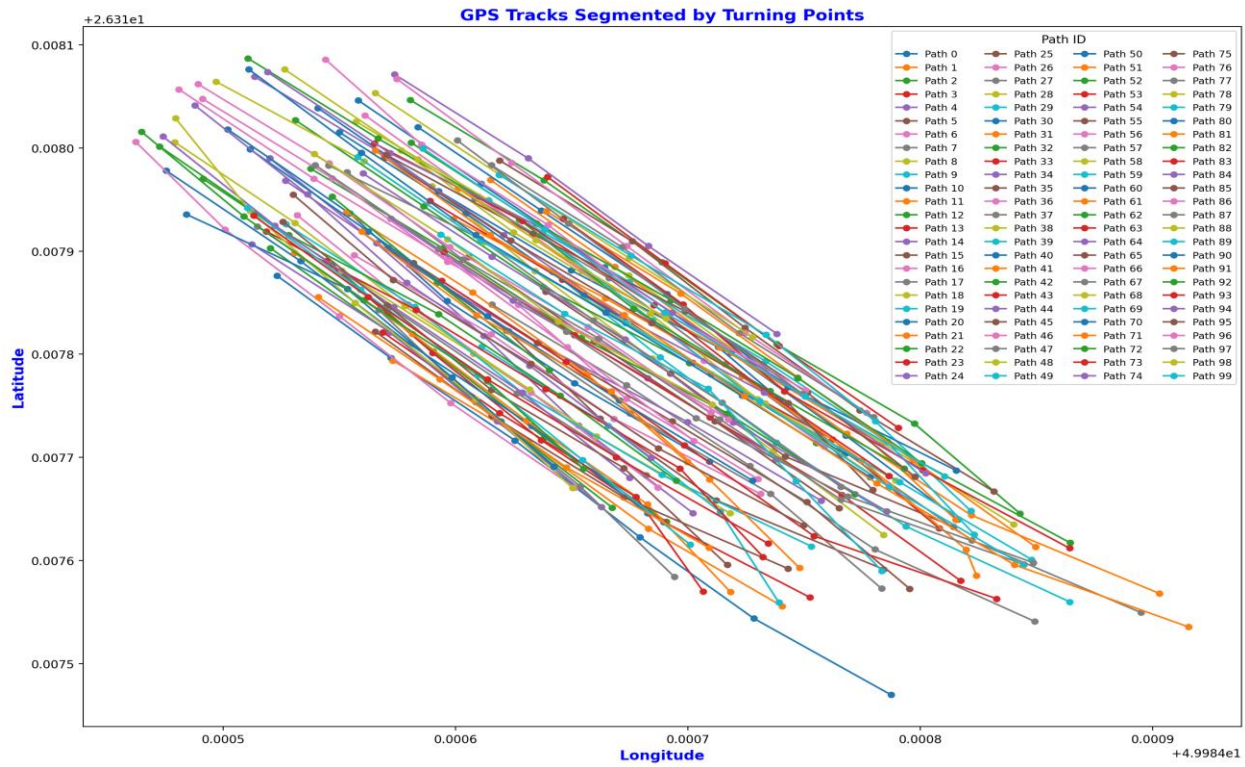


Figure 3.6: GPS Paths Segmented by Turning Points

3.5.2 Data Cleaning and Outlier Removal

After transforming and segmenting the GPS data, the dataset underwent a meticulous cleaning process to ensure accuracy and reliability. Outliers, such as unusually high or low speeds and excessively short or long cycles, were identified and removed. These anomalies, often caused by GPS inaccuracies or non-operational movements on-site, were filtered out to maintain the integrity of the analysis. By addressing these inconsistencies, the dataset was refined to provide a clear and precise foundation for evaluating bulldozer productivity.

Box Plot Analysis for Outlier Detection: A box plot, also known as a whisker plot, is a visualization technique used to depict the distribution of a dataset through its quartiles, making it a valuable tool for identifying and displaying outliers. This graphical representation is built around several key elements: the median (Q2), which indicates the dataset's central value; the interquartile range (IQR), spanning the middle 50% of the data from the first quartile (Q1) to the third quartile (Q3); and the whiskers, which extend to the smallest and largest data points within 1.5 times the IQR from Q1 and Q3. Data points that fall outside these whisker boundaries are highlighted as outliers, often marked with distinct symbols or dots. This method effectively summarizes the dataset's variability while highlighting deviations that could influence statistical analyses or model outcomes [98, 99].

The process of using a box plot for outlier detection involves calculating the quartiles and IQR, then setting thresholds for the lower and upper bounds using the formulas $\text{Lower Bound} = Q1 - 1.5 \times \text{IQR}$ and $\text{Upper Bound} = Q3 + 1.5 \times \text{IQR}$. Data points that fall outside these bounds are classified as outliers. This analysis is significant for ensuring

data integrity by detecting anomalies, improving modeling by removing data points that could distort results and providing a concise and insightful summary of the data distribution. The visual simplicity and statistical robustness of the box plot make it a versatile tool in exploratory data analysis and preprocessing [99, 100].

3.5.3 Data Analysis

3.5.3.1 Normalization Boxplots:

Normalization boxplots are a variation of the traditional box plot designed to evaluate and visualize the effects of data normalization techniques. Normalization is a preprocessing step that scales data to fit within a specified range or distribution, ensuring consistency and comparability between features with different scales or units. By using boxplots, the impact of normalization on the dataset's distribution can be effectively illustrated. In normalization boxplots, data is typically presented before and after applying normalization techniques, such as Min-Max scaling, Z-score normalization, or Robust scaling. The boxplot highlights changes in the central tendency (median), variability (interquartile range), and the presence of outliers. This visualization is particularly useful for detecting anomalies, ensuring uniformity in feature scales, and validating that normalization does not distort the underlying data distribution. In machine learning workflows, consistent data scales often improve the performance and convergence of algorithms sensitive to feature magnitudes, such as gradient-based methods. By providing a clear and concise visualization, normalization boxplots help practitioners evaluate whether their normalization approach is appropriate and effective for their specific dataset [101 - 103].

3.7.3.2 Pair-plot of Features:

A pair plot is a versatile visualization tool that combines scatter plots and histograms (or density plots) to provide a comprehensive view of relationships between features in a dataset. It generates scatter plots for each pair of numerical features, allowing users to explore pairwise interactions, identify patterns, detect correlations or clusters, and spot anomalies. Along the diagonal, it includes histograms or kernel density estimates (KDE) to display the distribution of individual features. This combination makes pair plots a powerful method for exploratory data analysis (EDA), offering a compact and detailed overview of feature interactions [104].

Scatter plots in a pair plot reveal the relationships between numerical features, highlighting potential linear or non-linear correlations and uncovering patterns that may inform data modeling. The diagonal elements, through histograms or KDE plots, provide insights into the data's spread and underlying distribution. Additionally, pair plots can incorporate color encoding to distinguish between different categories or groups in the dataset, enhancing their utility for understanding feature distributions across categorical variables. Pair-plots help researchers and practitioners better understand how features relate to one another. This insight can guide feature selection, engineering, and preprocessing decisions, ensuring more robust and efficient modeling. The ability to visualize these relationships in a single, comprehensive plot makes pair plots an invaluable tool for data exploration and analysis [104, 105].

3.7.3.3 Correlation Heatmap:

A correlation heatmap is a powerful visualization tool that represents the pairwise relationships between variables in a dataset through a color-coded matrix, making it easy to identify patterns, dependencies, and redundancies. Correlation, measured on a scale from -1 to 1, indicates the degree of linear association between variables, with -1 signifying a perfect negative correlation, 1 a perfect positive correlation, and 0 no correlation. The heatmap's cells, color-coded to reflect the strength and direction of correlations, enable quick identification of strong positive relationships (often shown in warm colors like red) or strong negative relationships (shown in cool colors like blue), while neutral tones signify weak or no correlation. Annotated heatmaps, which include correlation values within the cells, offer even more precision. Commonly used in EDA, correlation heatmaps help detect multicollinearity, a condition where independent variables are highly correlated, which can impair the performance of machine learning models. By identifying these relationships, analysts can make informed decisions about feature selection or transformation to enhance model robustness. Widely applied across domains such as finance, biology, and social sciences, heatmaps facilitate a deeper understanding of variable interactions and streamline the data preparation process. Tools like Python's Seaborn library simplify the creation of correlation heatmaps, making them an indispensable part of modern data analysis workflows [106]

3.5.4 Feature Engineering and Final Dataset Preparation

The data analysis process began with a detailed examination of the dataset, aimed at uncovering and confirming correlations among essential variables. Key relationships

emerged, such as those between speed, cycle time, and distance traveled, which are crucial for understanding bulldozer productivity. Building on these insights, feature engineering was applied to transform the raw data into meaningful input variables. Metrics such as average speed, cycle time, distance traveled, and actual productivity (measured in cubic meters per hour) were calculated for each segmented path. These engineered features offered a structured and actionable perspective on operational performance, ensuring the data accurately represented real-world bulldozer activities and was ready for further analysis and modeling.

With the transformed dataset, a solid foundation was laid for developing a predictive productivity model. The engineered features served as inputs, while measured productivity acted as the target variable, creating a dataset ready for training machine learning algorithms. The next steps involved selecting appropriate algorithms, fine-tuning their parameters, and rigorously evaluating their performance. The objective was to build a model capable of accurately predicting bulldozer productivity across various site conditions. By utilizing insights from data analysis and feature engineering, the modeling process was designed to ensure a framework that is both robust and adaptable to real-world applications. This seamless integration of data preparation and advanced modeling techniques paved the way for practical and impactful advancements in analyzing and optimizing construction productivity.

3.6 Model Development and Validation

The findings from the data analysis phase seamlessly transitioned into the model development stage, forming a strong foundation for building predictive models. By

conducting exploratory data analysis, key relationships between factors such as cycle time, speed, distance traveled, and productivity were uncovered, providing a nuanced understanding of the variables driving bulldozer performance. This process not only highlighted the most influential factors but also guided the creation of meaningful metrics through feature engineering, transforming raw data into actionable insights. Metrics like segmented paths, operational cycles, and calculated productivity rates were meticulously crafted to capture the complexities of real-world bulldozer operations. With this enriched dataset, the next step focused on developing models capable of predicting productivity with precision. The ANN was chosen as the primary predictive tool for its ability to capture nonlinear relationships, while SVM and RF models served as benchmarks for validation. This smooth progression from analysis to modeling ensured that the insights gained were fully leveraged, enabling the creation of robust and accurate models tailored to the unique challenges of construction site operations.

3.6.1 Artificial Neural Network (ANN) Model

The ANN model was developed as the central predictive tool to estimate bulldozer productivity by leveraging the comprehensive dataset of segmented paths and engineered metrics. The ANN model was specifically designed to capture the intricate relationships between key factors such as cycle time, average distance, operator skill, and environmental conditions variables that collectively influence bulldozer productivity. The flexibility and learning capacity of the ANN made it an ideal choice for modeling the complex, non-linear patterns inherent in construction operations.

The architecture of the ANN model consisted of an input layer to receive the data, multiple hidden layers to process it, and an output layer to deliver predictions. The hidden layers, composed of several neurons, used the ReLU activation function to introduce non-linearity and enhance the learning of complex patterns. This structure allowed the ANN to effectively model diverse scenarios, such as varying soil types or weather conditions, that impact productivity. The output layer employed a linear activation function to provide continuous predictions, aligning with the numerical nature of the target variable—measured productivity in cubic meters per hour. During the training process, the model optimized its neuron weights using backpropagation, a key algorithm that adjusts weights to minimize prediction errors. The MSE loss function was employed to measure performance, ensuring that the model progressively improved accuracy with each epoch of training.

Hyperparameter optimization played a crucial role in enhancing the ANN's performance. A grid search was conducted to fine-tune parameters such as the number of hidden layers, learning rate, and batch size. To ensure the model generalized well to unseen data, cross-validation techniques were employed, minimizing the risk of overfitting. Once trained, the ANN was validated using a separate test dataset to evaluate its prediction accuracy. This validation phase was pivotal in assessing the model's reliability before benchmarking it against other predictive models like SVM and Random Forest. By systematically designing, training, and optimizing the ANN, the study laid the groundwork for an accurate and robust predictive model, capable of offering actionable insights into bulldozer productivity under diverse operational conditions.

3.6.2 Support Vector Machine (SVM) Model

To assess the performance of the ANN model, an SVM with a regression approach (SVR) was developed as a comparative model. The SVM model offered a distinct methodology for productivity prediction by focusing on identifying a hyperplane in a higher-dimensional space that best fits the data. This approach made the SVM model particularly effective for regression tasks, allowing it to act as a benchmark for the ANN model. SVM functions by finding the optimal hyperplane that maximally separates data points in a high-dimensional space.

The SVM model utilized the RBF kernel to address the inherent complexity of productivity prediction. The RBF kernel maps input data into a higher-dimensional space, enabling the model to capture non-linear relationships between features—similar to the functionality of the ANN. The kernel's hyperparameter, γ (gamma), played a pivotal role in this process by determining the range of influence for each data point. Adjusting γ allowed the model to strike a balance between closely fitting the training data and generalizing effectively to unseen data. Another key component of the SVM model was the regularization parameter (C), which controlled the trade-off between achieving minimal error on the training data and maintaining a simpler, more generalizable model. A high value of C encouraged tighter fitting to the training data, while a lower value prioritized generalization.

Once the SVM model was trained, it was tested on the same dataset used to evaluate the ANN model. This comparative analysis offered insights into the strengths and limitations of each model. The SVM model's performance, particularly its ability to handle non-linear relationships, provided a valuable benchmark for validating the effectiveness of the ANN

model. By comparing the results from both models, the study demonstrated the superior ability of the ANN model to capture complex, multi-dimensional interactions in productivity prediction, reinforcing its position as the primary predictive tool in this research.

3.6.3 Random Forest (RF) Model

RF was chosen as the third model in the comparative analysis due to its robust ensemble learning approach, which aggregates the predictions of multiple decision trees to enhance accuracy and stability. This method offered a valuable alternative to the ANN and SVM models by leveraging the inherent diversity and collective decision-making of ensemble methods.

The RF model employs a technique known as bagging (Bootstrap Aggregating), which involves building multiple decision trees from random subsets of the training data. Each tree in the ensemble operates independently, generating its prediction based on its specific training sample. The final output of the RF model is then calculated as the average of all individual tree predictions, a process that significantly reduces the likelihood of overfitting. This ensemble approach creates a robust model that is less sensitive to noise or outliers in the dataset, offering consistent performance across different scenarios.

A key feature of RF is its introduction of randomness at the feature level. For each split in a tree, the model randomly selects a subset of features rather than considering all available features. This randomness increases the diversity among individual trees, enabling the ensemble to capture different patterns and nuances within the data. By combining these

diverse perspectives, the RF model becomes more effective at generalizing to unseen data. After training the RF model on the dataset, it was evaluated on the test set, and its performance was compared with the ANN and SVM models. This comparative analysis provided a comprehensive validation process, demonstrating how the ensemble approach complements the strengths of the other models and contributes to a holistic understanding of bulldozer productivity prediction.

3.6.4 Model Evaluation Metrics

To evaluate the performance of each model, several standard regression metrics were used to measure accuracy and reliability.

- i. **Mean Absolute Error (MAE):** The *MAE* calculates the average magnitude of prediction errors, offering a straightforward interpretation of how close predictions are to actual values [107, 108]. It is calculated as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \text{Eq. 3.2 [108]}$$

where y_i is the actual productivity and $\hat{y}_i$ is the predicted productivity. A lower MAE value indicates better accuracy.

- ii. **Root Mean Squared Error (RMSE):** The *RMSE* measures the square root of the average squared errors, giving more weight to larger errors [108, 109]. It is computed as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{Eq. 3.3[108, 109]}$$

Lower *RMSE* values signify higher model precision and fewer large deviations in predictions.

- iii. **Coefficient of Determination (R^2):** The R^2 assesses how well the model explains the variability in productivity [110]. It is given by:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{Eq. 3.4[110]}$$

Where $\bar{y}$ is the mean of actual productivity values. An R^2 value closer to 1 indicates a high degree of explanation, with higher values demonstrating a better model fit.

Using these metrics, the performance of the ANN, SVM, and RF models was thoroughly assessed to identify the most effective model for bulldozer productivity prediction.

3.6.5 Web-Based Platform Development

Following the successful validation of the predictive models, a web-based platform was developed to operationalize the ANN model and make its predictive capabilities accessible to end-users. This platform is designed specifically for construction project managers and equipment operators, enabling them to leverage cutting-edge predictive analytics without requiring in-depth technical expertise.

The platform's user interface was created to be intuitive and straightforward, prioritizing ease of use. Site-specific data, such as average speed, cycle time, soil type, and operator skill, can be entered into structured input fields, including dropdown menus and predefined options. This approach minimizes the chances of data entry errors and ensures consistency across different users. The interface is tailored to meet the practical needs of construction sites, ensuring that even non-technical users can quickly input data and obtain actionable insights.

In the backend, the trained ANN model was serialized and integrated seamlessly into the platform. This integration allows the platform to process user inputs in real-time, generating immediate productivity predictions. By embedding the ANN model into the platform's architecture, users can access the predictive power of the model directly through the platform, eliminating the need for manual calculations or external tools. To build trust and promote informed decision-making, the platform also provides key performance metrics such as MAE, RMSE, and R^2 values. These metrics give users insight into the reliability and accuracy of the predictions, ensuring that the platform is both transparent and effective in aiding productivity optimization efforts.

This web-based solution bridges the gap between advanced predictive analytics and practical field applications, empowering stakeholders to make data-driven decisions that improve bulldozer productivity and overall project efficiency.

3.7 Chapter Summary

This chapter presented a comprehensive methodology for developing a predictive model to estimate bulldozer productivity in construction projects. The process began with Finalizing the factors that affect the productivity of bulldozers and a detailed data collection phase, where GPS tracking devices were deployed on bulldozers across three active construction sites in the Eastern Region of Saudi Arabia. This enabled the recording of real-time operational data, while additional site-specific factors, such as fuel consumption, weather, operator skill, and soil conditions, were documented through structured record sheets. These measures ensured that the dataset accurately represented real-world conditions. After data collection, the raw GPS coordinates were transformed into segmented paths, identifying distinct operational cycles and calculating metrics like cycle time, speed, and distance. This cleaned and refined dataset provided a solid foundation for model training. The next step involved developing an ANN model as the primary predictive tool, with the model architecture carefully optimized to capture complex relationships in the data. For validation, the ANN model's performance was compared with SVM and RF models, using evaluation metrics such as *MAE*, *RMSE*, and the R^2 to assess predictive accuracy. Finally, a web-based platform was developed, incorporating the ANN model to enable construction managers and operators to input operational data and receive immediate productivity predictions, thus supporting real-time decision-making on-site. This structured methodology sets the stage for the next chapter, which will present and analyze the results of each model, demonstrating the practical utility of the developed platform in enhancing productivity and efficiency in construction operations.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the results of the survey and final RII ranking and will the predictive models developed to estimate bulldozer productivity, including the ANN, SVM, and RF models. The purpose of this chapter is to assess the effectiveness of each model based on several performance metrics, such as MAE, RMSE, and the R^2 . These metrics provide insight into the accuracy and reliability of the models when applied to the test dataset, allowing for a direct comparison of each model's predictive capabilities.

The chapter begins with a summary of the training and testing processes, including the dataset split and the evaluation setup, to establish the context for model performance analysis. Each model's results are then presented and analyzed, highlighting specific strengths and potential limitations in capturing productivity patterns. Following this, the chapter provides a comparative discussion of the models, identifying which model best aligns with the data characteristics and is most suitable for real-world applications in construction productivity management.

Finally, the effectiveness of the web-based platform, which integrates the trained ANN model for real-time productivity predictions, is discussed in terms of its practical utility for end-users. The insights gained from this analysis offer valuable contributions to the field of construction productivity modeling and demonstrate the potential for data-driven

decision-making in enhancing operational efficiency. This results discussion not only validates the developed model but also provides a foundation for future work and further improvements in predictive analytics for construction productivity.

4.2 Survey of Industry Professionals

After identifying the preliminary list of factors from the literature, the findings were validated through a survey conducted among industry professionals in Saudi Arabia. A structured survey was developed to collect feedback from 87 construction experts, ensuring practical insights into the relative importance of each factor. The participants included project managers with 5 to 15 years of experience, site engineers with over 5 years of field expertise, and equipment operators with more than 10 years of hands-on experience in bulldozer operations. This diverse group brought a wealth of practical knowledge and experience, ensuring that the survey results were grounded in real-world conditions and accurately reflected the factors influencing bulldozer productivity on construction sites.

In the survey, participants were asked to rank each factor's impact on bulldozer productivity on a scale, allowing us to calculate a Relative Importance Index (RII) for each factor. The RII is a commonly used metric in construction research, providing a quantitative measure of how important each factor is perceived to be by professionals in the field. The RII was calculated using the following formula:

$$\text{RII} = \frac{\sum W}{A \times N} \quad \text{Eq. 4.1 [108]}$$

Where:

- **W** represents the weight assigned by each respondent,
- **A** is the highest possible weight (in this case, 5), and
- **N** is the total number of respondents.

The RII scores presented in [Table 4.1](#) reveal a clear ranking of factors based on their perceived impact on bulldozer productivity, providing both theoretical validation and practical insights. Soil Type, ranked highest with an RII of 1, emphasizes the significance of environmental conditions, while operational factors like Blade Size and Type, Load Size, and Hauling Distance follow closely with high scores (0.91, 0.87, and 0.84, respectively), underscoring their critical role in determining productivity on construction sites. Notably, factors such as Project Planning (0.76) and Work Shifts and Rest Breaks (0.71) highlight the importance of strategic planning and labor management in maintaining efficient operations. The inclusion of Real-Time Monitoring System (0.67) and Temperature Extremes (0.6) further suggests that technological integration and environmental awareness are vital components of productivity optimization. Lower-ranked factors, like Equipment Utilization Rate (0.13) and Operator Fatigue (0.11), indicate elements that may have a more indirect or situational influence on productivity. Collectively, this ranking ensures that the productivity model is rooted in factors that are both theoretically validated and practically relevant, providing a robust foundation for applications on construction sites. By focusing on high-RII factors, this model offers a targeted approach to enhancing operational efficiency and aligns closely with the realities of construction productivity challenges.

Table 4.1: Survey Results RII for Each Factor

Rank	Factors	Number of Responses	Average Score	RII
1	Soil Type (Clay, Sand, Silt, etc.)	87	3.24	1
2	Blade Size and Type	87	3.2	0.91
3	Load Size	87	3.17	0.87
4	Hauling Distance	87	3.16	0.84
5	Project Planning	87	3.12	0.76
6	Work Shifts and Rest Breaks	87	3.09	0.71
7	Real-Time Monitoring System	87	3.07	0.67
8	Temperature Extremes (Hot, Cold)	87	3.03	0.6
9	Supervision Quality	87	3.02	0.58
10	Resource Allocation	87	2.99	0.51
11	Site Grade and Slope	87	2.97	0.47
12	GPS Tracking System	87	2.97	0.47
13	Fuel Costs	87	2.95	0.44
14	Operator Skill and Experience	87	2.94	0.42
15	Maintenance costs	87	2.94	0.42
16	Weather Conditions (Rain, Snow, Wind)	87	2.92	0.38
17	Equipment Rental vs. Ownership Costs	87	2.85	0.24
18	Equipment Age and Condition	87	2.83	0.2
19	Equipment Utilization Rate	87	2.79	0.13
20	Operator Fatigue	87	2.72	0.11

4.3 Validation of Productivity Factors with Project Experts

While the survey provided valuable quantitative insights, follow-up interviews were conducted with key project experts to fully contextualize each factor within Saudi construction projects. These interviews involved project managers, site engineers, and bulldozer operators working on active construction sites in Dammam, Qatif, and East Dhahran. This qualitative approach allowed for a deeper understanding of the day-to-day

realities of bulldozer operations and the nuanced ways in which each factor influences productivity. Project managers and site engineers elaborated on strategic elements like project planning, resource allocation, and supervision quality, providing concrete examples of their impact on bulldozer performance, particularly in complex or high-demand projects. Bulldozer operators contributed a practical perspective, detailing how factors such as skill, fatigue, and soil conditions affect their efficiency. They also highlighted challenges unique to Saudi construction sites, such as operating on sandy or compacted soil and maintaining equipment performance in extreme temperatures. These insights enriched the study, offering a comprehensive view of productivity influences across different operational levels.

This qualitative validation was essential for capturing the intricate dynamics of bulldozer productivity influenced by environmental and operational conditions. Observations at the construction sites revealed that soil type and texture significantly impact bulldozer efficiency. For example, bulldozers operating on sandy-clay mixtures, as shown in [Figure 4.1](#), required precise adjustments to blade positioning and load balance to ensure smooth operation. This figure showcases the interaction between the bulldozer and sandy-clay soil: (A) illustrates the bulldozer's blade working through the mixture, and (B) a close-up of the sandy-clay soil's granular composition. Similarly, on sandy soil, as depicted in [Figure 4.2](#), operational challenges included traction management and blade stability. In this figure, (A) shows the bulldozer's blade moving sandy soil, and (B) highlights the fine texture of the sandy soil. These images not only emphasize the diversity of soil conditions faced during operations but also underline the necessity of tailoring productivity prediction models to include such granular real-world factors.

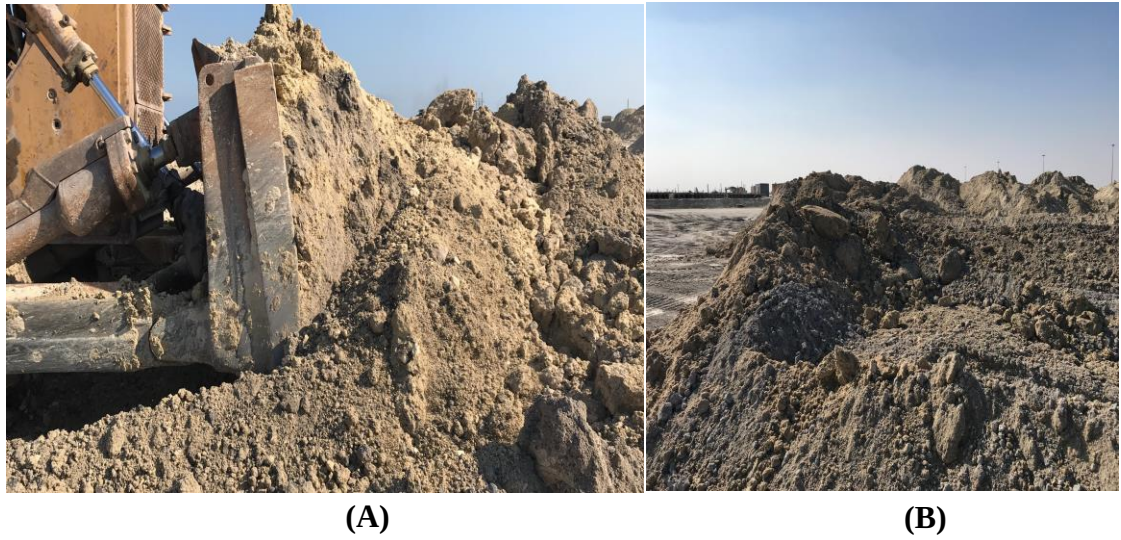


Figure 4.1: D155A Bulldozer, Qatif, Saudi Arabia. (A) Blade interaction, (B) Sandy-clay mix.

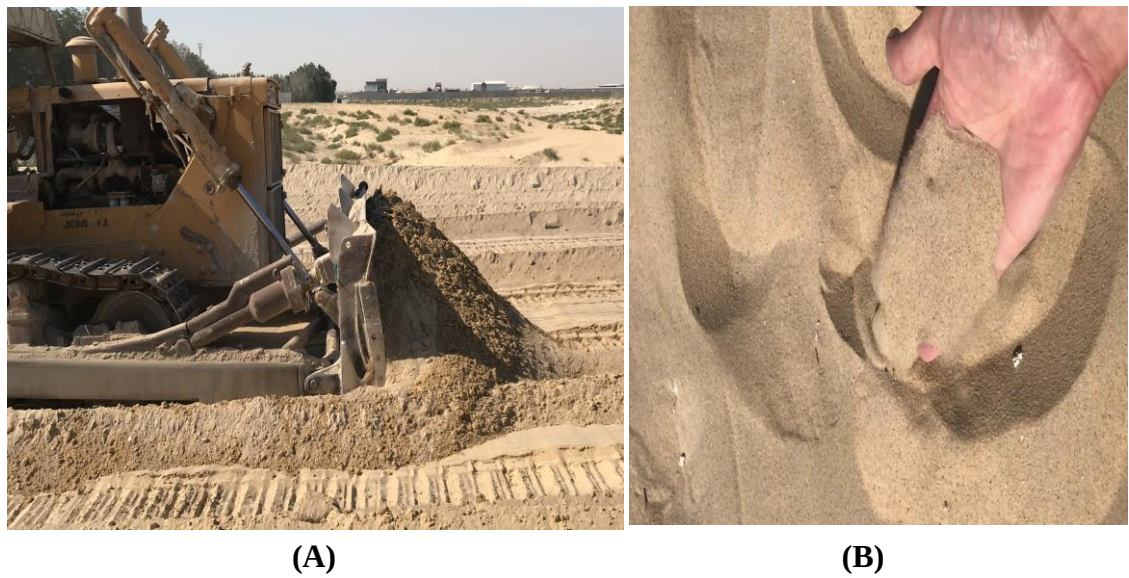


Figure 4.2: D155A Bulldozer, (A) Blade interaction, (B) Sandy soil texture.

4.4 Data Cleaning and Outlier Removal

Once the GPS data was transformed and segmented, the dataset underwent a thorough cleaning process to remove outliers and ensure data integrity. Outliers in the dataset, such as excessively high or low speeds and abnormally short or long cycles, were identified and removed. These outliers could have resulted from GPS inaccuracies or sudden, non-operational movements on-site and needed to be filtered out to avoid skewing the analysis.

- **Box Plot Analysis for Outlier Detection:** To systematically identify outliers, box plots were generated for key metrics. As shown in [Figure 4.3](#), box plots were used to compare the dataset before and after outlier removal, ensuring that only realistic values remained for analysis.

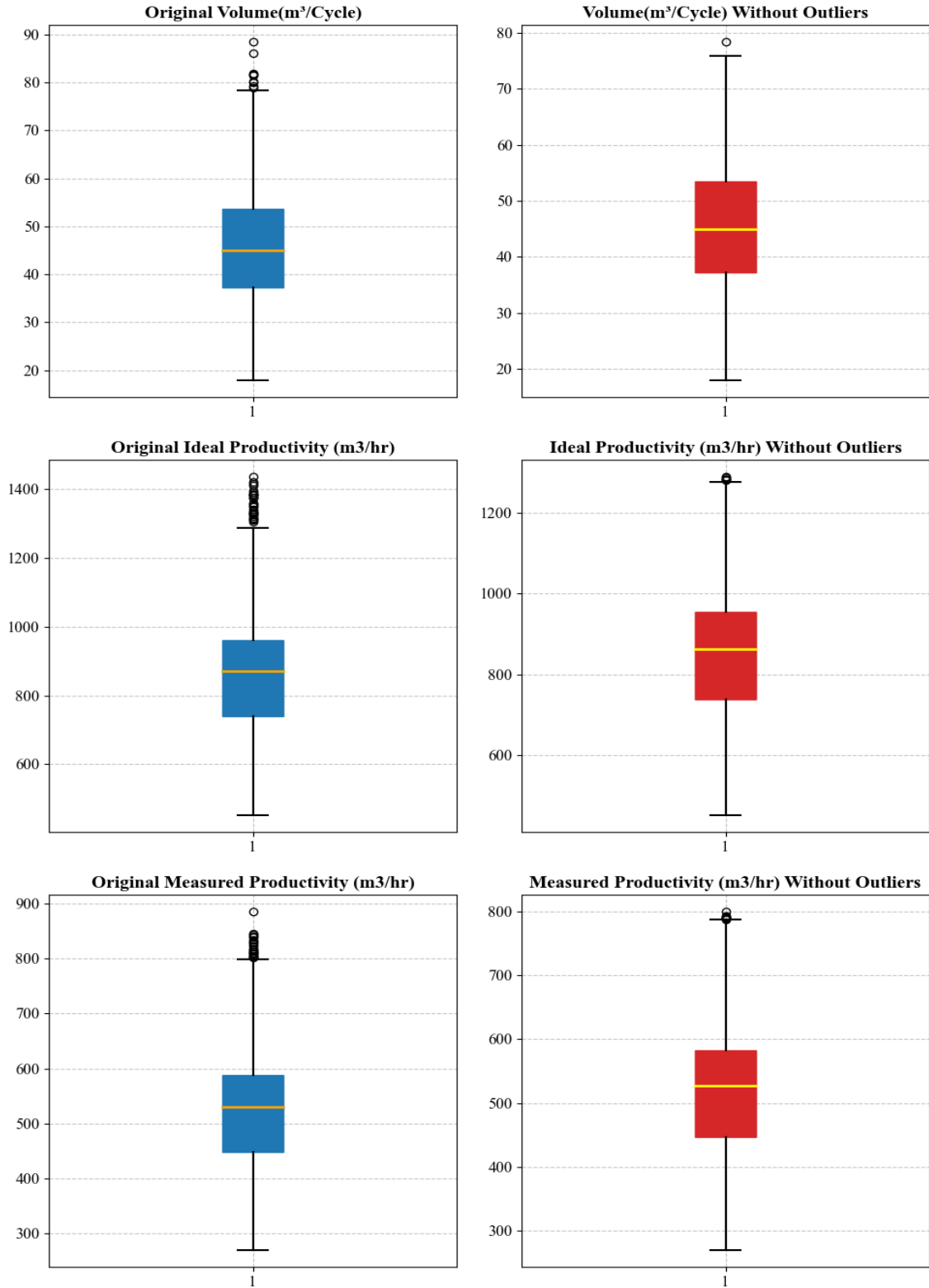


Figure 4.3: Box Plots of Before and After Outlier Removal

4.5 Data Analysis

4.5.1 Normalization Boxplots:

The "Before and After Normalization Boxplots the [Figure 4.4](#) illustrate the distribution of key productivity-related variables before and after the normalization process. Before normalization, the boxplots demonstrate significant variability in the scales of features such as cycle time, speed, and adjusted productivity, making direct comparisons between variables challenging. After normalization, the boxplots show that all variables have been scaled to a uniform range (0–1), ensuring that no feature dominates others in the modeling process. This step is crucial for improving the performance and reliability of machine learning algorithms, as it standardizes the input data for consistent analysis.

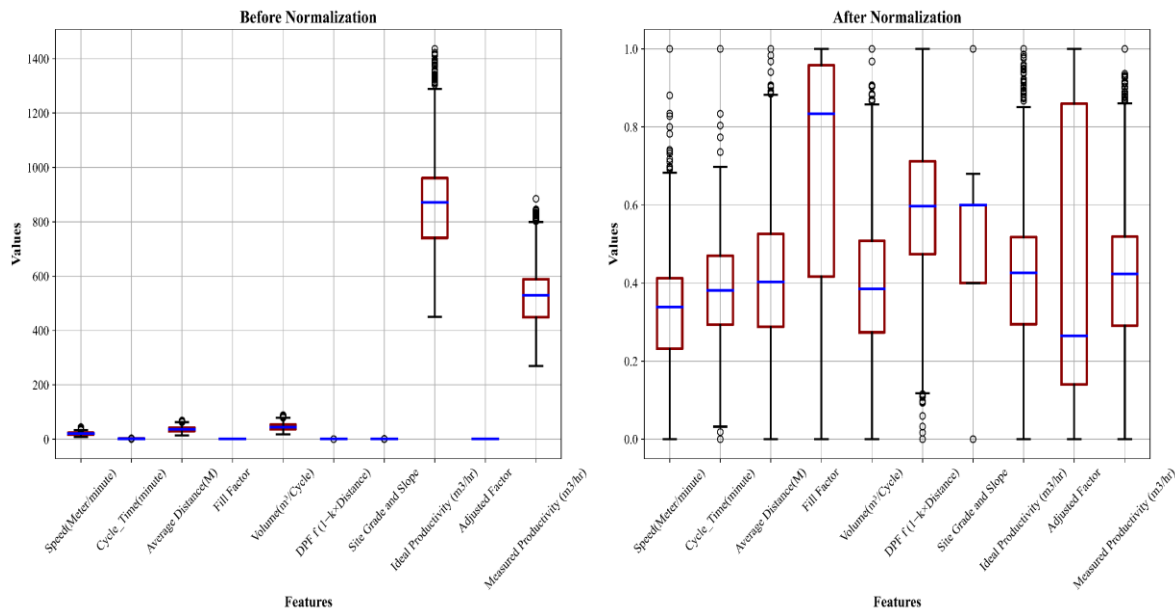


Figure 4.4: Normalized and Unnormalized Boxplots

4.5.2 Pairplot of Features:

The Customized Pairplot of Features [Figure 4.5](#) showcases scatterplots and density distributions of key features such as average distance, speed, cycle time, and measured productivity. The scatterplots reveal linear and nonlinear relationships, with some variables showing strong positive or negative correlations. For example, as the average distance increases, the scatterplot suggests a decrease in productivity due to extended cycle durations. The density distributions along the diagonal provide insights into the range and frequency of values for each feature, highlighting any skewness or concentration of data points.

Customized Pairplot of Selected Features

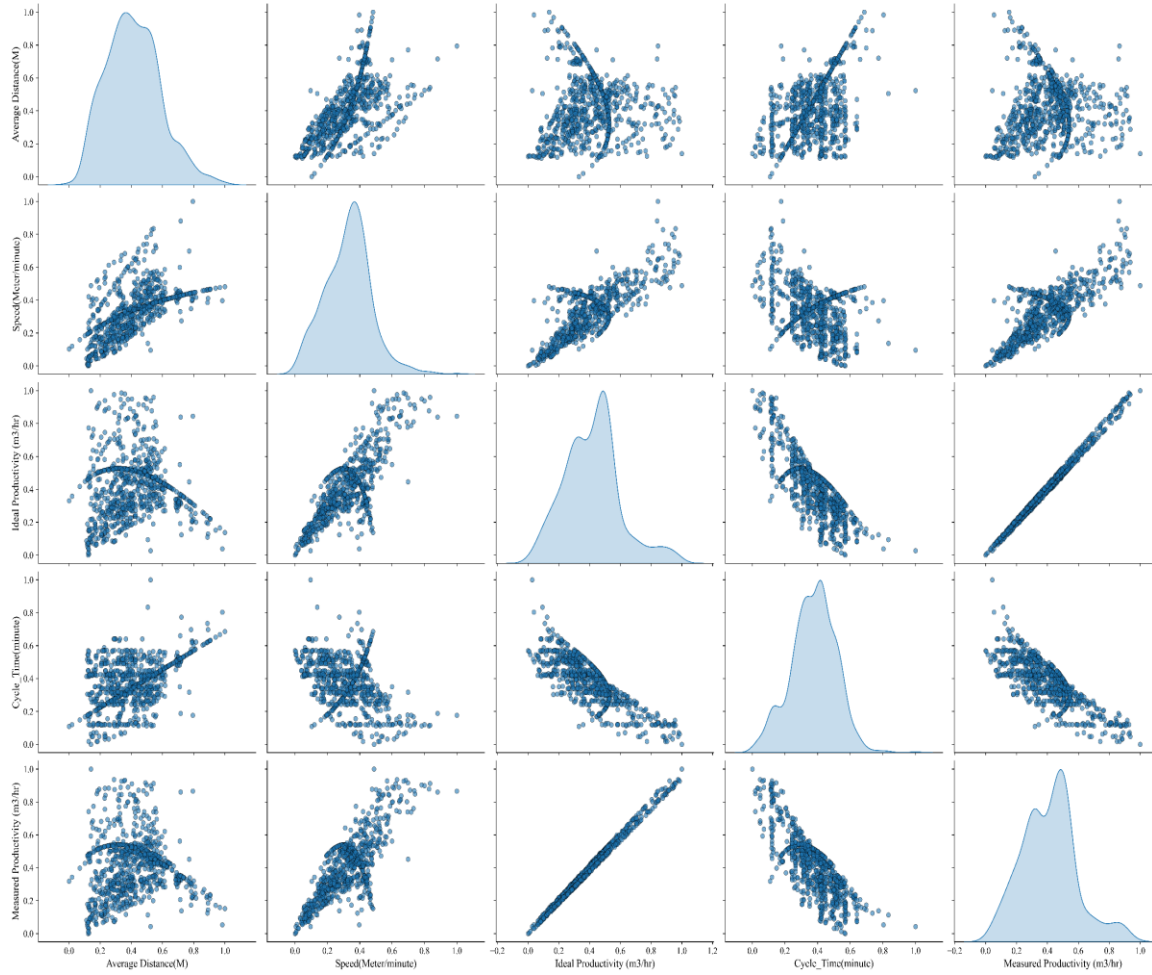


Figure 4.5: Pairplot of Key Variables

4.5.3 Correlation Heatmap:

The Correlation Matrix Heatmap [Figure 4.6](#) provides a detailed visualization of the correlation coefficients between different productivity-related features. This heatmap identifies significant relationships, such as the strong positive correlation between measured productivity and speed, and the negative correlation between cycle time and productivity. Additionally, features like "Distance per Cycle" exhibit negative correlations

with productivity metrics, signaling inefficiencies caused by excessive movement. This heatmap guides the feature selection process, ensuring that highly correlated predictors are prioritized for model development while avoiding multicollinearity.

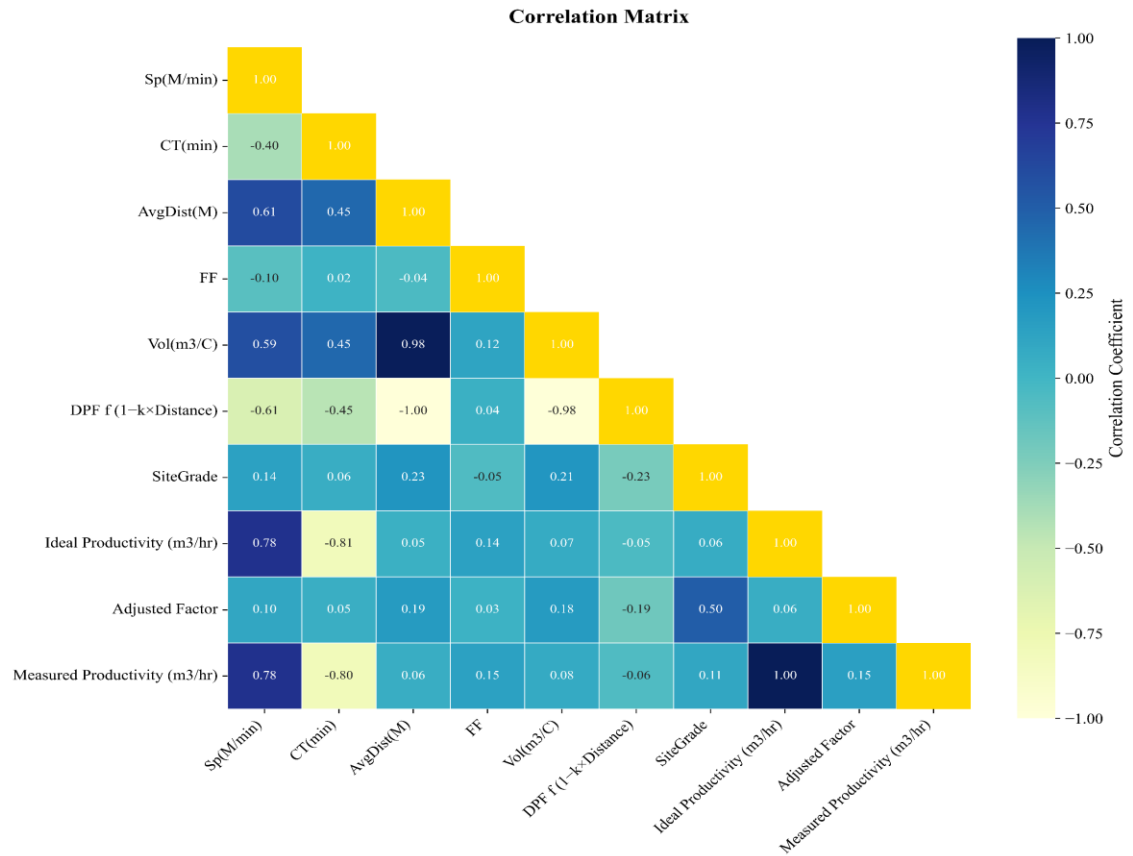


Figure 4.6: Heatmap of Feature Correlations

Table 4.2 provides a detailed description of all the features included in the dataset. It lists the feature names, their corresponding labels, and a concise explanation of each feature. These features encompass all variables collected during the study, which serve as input data for analyzing and predicting the productivity of heavy equipment.

Table 4.2: Description and Labels of Dataset Features

Label	Description
Sp(M/min)	Speed (Meter/minute)
CT(min)	Cycle Time (minute)
AvgDist(M)	Average Distance (M)
FF	Fill Factor
Vol(m ³ /C)	Volume (m ³ /Cycle)
DPF	Distance Penalty Factor $f(1-k \times \text{Distance})$
SoilType	Soil Type (Clay, Sand, Silt, etc.)
ProjPlan	Project Planning
JobEff(min/hr)	Job Efficiency (minute/hour)
OpSkillExp	Operator Skill and Experience
WeatherCond	Weather Conditions (Rain, Snow, Wind)
SiteGrade	Site Grade and Slope
EquipAgeCond	Equipment Age and Condition
OpFatigue	Operator Fatigue
IdealProd(m ³ /hr)	Ideal Productivity (m ³ /hr)
MeasProd(m ³ /hr)	Measured Productivity (m ³ /hr)

4.6 Models Performance

In this section, the predictive accuracy of the ANN, SVM, and RF models was evaluated using the test dataset. The dataset was divided into 80% training data and 20% test data, ensuring that the models were trained on a substantial portion of the data while being validated on an independent test set. This methodology provided a robust framework to assess how effectively each model generalized to new, unseen data. Performance metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 , were used to evaluate the models. These metrics captured the magnitude of prediction errors and

the proportion of variance explained by the models, offering a comprehensive view of their predictive performance.

4.6.1 The ANN Model

The ANN model was meticulously designed as the primary predictive tool for this study due to its exceptional ability to model non-linear relationships. This robust architecture consisted of multiple layers, each playing a crucial role in capturing complex patterns within the dataset. The input layer served as the entry point, taking in all relevant productivity features such as cycle time, average distance, operator skill, and environmental conditions. Each feature was represented by an individual neuron, ensuring comprehensive coverage of the factors influencing bulldozer productivity.

Hidden layers were at the heart of the model's learning capacity, composed of multiple neurons that progressively transformed input data through weighted connections and activation functions. These layers were configured after extensive experimentation to strike a balance between capturing intricate data relationships and avoiding overfitting. The ReLU activation function was employed within these layers to introduce non-linearity and mitigate issues such as the vanishing gradient problem. This activation choice enhanced the model's ability to identify subtle interactions between features. The output layer consisted of a single neuron tasked with predicting continuous productivity values (m^3/hr). A linear activation function was used to ensure accurate continuous outputs aligned with the nature of the target variable.

The ANN model's training process further reinforced its capabilities. The Adam optimizer, known for its adaptability and efficiency, was selected to adjust learning rates dynamically based on gradient information. The MSE loss function quantified prediction errors, with lower values indicating better model performance. The training was conducted over several epochs with a carefully chosen batch size to balance computational efficiency and convergence.

4.6.2 The ANN model Parameters:

Table 4.3 outlines the architecture of a deep learning model, providing details on the layers, their types, output shapes, and the number of trainable parameters at each stage. The model follows a structured design, starting with a fully connected layer (Dense) containing 128 neurons, followed by a sequence of normalization, activation, and dropout layers. Each Dense layer reduces the number of neurons progressively, forming a funnel-shaped architecture, ultimately culminating in a single-neuron output layer designed for tasks like regression or binary classification.

Each Dense layer computes outputs by connecting every input neuron to every output neuron, with trainable weights and biases that adapt during training. The batch normalization layers standardize these outputs to stabilize learning and accelerate convergence. Activation layers introduce non-linearity into the model using predefined functions like ReLU, enabling the network to learn complex patterns. Dropout layers are strategically placed to combat overfitting by randomly deactivating neurons during training, ensuring the model generalizes well to unseen data.

The output shapes indicate the number of neurons in each layer, with the first dimension (None) allowing the model to handle variable batch sizes. For instance, the initial Dense layer outputs 128 neurons with 2,944 trainable parameters, derived from the formula $(\text{input neurons} \times \text{output neurons}) + \text{biases}$. The model gradually narrows, with subsequent layers reducing the neuron count to 64, then 32, and so on, until the final Dense layer outputs a single value, with just 17 parameters connecting the 16 neurons from the previous layer.

This design emphasizes efficiency and scalability. Batch normalization layers (e.g., `batch_normalization_8`) include scale and shift parameters to adjust outputs dynamically, while dropout layers ensure robustness without adding trainable parameters. The model's progressive reduction in neurons and its use of regularization and normalization techniques highlight a well-structured approach to balancing complexity and generalization, making it suitable for tasks requiring precise predictions with minimal overfitting. This detailed table offers insights into the model's capacity, making it easier to understand its functional and computational characteristics.

Table 4.3: The ANN Model Parameters

Layer (type)	Output Shape	Param #
dense_10 (Dense)	(None, 128)	2,944
batch_normalization_8 (BatchNormalization)	(None, 128)	512
activation_8 (Activation)	(None, 128)	0
dropout_8 (Dropout)	(None, 128)	0
dense_11 (Dense)	(None, 64)	8,256
batch_normalization_9 (BatchNormalization)	(None, 64)	256
activation_9 (Activation)	(None, 64)	0
dropout_9 (Dropout)	(None, 64)	0
dense_12 (Dense)	(None, 32)	2,080
batch_normalization_10 (BatchNormalization)	(None, 32)	128
activation_10 (Activation)	(None, 32)	0
dropout_10 (Dropout)	(None, 32)	0
dense_13 (Dense)	(None, 16)	528
batch_normalization_11 (BatchNormalization)	(None, 16)	64
activation_11 (Activation)	(None, 16)	0
dropout_11 (Dropout)	(None, 16)	0
dense_14 (Dense)	(None, 1)	17

Finally, the models' parameters are summarized in the following table:

Table 4.4: Summary of the parameters

Total params:	14,785
Trainable params:	14,305
Non-trainable params:	480

4.6.3 The ANN Architecture

The network is formed of an input layer, three fully connected hidden layers, and a single-neuron output layer. This design facilitates the progressive transformation of input features into meaningful predictions through a systematic computational process.

Input Layer: The input layer serves as the entry point for data into the model, comprising 128 neurons, each representing a unique feature or variable from the dataset. These neurons function as placeholders, relaying raw data into the network without performing any computations. By matching the input layer size to the dimensionality of the dataset, the model ensures all relevant features are utilized in the prediction process.

Hidden Layers: The model features three fully connected hidden layers, each consisting of 128 neurons. These layers form the computational core of the network, learning patterns and relationships within the data.

- **Hidden Layer 1:** The first hidden layer processes the input features by applying weights, biases, and a non-linear activation function. This initial layer extracts basic patterns and prepares the data for deeper analysis.
- **Hidden Layer 2 and Hidden Layer 3:** These subsequent layers refine the representations learned in the first layer. By applying additional transformations and non-linearities, they uncover complex relationships and high-level features crucial for accurate predictions.

Output Layer: The output layer contains a single neuron, providing the final predicted value based on the processed information from the hidden layers. Designed for regression

tasks, this layer employs a linear activation function to produce continuous outputs. This architecture is ideal for tasks requiring precise numerical predictions, such as productivity or performance metrics.

Key Features and Advantages: The consistent size of 128 neurons across the hidden layers ensures the network has sufficient capacity to capture complex relationships while avoiding overfitting. The progressive reduction of data dimensions through these layers allows the model to focus on the most critical patterns in the dataset. Regularization techniques, such as dropout or batch normalization (not explicitly shown in the figure), can be integrated to further improve model generalization and stability. Figure 4.7 visually demonstrates the systematic flow of data through the model, from the input layer to the output layer. This architecture balances computational efficiency and predictive power, making it well-suited for a wide range of structured data applications. Its flexibility and robustness ensure that it can handle diverse datasets and deliver reliable predictions.

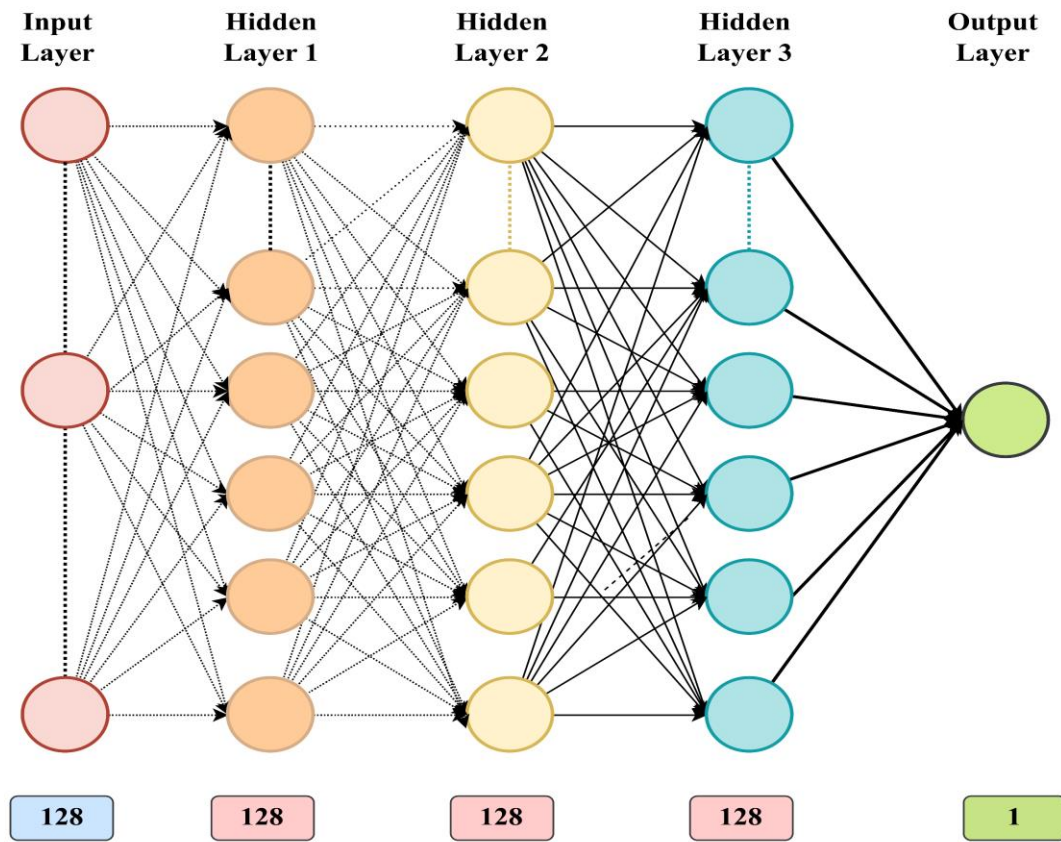


Figure 4.7: Architecture of the ANN model

4.6.4 The ANN Model Performance Evaluation

The performance evaluation of the ANN model is summarized through its results on both the training and testing datasets, as shown in [Figures 4.8 and 4.9](#). These figures illustrate the correlation between the measured productivity values and the predictions made by the model, while also reporting key performance metrics such as the R^2 , RMSE, and MAE. Together, these results provide a detailed assessment of the model's accuracy and generalization capability.

The scatter plot indicates a strong linear relationship between the measured and predicted values, as evidenced by an R^2 score of 0.98. This score suggests that 98% of the variance

in the measured productivity values is explained by the model on the training data, highlighting its strong learning capacity. Additionally, the RMSE value of 14.59 and MAE value of 10.37 reflect the model's accuracy in predicting productivity, with minimal average prediction errors. The clustering of points closely along the diagonal line reinforces the model's ability to accurately capture patterns in the training dataset, suggesting a well-tuned model that effectively learns the underlying relationships in the data. [Figure 4.8](#) demonstrates the ANN model's performance on the training dataset.

The performance of the ANN model on the testing dataset is critical for evaluating its ability to generalize to unseen data. The results are similarly impressive, with an R^2 score of 0.96, indicating that 96% of the variance in the test data is captured by the model. While the RMSE of 20.92 and MAE of 11.74 are slightly higher than those observed in the training dataset, this is expected due to the variability and potential noise in the testing data. Despite this, the distribution of points remains closely aligned along the diagonal, signifying that the model performs consistently well when predicting unseen productivity values. [Figure 4.9](#) shows this performance underscores the hardness and reliability of the model for generalization tasks.

In conclusion, the evaluation metrics and visualizations in [Figures 4.8](#) and [4.9](#) confirm the ANN model's ability to accurately predict productivity values. The slight increase in error values for the testing dataset is within acceptable limits, suggesting that the model avoids overfitting while maintaining high accuracy. These results validate the effectiveness of the model for regression tasks, demonstrating its potential as a reliable tool for productivity prediction.

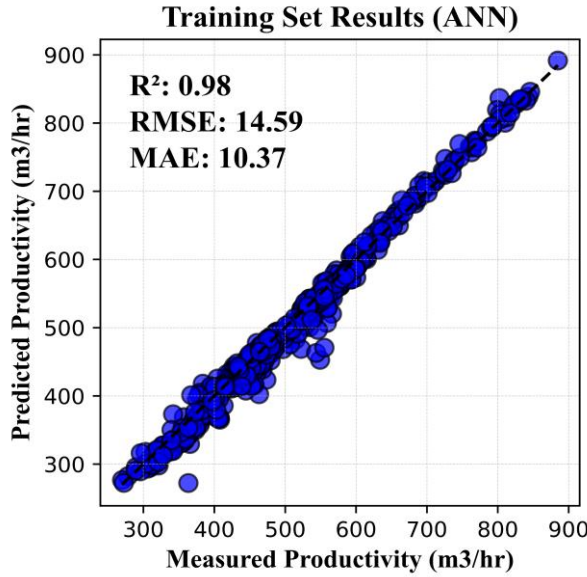


Figure 4.8: Training set results of the ANN

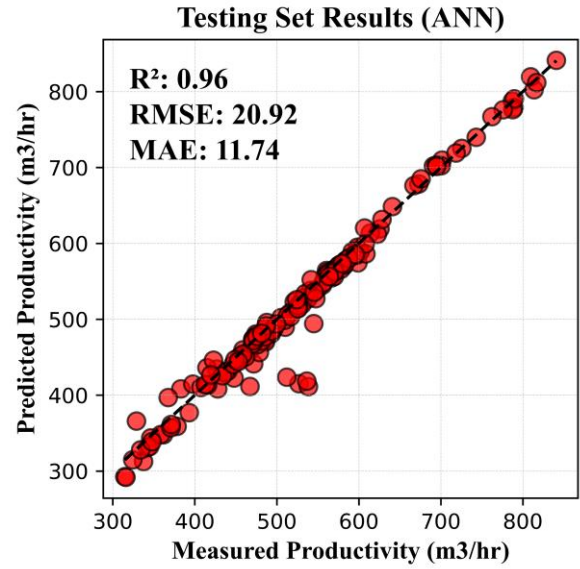


Figure 4.9: Testing set results of the ANN

4.6.5 The SVM Model

The SVM model was utilized as a baseline model to compare its performance against the primary ANN model. To accommodate the non-linear relationships present in the dataset, the RBF kernel was employed, providing the model with the ability to map input features into a higher-dimensional space and effectively identify complex patterns.

4.6.6 The SVM model Parameters

The SVM model used in this study was configured with specific parameters to optimize its performance for the given prediction task. The selected kernel function was the RBF, which is particularly effective for capturing non-linear relationships in the data. The RBF kernel maps input features into a higher-dimensional space, allowing the model to find a more complex decision boundary that separates or predicts data points accurately.

The regularization parameter (C) was set to 5, which controls the trade-off between maximizing the margin and minimizing classification errors. A higher C value places greater emphasis on correctly predicting all training points, thereby reducing misclassification but potentially leading to a more complex model. This value was chosen to balance model accuracy with generalization.

The gamma parameter was set to '*scale*', an automatic setting in modern SVM implementations that adjusts gamma based on the number of features in the dataset and their variance. This setting ensures that the influence of each data point is appropriately scaled, allowing the model to maintain flexibility without overfitting. Additionally, the epsilon parameter (ϵ) was set to 0.1, a crucial hyperparameter for regression tasks that defines a margin of tolerance where no penalty is given for predictions within a certain range of the actual values. This parameter helps to fine-tune the model's precision in predicting continuous values.

These carefully selected parameters ensure that the SVM model is well-suited for the task, balancing complexity and generalization to deliver accurate and reliable predictions. By using the RBF kernel and fine-tuning C , gamma, and epsilon, the model is optimized for capturing both linear and non-linear patterns in the dataset.

4.6.7 The SVM Model Representation

The SVM's primary objective is to maximize the margin, and the space between the closest points of each class, while ensuring correct classification. The central line, known as the max-margin hyperplane, is equidistant from the closest points of both classes, which are referred to as support vectors. These critical points define the position and orientation of

the hyperplane, making them the only data points influencing the model. The dashed lines represent the positive and negative hyperplanes, marking the boundaries of the margin. By maximizing the margin, the SVM enhances its ability to generalize to new, unseen data. This robust approach is effective for both linearly separable data, as shown, and non-linearly separable data through the use of kernel functions. Figure 4.10 illustrates the fundamental concept of the SVM algorithm, which identifies an optimal hyperplane to separate two classes of data points—represented here as red circles and green squares.

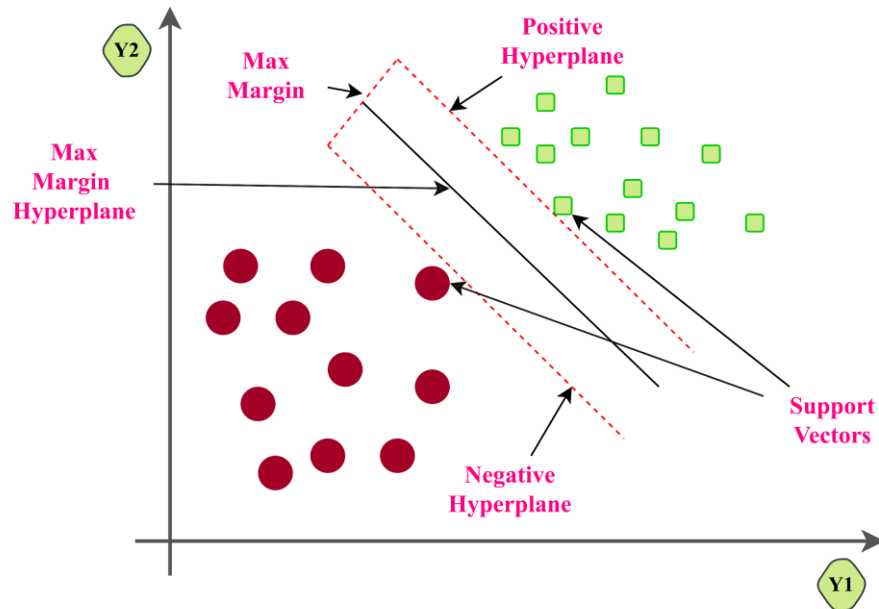


Figure 4.10: Conceptual Representation of SVM

4.6.8 The SVM Model Performance Evaluation

The performance of the SVM model is evaluated using both the training and testing datasets, as shown in Figures 4.11 and 4.12. These results provide a detailed assessment of the model's ability to predict productivity values accurately, including key performance

metrics such as the R^2 , RMSE, and MAE. The analysis of these metrics sheds light on the model's learning capability, accuracy, and ability to generalize to unseen data.

The training dataset results, illustrated in [Figure 4.11](#), indicate that the SVM model achieves a reasonable level of accuracy. With an R^2 value of 0.86, the model is able to explain 86% of the variance in the measured productivity values. This demonstrates that the SVM effectively captures most of the underlying patterns within the training data. However, the RMSE of 41.79 and MAE of 19.23 highlight some variability in the model's predictions, with an average absolute error of 19.23 m³/hr. While the scatter plot shows a general alignment of the predicted values along the diagonal line, suggesting that the model is capable of learning the relationships between input features and target values, there is some noticeable scatter, particularly at higher productivity values. This suggests that the model struggles slightly with extreme cases or outliers even within the training data.

The testing dataset results, presented in [Figure 4.12](#), provide a measure of the SVM model's generalization capability. The R^2 value of 0.83 on the test set is slightly lower than that of the training set, which is expected, as the test data consists of unseen examples. This value indicates that the model captures 83% of the variance in the test set, demonstrating reasonable generalization to new data. However, the RMSE of 45.10 and MAE of 21.15 are slightly higher than the corresponding training metrics, reflecting a modest increase in prediction error. The scatter plot for the testing results shows that while most points are closely aligned with the diagonal line, there are noticeable deviations, particularly for higher productivity values. This indicates that the model's predictions for higher

productivity values are less accurate, suggesting some limitations in its ability to handle complex or extreme scenarios.

Overall, the SVM model demonstrates satisfactory performance in predicting productivity values, with good alignment between measured and predicted values in both the training and testing datasets. The relatively small difference in R^2 scores and error metrics between the two datasets suggests that the model generalizes reasonably well, avoiding significant overfitting. However, compared to more advanced models like ANN, the SVM may have limitations when handling highly non-linear relationships or extreme cases in the dataset.

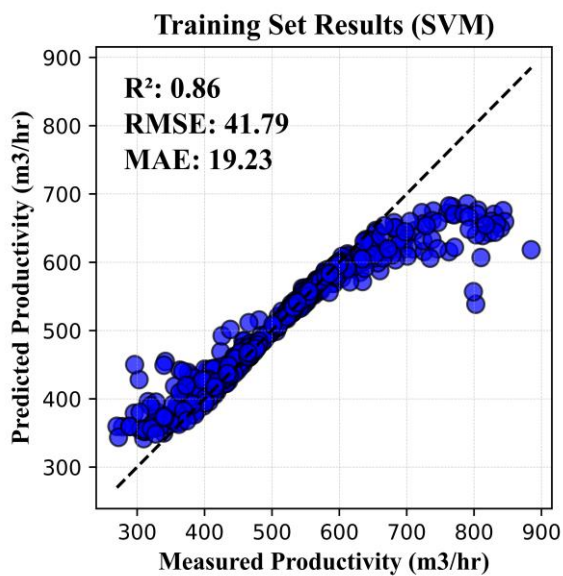


Figure 4.11: Training set results of the SVM

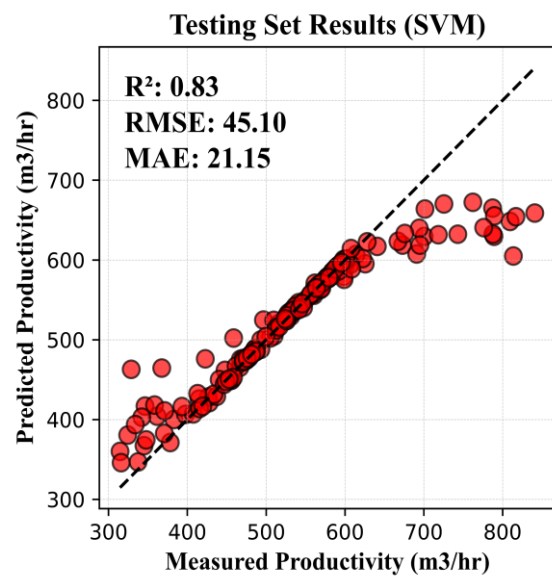


Figure 4.12: Testing set results of the SVM

4.4.9 The RF Model

The RF model was employed as a comparative baseline due to its ensemble-based approach, which aggregates predictions from multiple decision trees to enhance accuracy

and robustness. This method reduces variance and mitigates overfitting, making RF particularly effective for structured datasets with moderate non-linearity. The ensemble's ability to generalize well to unseen data was particularly advantageous for modeling the complexities of bulldozer productivity.

4.6.10 The RF model Parameters

The RF model used in this study was configured with a set of carefully chosen hyperparameters to optimize its performance for the prediction task. The *n_estimators* parameter was set to 100, specifying the number of decision trees in the ensemble, ensuring robustness and diversity in predictions. The *max_depth*, limited to 5, controls the maximum depth of each tree to prevent overfitting by restricting individual tree complexity. A *random state* of 42 was used to ensure reproducibility by controlling the randomness of the model's operations. The *min_samples_split* parameter, set to 30, defines the minimum number of samples required to split an internal node, ensuring splits occur only when the data volume is sufficient for stability. Similarly, the *min_samples_leaf*, set to 20, specifies the minimum number of samples required at a leaf node, preventing unreliable and overly small leaf nodes. Finally, the *max_features* parameter, set to 'sqrt', determines the number of features considered for the best split, balancing computational efficiency and predictive accuracy. These parameter values are summarized in [Table 4.5](#) below.

Table 4.5: Hyperparameters and descriptions for the Random Forest model

Parameter	Value	Description
n_estimators	100	The number of trees in the forest.
max_depth	5	The maximum depth of the tree.
random_state	42	Controls the randomness of the estimator.
min_samples_split	30	The minimum number of samples required to split an internal node.
min_samples_leaf	20	The minimum number of samples required to be at a leaf node.
max_features	'sqrt'	The number of features to consider when looking for the best split.

4.6.11 The RF Model Representation

The diagram in [Figure 4.13](#) illustrates the fundamental workflow of the RF model, an ensemble learning technique designed to improve predictive performance by combining the outputs of multiple decision trees. In this approach, the dataset is used to build multiple decision trees, each trained independently on a random subset of the data and features. This process introduces diversity among the trees, ensuring that each tree captures slightly different patterns in the data. As shown in the figure, the predictions from the individual trees Results 1, 2, and 3 are aggregated, typically through averaging in regression tasks or majority voting in classification tasks, to produce the final RF prediction. This ensemble method reduces the variance and enhances the robustness of the model, making it less prone to overfitting compared to individual decision trees. The visual representation highlights

how RF leverages the strengths of multiple decision trees to achieve accurate and reliable predictions.

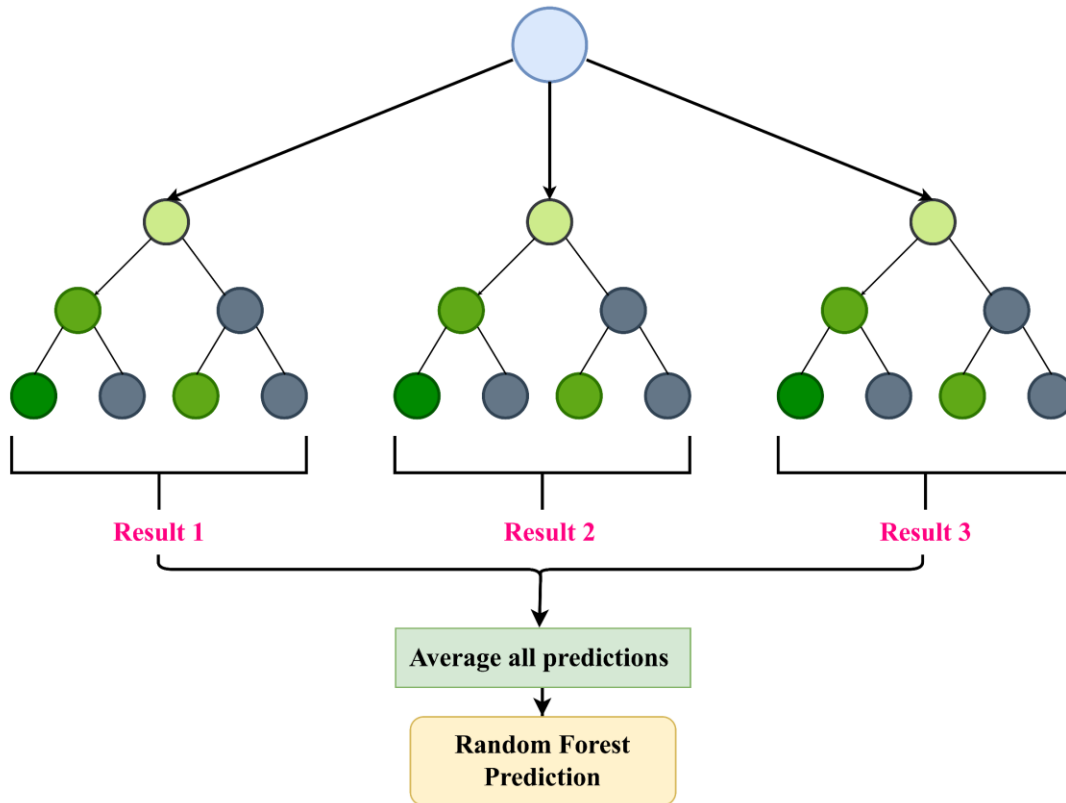


Figure 4.13: Random Forest Model Architecture

4.6.12 The RF Model Performance Evaluation

The performance of the RF model was evaluated using both the training and testing datasets, as illustrated in Figures 4.14 and 4.15. These figures visually present the correlation between measured and predicted productivity values, accompanied by key performance metrics such as the R^2 , RMSE, and MAE. Together, these metrics provide insights into the model's accuracy, ability to generalize to unseen data, and overall effectiveness in capturing patterns within the dataset.

On the training dataset, as shown in [Figure 4.14](#), the RF model achieved an R^2 value of 0.92. This indicates that 92% of the variance in the measured productivity values was explained by the model during training. While this performance is reasonable, it is lower compared to other models, suggesting that the RF model captures a substantial but not exhaustive amount of information from the training data. The RMSE of 32.01 and MAE of 19.68 reflect the magnitude of prediction errors, indicating moderate accuracy. The scatter plot reveals a reasonable alignment of predicted values along the diagonal line, which represents perfect predictions. However, some deviation is observed, particularly for higher productivity values, where the model tends to underestimate or overestimate. This variability suggests that the RF model has limitations in learning the complete complexity of the dataset, possibly due to restrictions in its parameter settings or inherent constraints in ensemble learning.

For the testing dataset, as depicted in [Figure 4.15](#), the RF model produced an R^2 value of 0.89, which is slightly lower than the training set R^2 , indicating a moderate decrease in performance when applied to unseen data. This drop highlights the model's limited ability to generalize fully to new scenarios. The RMSE increased to 35.55, and the MAE rose to 22.60, reflecting higher prediction errors compared to the training set. The scatter plot for the testing data demonstrates a similar trend, with most points clustering around the diagonal but with noticeable deviations, especially at higher productivity values. This indicates that the model struggles more with extreme values in the test set, which further underscores its limited capacity to handle variability and outliers effectively.

In conclusion, the performance evaluation of the RF model demonstrates that it performs reasonably well on both training and testing datasets, capturing a substantial amount of variance in productivity predictions. However, the observed gaps between the R^2 values, RMSE, and MAE for the training and testing sets indicate limitations in its generalization ability, particularly when predicting higher productivity values or handling more complex data patterns. These results suggest that while the RF model is a robust ensemble learning technique, it may not be the most suitable model for highly non-linear or extreme-value scenarios in this dataset. Further optimization of hyperparameters or alternative modeling techniques could enhance its predictive performance.

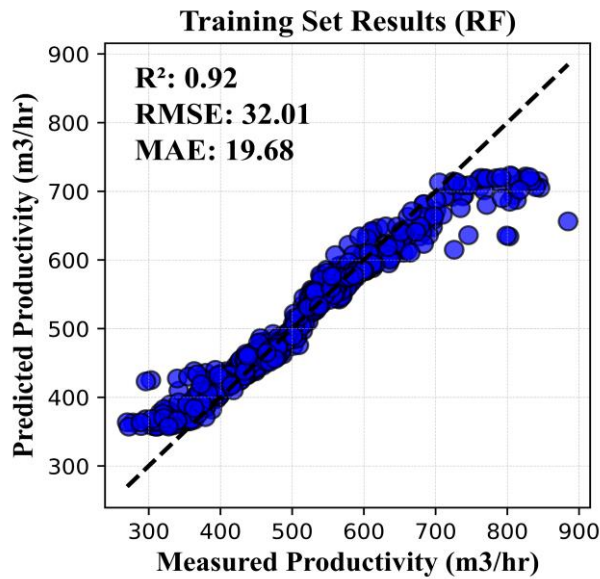


Figure 4.14: Training set results of the RF

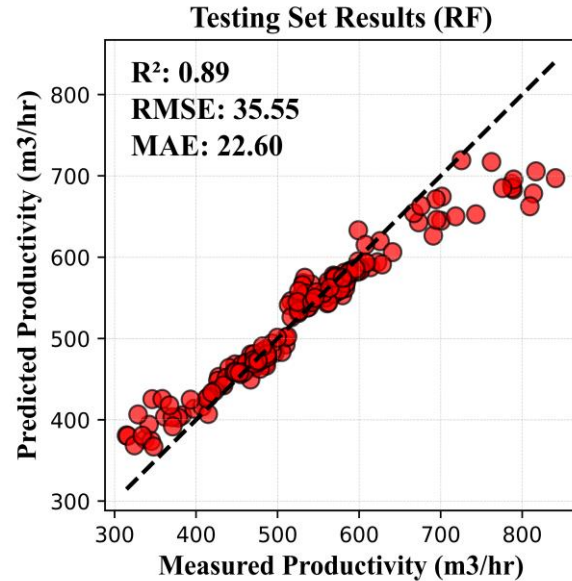


Figure 4.15: Testing set results of the RF

4.7 Models Comparison Summary

To provide a clear comparison of the performance of the three models SVM RF, and ANN their key performance metrics across training and testing datasets are summarized in [Table](#)

4.6 below. These metrics include the R^2 , RMSE, and MAE, which collectively highlight the models' accuracy, generalization capabilities, and prediction reliability.

Table 4.6: Performance Comparison of SVM, RF, and ANN Models

Model	Training R^2	Training RMSE	Training MAE	Testing R^2	Testing RMSE	Testing MAE
Support Vector Machine (SVM)	0.86	41.79	19.23	0.83	45.1	21.15
Random Forest Regressor (RF)	0.92	32.01	19.68	0.89	35.55	22.60
Neural Network (ANN)	0.98	14.59	10.37	0.96	20.92	11.74

The ANN continues to demonstrate the best performance across all metrics. It achieves the highest R^2 values for both training (0.98) and testing (0.96), reflecting its superior ability to capture the variance in the data and generalize effectively to unseen examples. Moreover, the ANN produced the lowest RMSE and MAE values for both datasets, with a testing RMSE of 20.92 and MAE of 11.74, indicating minimal prediction errors. The close alignment between the training and testing results suggests that the ANN model is well-optimized and effectively avoids overfitting.

The RF Regressor shows moderate performance, with an R^2 value of 0.92 for training and 0.89 for testing. Its RMSE and MAE values are higher than those of the ANN, with a testing RMSE of 35.55 and MAE of 22.60, indicating a larger margin of prediction errors. This suggests that while RF is robust in handling variability through ensemble learning, it struggles with the complexity and non-linearity of the dataset compared to the ANN.

The SVM model performed slightly better than RF but significantly worse than ANN. With an R^2 value of 0.86 for training and 0.83 for testing, the SVM displayed reasonable generalization. However, its testing RMSE of 45.10 and MAE of 21.15 highlight its inability to achieve the same level of accuracy as the ANN. While SVM's RBF kernel allows it to handle non-linear relationships effectively, its performance is constrained by its sensitivity to parameter tuning and reliance on support vectors.

The graph compares the observed productivity values with the predictions made by three machine learning models: ANN, SVM, and RF. The observed values, shown as a black solid line with circular markers, serve as the benchmark for evaluating the models' performance. The ANN predictions (red dashed line) generally follow the observed trend but tend to overestimate productivity in some sections, particularly at higher average distances. The SVM model (green dashed line) shows consistent alignment with the observed data across most of the range, especially in the mid-range distances. In contrast, the RF predictions (blue dotted line) demonstrate good agreement in certain regions but exhibit noticeable deviations when the observed productivity experiences sharp declines.

This comparison highlights the strengths and weaknesses of each model. The ANN model excels in capturing broader trends but shows a tendency to overshoot or undershoot in certain scenarios. The SVM model demonstrates the most reliable performance, closely mirroring the observed data across varying distances, making it an effective choice for balancing accuracy and stability. The RF model, while effective in capturing general trends, struggles to adapt to abrupt changes in productivity, suggesting the need for further refinement in handling variability in the data. Overall, these insights emphasize the

importance of selecting the appropriate model based on the specific characteristics of the dataset and prediction goals as shown in Figure 4.16.

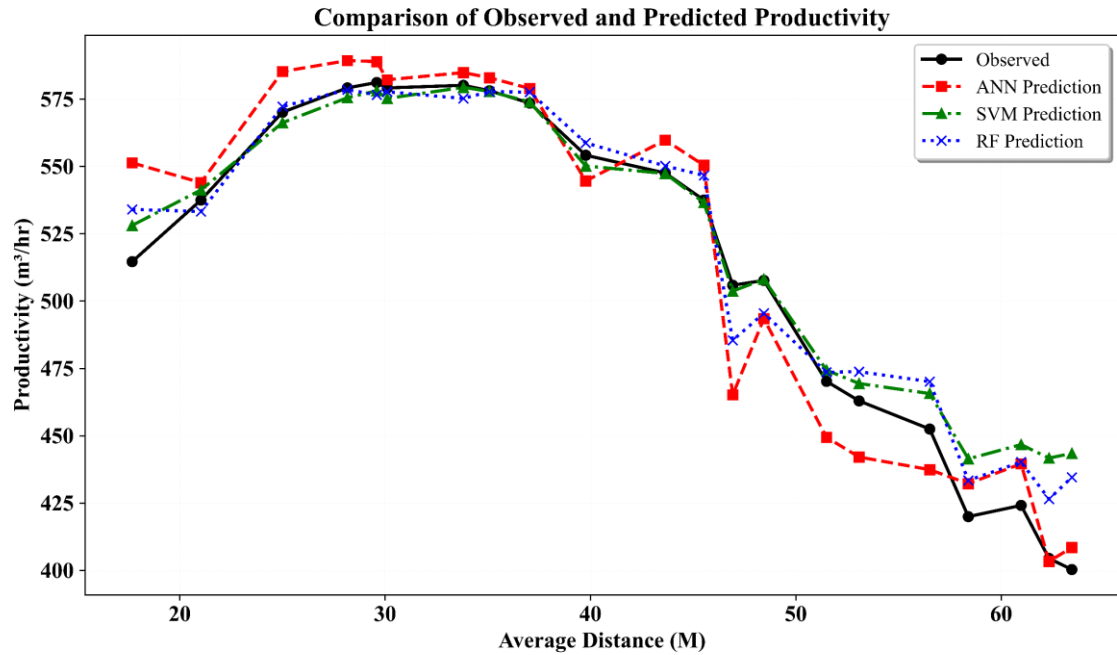


Figure 4.16: Comparison of Observed and Predicted Productivity

In conclusion, the ANN model remains the most effective model for this task, providing the best balance of accuracy and generalization across all datasets. The RF and SVM models, while competitive, are less suited for capturing the full complexity of the data, with RF showing moderate prediction reliability and SVM slightly outperforming RF in error metrics but falling behind ANN overall. These results further reinforce the suitability of advanced architectures like ANN for tasks involving complex, non-linear data.

4.8 Validation of the ANN Model

The ANN model underwent two critical validation steps: k-fold cross-validation and evaluation using random sample data. These validations were designed to assess the

model's robustness, accuracy, and ability to generalize to unseen data. The results, presented in [Table 4.7](#) and [Figure 4.17](#), demonstrate the model's superior performance and reliability across different evaluation techniques.

4.8.1 K-Fold Cross-Validation Results

K-fold cross-validation was conducted to rigorously test the model's performance across multiple subsets of the data. The dataset was divided into five folds, with the model trained on four folds and validated on the fifth in each iteration. This process was repeated five times, ensuring every subset was used once as a validation set. The R^2 , MAE, and RMSE were calculated for each fold, and their averages were computed to provide a comprehensive measure of the model's predictive capabilities. The results, summarized in [Table 4.7](#), highlight the ANN model's consistent performance across all folds. The R^2 values remained uniformly high, with an average of 0.97, indicating that the model explained 97% of the variance in productivity data. This consistency demonstrates that the model is not overfitting to any specific subset and has a strong ability to generalize across different parts of the dataset.

The average MAE was 12.05, and the RMSE was 17.50, reflecting the low prediction errors across all folds. These metrics indicate that the ANN model maintained a high degree of accuracy, regardless of the data used for training and validation. While minor variations in MAE and RMSE were observed between folds, these can be attributed to natural differences in data distribution within each fold and do not detract from the model's overall reliability. The cross-validation process affirms that the ANN model performs consistently well across the dataset, making it a robust choice for predictive tasks.

Table 4.7: Cross-Validation Result for the ANN Model

Fold	R²	MAE	RMSE
1	0.97	12.45	17.50
2	0.97	9.78	14.70
3	0.98	9.10	15.34
4	0.98	12.85	17.44
5	0.96	16.09	22.51
Avg	0.97	12.05	17.50

4.8.2 Validation Using Random Sample Data

To further validate the model's performance, it was tested on a separate set of randomly sampled data, simulating real-world conditions. This evaluation was essential to understand the model's behavior when exposed to entirely new data that was not part of the training or testing sets. The results are visually represented in [Figure 4.17](#), which plots the measured productivity values against the model's predictions. The ANN model achieved an exceptional R^2 value of 0.98 on the validation set, demonstrating its capability to capture nearly all the variance in the new data. This high correlation between measured and predicted values confirms the model's ability to generalize effectively beyond the training and testing datasets. Furthermore, the RMSE of 13.97 and MAE of 11.23 highlight the model's precision, with minimal prediction errors in the random sample data.

The scatter plot in [Figure 4.17](#) provides a visual validation of the model's accuracy. The predicted values align closely with the diagonal line, which represents perfect predictions. The minimal deviations from the line indicate that the ANN model produces highly reliable

predictions across a wide range of productivity values. These results confirm that the model is not only accurate but also robust, and capable of maintaining its performance in varying data scenarios.

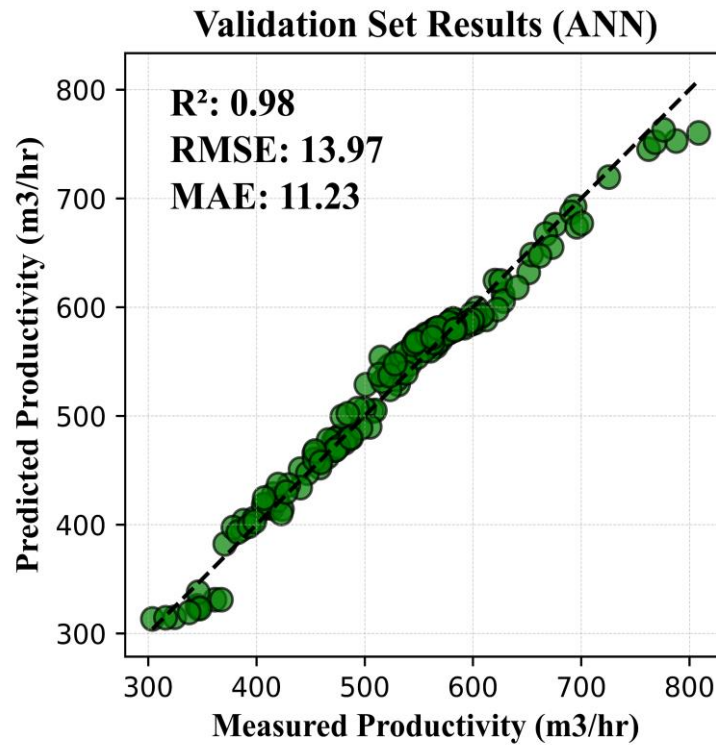


Figure 4.17: Validation Set Results for the ANN Model

The combined results from k-fold cross-validation and random sample validation provide comprehensive evidence of the ANN model's reliability and robustness. The consistently high R^2 values, low RMSE, and MAE, and minimal variability across folds underscore the model's ability to generalize across different subsets of the dataset. The validation of random sample data further reinforces its accuracy and adaptability, proving its effectiveness in real-world scenarios. These validation steps confirm that the ANN model is well-suited for deployment. Its robust performance across multiple evaluation techniques

makes it an excellent candidate for integration into a platform for productivity prediction, enabling accurate and reliable decision-making in practical applications.

4.9 Development of the Web-Based Productivity Prediction Platform

The development of the productivity prediction platform was a key objective of this research. The platform was designed to provide an intuitive and efficient tool for estimating the productivity of bulldozers based on real-world data. Integrating advanced machine learning models with a user-friendly interface, the platform ensures accessibility and reliability for engineers and decision-makers in the construction industry. The use of web-based architecture allows seamless deployment and utilization across multiple devices, providing a scalable and adaptable solution for diverse operational contexts.

The platform workflow is illustrated in the Productivity Prediction Platform Diagram shown in [Figure 4.18](#). This diagram outlines the three key stages of the platform's functionality: Input, Prediction Model, and Output. At the Input stage, users upload datasets containing operational parameters of bulldozers, such as job efficiency, weather conditions, equipment age, and other influencing factors. These datasets undergo preprocessing to ensure compatibility with the integrated ANN model. The Prediction Model stage involves utilizing the ANN to generate productivity predictions based on the input parameters. Finally, the Output stage presents the results in an accessible format, including prediction metrics, tabular results, and visual insights for further analysis and decision-making.

Diagram Description: The Productivity Prediction Platform Diagram (Figure 4.18) visually represents the platform's structure and workflow. It is divided into three distinct sections:

- i. **Input:** The first section, depicted on the left side, illustrates the dataset preparation and upload process. Various files, such as CSV or Excel formats, are consolidated into a server where the preprocessing steps are performed. The inclusion of Excel logos and file flow arrows emphasizes the ease of data handling, reflecting the platform's user-centric design.
- ii. **Prediction Model:** At the center, the machine learning model is represented by a multi-layer ANN. This neural network comprises multiple hidden layers, reflecting its capability to process complex relationships among input features. The gears and speedometer icons symbolize the computational aspects of the model, indicating efficiency and precision in prediction. The flow arrows connecting the input and output sections emphasize the continuous workflow of the system.
- iii. **Output:** On the right side, the results section showcases an analyst or decision-maker interpreting the platform's outputs. Scatter plots and performance metrics (R^2 , MAE, RMSE) are displayed, demonstrating the platform's reliability and accuracy. This section underscores the actionable insights provided by the platform, empowering users to make data-driven decisions.

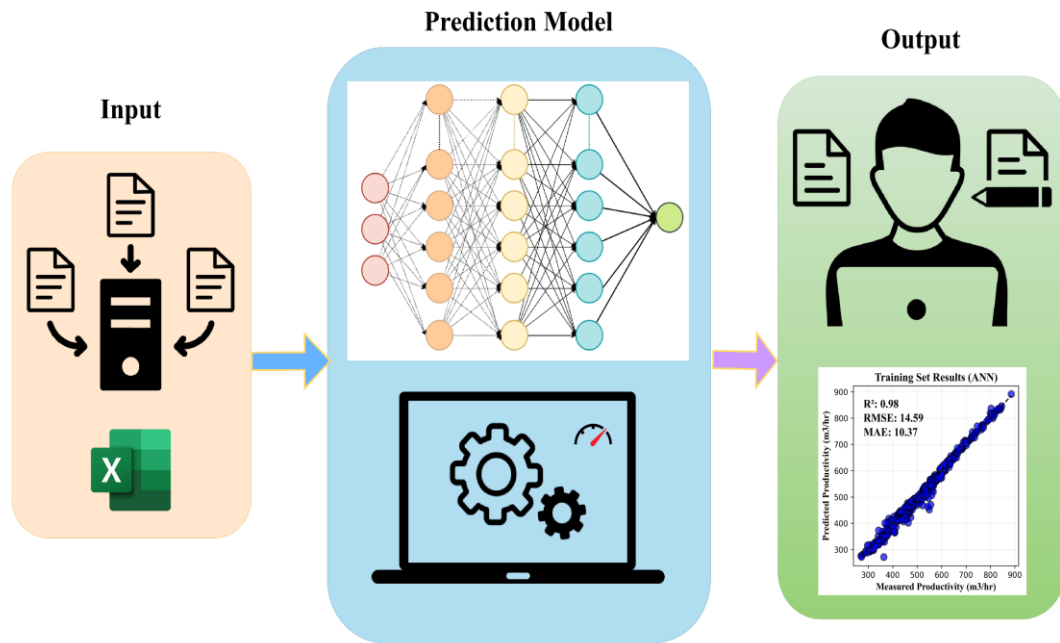


Figure 4.18: A Visual Representation of the Platform Workflow

Platform Surfaces: The platform consists of two main user interfaces: the Upload Page as shown in Figure 4.19 and the Result Page Figure 4.20, described as follows:

1. **Upload Page:** The Upload Page serves as the entry point for users. It includes a clean and structured layout where users can upload their datasets. A prominent note is displayed, specifying that the platform is calibrated for datasets with *Dozer Type D155A*. This ensures users are guided appropriately during data preparation. The page also features institutional branding, including the KFUPM logo and Vision 2030 imagery, reinforcing the platform's alignment with institutional and national priorities.

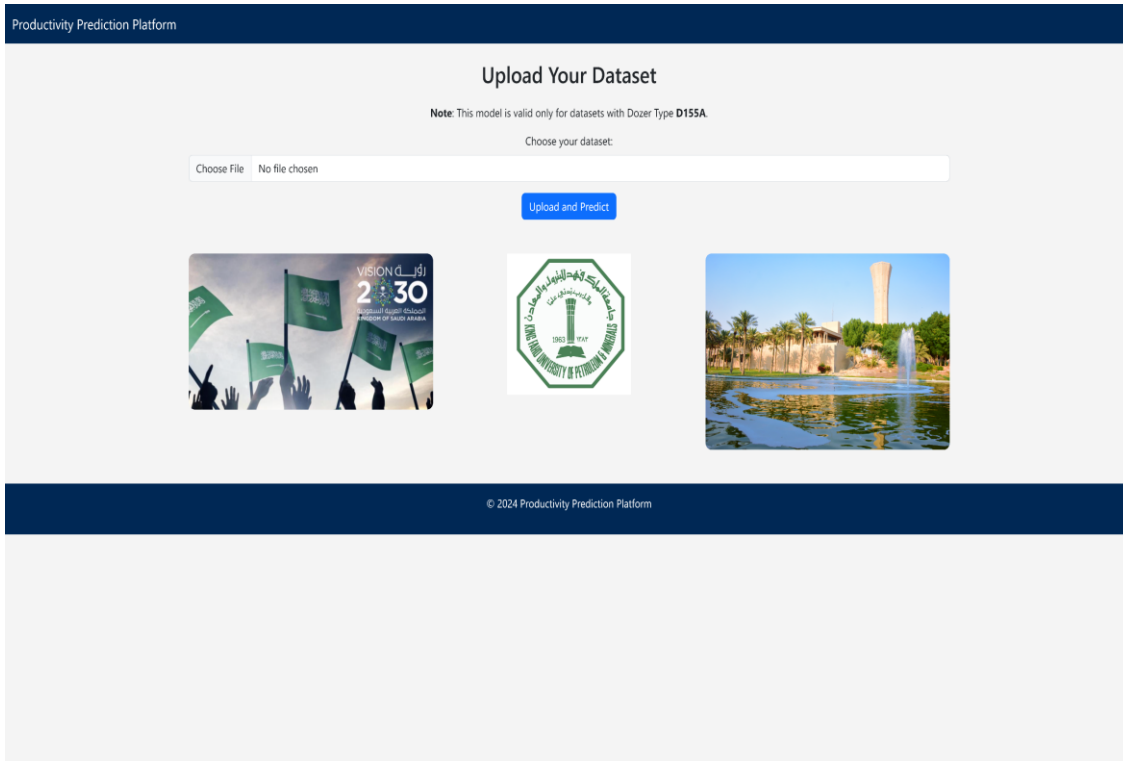


Figure 4.19: The Platform’s Upload Interface

2. **Result Page:** After processing, users are directed to the Result Page, which displays the platform's outputs. The page includes:
 - i. **Metrics Summary:** Key model evaluation metrics (R^2 , MAE, RMSE) are presented in an elegant card layout for instant comprehension.
 - ii. **Visual Elements:** Institutional branding and supportive graphics provide a professional and engaging interface.
 - iii. **Detailed Results:** A tabular display of input parameters, adjusted factors, measured productivity, and predicted productivity. This enables users to review the outputs comprehensively and interpret the results in the context of their datasets.

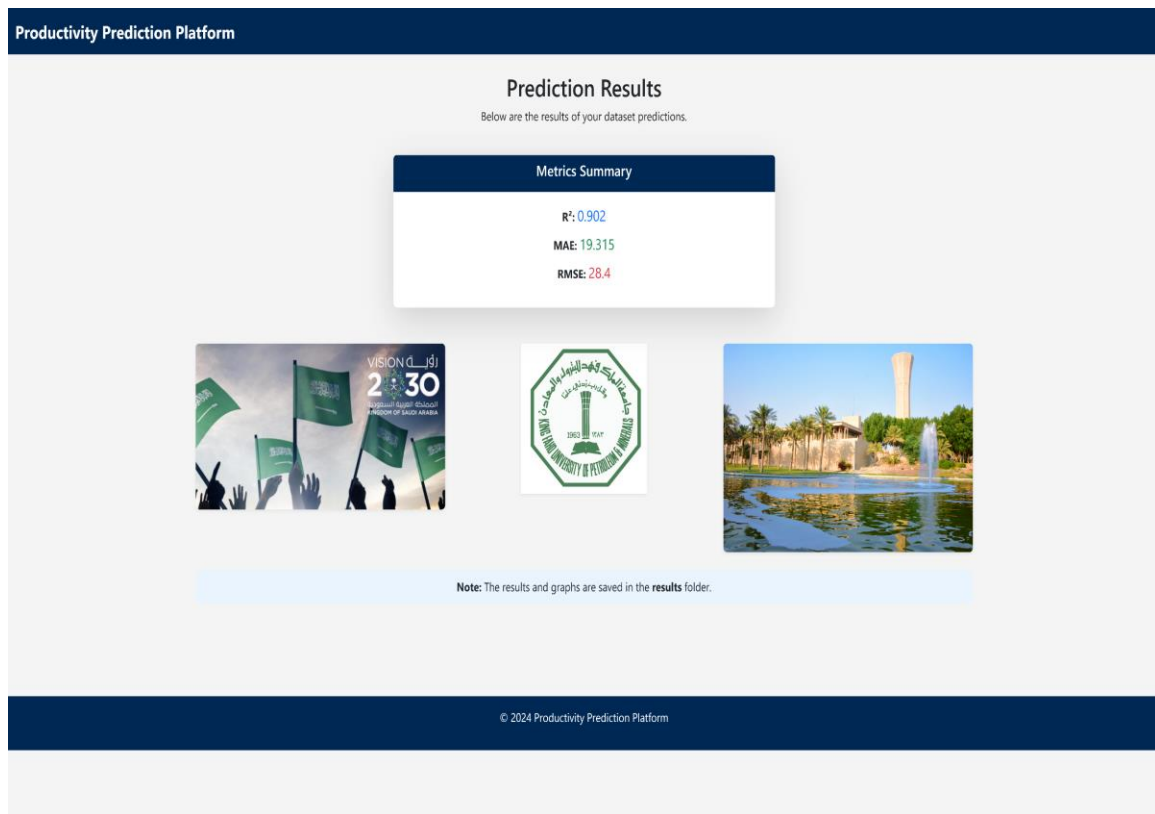


Figure 4.20: The Platform’s Result Interface

These pages, combined with the platform's robust architecture, ensure functionality, reliability, and a professional presentation, aligning with the research objectives of developing an innovative and user-friendly prediction tool.

4.10 Discussion

The results of this study underscore the potential of machine learning, particularly the ANN model, as a powerful tool for predicting bulldozer productivity in construction projects. This section explores the comparative performance of the models, practical implications

for construction management, alignment with industry trends and existing research, and the limitations and future directions for further enhancing the model. Together, these insights provide a comprehensive view of how predictive models can inform and transform productivity management in the construction industry.

4.10.1 Comparative Analysis of Model Performance

The comparative analysis of the ANN, SVM, and RF models highlights the ANN model's strong performance in predicting bulldozer productivity. The ANN achieved the highest R^2 score and lowest MAE and RMSE, indicating a superior ability to capture complex, non-linear relationships within the dataset. These results suggest that the ANN model's layered architecture and capacity for learning intricate patterns allow it to interpret productivity variations more accurately than the other models.

In contrast, the SVM model, while effective for certain non-linear patterns due to its RBF kernel, struggled to accurately predict higher productivity values, which may stem from its sensitivity to high-dimensional data and inability to fully capture complex interactions among productivity factors. The Random Forest model, known for its stability and ability to handle missing data, also fell short in predicting extreme productivity values, tending to “average out” predictions in high-productivity scenarios. This smoothing effect is inherent to ensemble models like Random Forest, which aggregates predictions from multiple trees and may sometimes sacrifice specificity for general accuracy. Overall, the ANN's performance suggests that neural networks, with their depth and flexibility, are particularly well-suited for applications where multiple factors interact dynamically to influence outcomes.

These comparative findings not only reinforce the ANN model's status as the primary predictive tool in this study but also illustrate how machine learning approaches can be tailored based on the complexity of the data and the specific needs of the application. For construction productivity predictions, models capable of capturing subtle, non-linear relationships such as the ANN are likely to yield the most reliable results, especially in scenarios involving varied site conditions and operator influences.

4.10.2 Practical Implications for Construction Management

The high accuracy of the ANN model in predicting productivity has profound implications for construction management, particularly in enabling project managers to make informed, data-driven decisions. This capability not only improves operational efficiency but also enhances resource management, ensures accurate estimation, and strengthens overall project control. The model highlights cycle time, operator skill, and soil type as the most influential factors in productivity, providing actionable insights for construction managers to focus their efforts. For instance, optimizing cycle times by reducing idle periods, assigning skilled operators to handle complex tasks, and tailoring bulldozer settings to suit specific soil conditions can lead to substantial productivity gains. Moreover, these insights support better resource allocation, such as scheduling equipment more effectively, predicting potential bottlenecks, and preemptively addressing issues that could delay progress. By leveraging these findings, project managers can achieve more accurate project estimates, enhance timeline adherence, and maintain tighter control over costs and resources, ultimately driving improved project outcomes.

Moreover, by using the ANN model within a web-based platform, construction managers can predict productivity under different conditions before they occur. This ability to simulate various scenarios—such as the impact of adverse weather or changes in operator experience empowers managers to proactively adjust resources, labor allocation, and equipment deployment in response to anticipated challenges. Rather than reacting to productivity issues as they arise, managers can implement preventive strategies that mitigate delays, optimize resource usage, and reduce project costs. This predictive, forward-looking approach represents a departure from traditional reactive management styles, allowing companies to capitalize on productivity forecasting as a strategic asset.

Furthermore, the model's practical applications extend beyond on-site adjustments to inform long-term planning and resource allocation. By analyzing productivity forecasts for different project phases, managers can anticipate high or low productivity periods, allowing for efficient scheduling of equipment and personnel. The insights gained from these forecasts support a more strategic approach to project planning, helping firms to better align project schedules with resource availability, minimize idle time, and ultimately enhance project timelines. The ANN model's accuracy in productivity predictions, particularly under typical operational conditions, thus serves as a valuable asset in both day-to-day and long-term project planning, offering managers a clearer understanding of how to navigate productivity challenges effectively.

4.10.3 Alignment with Existing Research and Industry Trends

The findings in this study align well with existing literature and industry trends, further validating the model's practical value. Research has consistently highlighted the influence

of operator skill, equipment specifications, and site conditions on construction productivity, and these findings are echoed in the model's sensitivity to factors like operator skill and soil type. Previous studies have demonstrated that experienced operators contribute to higher productivity through their efficiency in handling equipment, which correlates with the ANN model's identification of operator skill as a top variable. This alignment not only reinforces the model's reliability but also highlights the importance of training and skill development as key drivers of productivity in construction.

The emphasis on site-specific adjustments, particularly in soil type and terrain conditions, mirrors existing studies on the need for adaptive productivity strategies. Different soil compositions, such as sand, clay, and rock, require distinct operational approaches due to their varying resistance levels, which influence fuel consumption and cycle times. By capturing these variations in its predictions, the model supports the industry's shift toward a more data-driven approach to construction management, where adjustments are based on concrete productivity forecasts rather than estimates or assumptions.

Moreover, the integration of machine learning models into construction management reflects broader industry trends toward "smart construction" practices. The construction industry has increasingly recognized the value of predictive modeling and data analytics in project management, and the application of this ANN-based model aligns with that trend. By offering a reliable predictive tool, this model contributes to the growing ecosystem of data-driven technologies that aim to streamline construction operations, improve efficiency, and reduce costs.

4.11 Chapter Summary

Chapter 4 has provided an in-depth evaluation of machine learning models for predicting bulldozer productivity in construction, highlighting the potential of the ANN model as a robust predictive tool. Through comparative analysis, sensitivity exploration, practical application, and a thorough discussion of the results, this chapter underscores the model's strengths, its applicability to real-world construction management, and areas for future enhancement.

The comparative analysis of the ANN, SVM, and RF models demonstrated that the ANN model consistently outperformed the others, achieving the highest R^2 score and lowest error metrics. Its ability to capture complex, non-linear relationships among input variables underscores the ANN model's suitability for dynamic and multifaceted productivity data, where numerous factors ranging from operator skill to cycle time and soil type interact in unpredictable ways. The SVM and RF models, while useful for simpler data structures, were limited in capturing extreme productivity values and subtle interactions. This analysis reaffirms that deep learning architectures, such as the ANN, offer unique advantages for construction productivity forecasting, where accuracy and adaptability are paramount.

In terms of practical application, integrating the ANN model into a web-based platform allows for productivity predictions that can be accessed in real-time on construction sites. This capability enables managers to anticipate productivity outcomes under various conditions, adjust resources proactively, and plan contingencies for potential productivity bottlenecks. The ability to simulate different scenarios—such as adverse weather or high operator turnover—demonstrates a shift from reactive management to predictive, strategic

planning. This application is particularly valuable in large-scale projects where optimizing equipment deployment and scheduling based on reliable predictions can lead to substantial cost savings and efficiency improvements.

The discussion further contextualizes the model's findings within the construction industry's current trends towards data-driven management. The results align with existing literature that emphasizes the role of operator skill, site conditions, and equipment specifications in driving productivity. By confirming these factors and demonstrating their predictive power within a machine learning framework, this research contributes to the industry's evolving landscape, where technology and data analytics play an increasingly central role in project optimization. Moreover, the model supports the "smart construction" movement, equipping managers with an accessible, predictive tool that enhances decision-making and operational control.

While the ANN model demonstrates high predictive accuracy, this chapter also addressed the model's limitations. The dataset's limited geographic and project diversity suggests that the model may require further refinement to generalize effectively to diverse construction environments. Additionally, the ANN model occasionally underpredicts extreme productivity values, indicating that it may struggle with rare, high-productivity scenarios. Future research, therefore, should focus on expanding data collection across different regions and project types, as well as enhancing the model to better handle outliers. Such improvements would strengthen the model's flexibility and ensure it remains relevant across varied conditions.

In conclusion, Chapter 4 establishes the ANN model as a powerful predictive tool for construction productivity, offering actionable insights, practical applications, and alignment with broader industry trends. By addressing key limitations and identifying areas for future research, this chapter provides a foundation for ongoing model improvements. The findings reinforce the potential of machine learning to transform productivity management, supporting construction companies in adopting predictive, data-driven strategies that drive efficiency, reduce costs, and improve project outcomes.

Building on these results, the following chapter Conclusion will integrate the key insights from this study, drawing together the model's strengths, practical contributions, and limitations. This concluding chapter will offer recommendations for further model development and implementation, highlighting the potential impact of predictive tools on construction productivity and providing strategic guidance for industry adoption.

CHAPTER 5

CONCLUSION

5.1 Conclusion

This research has successfully demonstrated the potential of predictive modeling to enhance productivity management in the construction industry, specifically focusing on bulldozer operations in Saudi Arabian projects. By addressing key research objectives, the study identified critical productivity factors, developed a robust ANN model, and proposed specifications for a web-based productivity management platform. Key findings include the identification of cycle time, operator skill, soil type, and weather conditions as the most influential factors in bulldozer productivity. The ANN model outperformed other approaches, such as SVM and RF, achieving the highest R^2 and lowest error metrics (RMSE and MAE), thus confirming its effectiveness in capturing complex, non-linear relationships between productivity variables. The web-based platform, designed to integrate the ANN model, enables real-time productivity forecasting and scenario analysis, offering construction managers a practical tool for proactive decision-making and resource optimization.

The first objective was to identify the factors most affecting the productivity of heavy construction equipment. The current findings add to a growing body of literature achieved through expert consultations, which allowed for a rich understanding of the variables influencing bulldozer productivity factors that affect construction projects. Cycle time was

found to be indicating that optimizing the time taken for each operational cycle could yield significant productivity improvements. Operator skill, emphasizes the role of trained personnel in achieving high productivity rates. Soil type, with its direct impact on equipment movement and fuel consumption, emerged as a critical environmental factor, underscoring the need for site-specific adjustments. These insights equip construction managers with targeted areas to focus on, allowing them to prioritize improvements that have the highest potential for productivity increases.

The second objective involved the development of an ANN model to predict productivity. Through iterative testing and evaluation, the ANN model demonstrated a strong ability to accurately predict bulldozer productivity, outperforming other models such as SVM and RF. The model achieved the highest R^2 value and the lowest error metrics (MAE and RMSE) across the dataset, confirming its suitability for capturing the complex, non-linear relationships inherent in construction productivity data. The ANN model's performance underscores its hardiness and adaptability, as it effectively manages the dynamic interactions between variables like cycle time, operator skill, and environmental factors. By learning from diverse operational scenarios, the model proved reliable for predicting productivity across a range of conditions. This finding suggests that the ANN model is well-suited for real-time productivity estimation, providing managers with accurate, timely insights to make informed operational decisions. The model's success aligns with the construction industry's need for data-driven solutions that can handle complexity and variability, reinforcing its practical value.

The third objective focused on developing the specifications for a web-based system that would utilize the ANN model to estimate productivity in real time on construction sites. Based on industry needs and user requirements identified through consultations with construction professionals, this platform was designed to integrate productivity forecasting, scenario analysis, and real-time decision-making support. This web-based system is intended to serve as a bridge between predictive analytics and practical application, helping managers optimize productivity and address potential challenges before they arise.

The current investigation was limited by the dataset primarily consisting of a limited number of projects, which may restrict the model's generalizability across diverse construction environments. Additionally, variations in bulldozer types, soil conditions, and site characteristics were underrepresented, potentially limiting the model's ability to predict productivity under unique operational scenarios. The model's performance is also highly dependent on the quality and timeliness of real-time data inputs, such as cycle times, fuel consumption, and weather conditions, emphasizing the need for robust data validation protocols and advanced tracking systems. These limitations highlight the importance of expanding future data collection efforts to include a broader range of projects, diverse equipment types, and varied soil conditions, thereby improving the model's adaptability and reliability.

To further refine this approach, future research should focus on collecting data from diverse construction sites and incorporating additional variables such as equipment maintenance status, detailed weather conditions, and soil resistance. Furthermore, incorporating

reinforcement learning techniques may allow the model to dynamically adapt to real-time conditions, making it more responsive to operational changes. Expanding the platform's scope to include other heavy equipment types, such as excavators and loaders, could provide a comprehensive pass productivity management tool, addressing broader industry needs. Finally, user feedback mechanisms could be integrated into the web-based platform to ensure continuous model improvement, aligning predictions with real-world outcomes. These recommendations aim to bridge existing gaps, enhance model performance, and support the construction industry's ongoing transition toward smarter, data-driven decision-making practices.

References:

- [1] M. J. Farrell, “The Measurement of Productive Efficiency,” *J R Stat Soc Ser A*, vol. 120, no. 3, pp. 253–281, May 1957, doi: 10.2307/2343100.
- [2] H. Jang, K. Kim, J. Kim, and J. Kim, “Labour productivity model for reinforced concrete construction projects,” *Construction Innovation*, vol. 11, no. 1, pp. 92–113, 2011, doi: 10.1108/14714171111104655.
- [3] A. Hasan, B. Baroudi, A. Elmualim, and R. Rameezdeen, “Factors affecting construction productivity: a 30 year systematic review,” Aug. 10, 2018, *Emerald Group Holdings Ltd*. doi: 10.1108/ECAM-02-2017-0035.
- [4] I. Naskoudakis and K. Petroutsatou, “A Thematic Review of Main Researches on Construction Equipment over the Recent Years,” in *Procedia Engineering*, Elsevier Ltd, 2016, pp. 206–213. doi: 10.1016/j.proeng.2016.11.611.
- [5] R. L. (Robert L. Peurifoy, C. J. Schexnayder, and Aviad. Shapira, *Construction planning, equipment, and methods*. McGraw-Hill Higher Education, 2006.
- [6] A. Salem, “Study of Factors Influencing Productivity of Hauling Equipment in Earthmoving Projects using Fuzzy Set Theory,” *International Journal of Innovation, Management and Technology*, pp. 151–154, 2017, doi: 10.18178/ijimt.2017.8.2.719.
- [7] J. Kim and S. Chi, “Multi-camera vision-based productivity monitoring of earthmoving operations,” *Autom Constr*, vol. 112, Apr. 2020, doi: 10.1016/j.autcon.2020.103121.
- [8] J. Kim, S. Chi, and J. Seo, “Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks,” *Autom Constr*, vol. 87, pp. 297–308, Mar. 2018, doi: 10.1016/j.autcon.2017.12.016.
- [9] A. Montaser and O. Moselhi, “Truck+ for earthmoving operations,” 2014. [Online]. Available: <http://www.itcon.org/2014/25>
- [10] S. Lek, J. G.-E. modelling, and undefined 1999, “Artificial neural networks as a tool in ecological modelling, an introduction,” *Elsevier*, Accessed: May 28, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0304380099000927>
- [11] X. Yao, “A review of evolutionary artificial neural networks,” *International Journal of Intelligent Systems*, vol. 8, no. 4, pp. 539–567, Jan. 1993, doi: 10.1002/INT.4550080406.

- [12] G. G. Towell and J. W. Shavlik, "Knowledge-based artificial neural networks," *Artif Intell*, vol. 70, no. 1–2, pp. 119–165, Oct. 1994, doi: 10.1016/0004-3702(94)90105-8.
- [13] S. Golnaraghi, Z. Zangenehmadar, O. Moselhi, S. Alkass, and A. R. Vosoughi, "Application of Artificial Neural Network(s) in Predicting Formwork Labour Productivity," *Advances in Civil Engineering*, vol. 2019, 2019, doi: 10.1155/2019/5972620.
- [14] A. Rashidi, H. R. Nejad, and M. Maghiar, "Productivity estimation of bulldozers using generalized linear mixed models," *KSCE Journal of Civil Engineering*, vol. 18, no. 6, pp. 1580–1589, 2014, doi: 10.1007/s12205-014-0354-0.
- [15] L. Flood, N. Kartam, and A. Members, "NEURAL NETWORKS IN CIVIL ENGINEERING. h PRINCIPLES AND UNDERSTANDING."
- [16] B. Li-Chung Chao, S. Member, and M. J. Skibniewski, "ESTIMATING CONSTRUCTION PRODUCTIVITY" NEURAL-NETWORK-BASED APPROACH."
- [17] J. J. Shi, "A neural network based system for predicting earthmoving production," *Construction Management and Economics*, vol. 17, no. 4, pp. 463–471, 1999, doi: 10.1080/014461999371385.
- [18] C. M. Tam, O. K. L. Tong, S. Haron, and L. Tse, "Artificial neural networks model for predicting excavator productivity."
- [19] K. Schabowicz and B. Hoła, "Application of artificial neural networks in predicting earthmoving machinery effectiveness ratios," *Archives of Civil and Mechanical Engineering*, vol. 8, no. 4, pp. 73–84, 2008, doi: 10.1016/s1644-9665(12)60123-x.
- [20] A. Rashidi, S. M. Asce, ; Fateme Jazebi, I. Brilakis, and M. Asce, "Neurofuzzy Genetic System for Selection of Construction Project Managers", doi: 10.1061/ASCECO.1943-7862.0000200.
- [21] F. Jazebi and A. Rashidi, "An automated procedure for selecting project managers in construction firms," *Journal of Civil Engineering and Management*, vol. 19, no. 1, pp. 97–106, Feb. 2013, doi: 10.3846/13923730.2012.738707.
- [22] S. C. Ok and S. K. Sinha, "Construction equipment productivity estimation using artificial neural network model," *Construction Management and Economics*, vol. 24, no. 10, pp. 1029–1044, 2006, doi: 10.1080/01446190600851033.

- [23] K. Schabowicz and B. Hola, “Mathematical-neural model for assessing productivity of earthmoving machinery,” *Journal of Civil Engineering and Management*, vol. 13, no. 1, pp. 47–54, 2007, doi: 10.1080/13923730.2007.9636418.
- [24] M. Obach, R. Wagner, H. Werner, H. S.-E. modelling, and undefined 2001, “Modelling population dynamics of aquatic insects with artificial neural networks,” *Elsevier*, Accessed: May 28, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0304380001003076>
- [25] V. K. Ojha, A. Abraham, and V. Snášel, “Metaheuristic design of feedforward neural networks: A review of two decades of research,” *Eng Appl Artif Intell*, vol. 60, pp. 97–116, Apr. 2017, doi: 10.1016/J.ENGAPPAI.2017.01.013.
- [26] P. Dhruv and S. Naskar, “Image Classification Using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN): A Review,” *Advances in Intelligent Systems and Computing*, vol. 1101, pp. 367–381, 2020, doi: 10.1007/978-981-15-1884-3_34.
- [27] A. Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network,” *Physica D*, vol. 404, p. 132306, Mar. 2020, doi: 10.1016/J.PHYSD.2019.132306.
- [28] Y. Yu, X. Si, C. Hu, and J. Zhang, “A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures,” *Neural Comput*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi: 10.1162/NECO_A_01199.
- [29] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, “Generative Adversarial Networks: An Overview,” *IEEE Signal Process Mag*, vol. 35, no. 1, pp. 53–65, Jan. 2018, doi: 10.1109/MSP.2017.2765202.
- [30] J. Vesanto and E. Alhoniemi, “Clustering of the self-organizing map,” *IEEE Trans Neural Netw*, vol. 11, no. 3, pp. 586–600, May 2000, doi: 10.1109/72.846731.
- [31] M. Šopić, M. Vukomanović, D. Car-Pušić, and I. Završki, “Estimation of the excavator actual productivity at the construction site using video analysis,” *Organization, Technology and Management in Construction*, vol. 13, no. 1, pp. 2341–2352, Jan. 2021, doi: 10.2478/otmcj-2021-0003.
- [32] C. Chen, Z. Zhu, and A. Hammad, “Automated excavators activity recognition and productivity analysis from construction site surveillance videos,” *Autom Constr*, vol. 110, Feb. 2020, doi: 10.1016/j.autcon.2019.103045.
- [33] B. Hola and K. Schabowicz, “Estimation of earthworks execution time cost by means of artificial neural networks,” Aug. 2010. doi: 10.1016/j.autcon.2010.02.004.

- [34] B. Sherafat *et al.*, “Automated Methods for Activity Recognition of Construction Workers and Equipment: State-of-the-Art Review,” *J Constr Eng Manag*, vol. 146, no. 6, 2020, doi: 10.1061/(asce)co.1943-7862.0001843.
- [35] H.-S. Park, “Conceptual Framework of Construction Productivity Estimation,” 2006.
- [36] S. Durdyev, S. Ismail, and N. Kandymov, “Structural Equation Model of the Factors Affecting Construction Labor Productivity,” *J Constr Eng Manag*, vol. 144, no. 4, Apr. 2018, doi: 10.1061/(asce)co.1943-7862.0001452.
- [37] A. Panas and J. P. Pantouvakis, “Comparative analysis of operational coefficients’ impact on excavation operations,” *Engineering, Construction and Architectural Management*, vol. 17, no. 5, pp. 461–475, 2010, doi: 10.1108/09699981011074565.
- [38] H. Cai, A. Kandil, M. Hastak, and P. S. Dunston, *Construction Research Congress 2012*. American Society of Civil Engineers, 2012. doi: 10.1061/9780784412329.
- [39] Saudi Vision 2030 Program, “Vision 2030: Saudi Arabia’s Economic Transformation,” 2016.
- [40] A. S. Rao *et al.*, “Real-time monitoring of construction sites: Sensors, methods, and applications,” Apr. 01, 2022, *Elsevier B.V.* doi: 10.1016/j.autcon.2021.104099.
- [41] A. Rossi *et al.*, “Embedded smart sensor device in construction site machinery,” *Comput Ind*, vol. 108, pp. 12–20, Jun. 2019, doi: 10.1016/j.compind.2019.02.008.
- [42] J. Tang *et al.*, “Full-body pose estimation for excavators based on data fusion of multiple onboard sensors,” *Autom Constr*, vol. 147, no. November 2022, p. 104694, 2023, doi: 10.1016/j.autcon.2022.104694.
- [43] M. Menolotto, D. S. Komaris, S. Tedesco, B. O’flynn, and M. Walsh, “Motion capture technology in industrial applications: A systematic review,” Oct. 01, 2020, *MDPI AG*. doi: 10.3390/s20195687.
- [44] M. Haist *et al.*, “DIGITIZATION OF THE CONCRETE PRODUCTION CHAIN USING COMPUTER VISION AND ARTIFICIAL INTELLIGENCE.” [Online]. Available: <https://www.recycontrol.uni-hannover.de/>
- [45] K. Mochida *et al.*, “Computer vision-based phenotyping for improvement of plant productivity: A machine learning perspective,” Dec. 06, 2018, *Oxford University Press*. doi: 10.1093/gigascience/giy153.

- [46] K. Mochida *et al.*, “Computer vision-based phenotyping for improvement of plant productivity: A machine learning perspective,” *Gigascience*, vol. 8, no. 1, pp. 1–12, 2018, doi: 10.1093/gigascience/giy153.
- [47] Y. J. Park, H. S. Gwak, B. S. Kim, and D. E. Lee, “Stochastic dozer productivity estimation method,” *KSCE Journal of Civil Engineering*, vol. 21, no. 5, pp. 1573–1580, 2017, doi: 10.1007/s12205-016-1596-9.
- [48] A. S. Rao *et al.*, “Real-time monitoring of construction sites: Sensors, methods, and applications,” *Autom Constr*, vol. 136, no. August 2021, p. 104099, 2022, doi: 10.1016/j.autcon.2021.104099.
- [49] Z. Zheng, F. Wang, G. Gong, H. Yang, and D. Han, “Intelligent technologies for construction machinery using data-driven methods,” *Autom Constr*, vol. 147, no. December 2022, p. 104711, 2023, doi: 10.1016/j.autcon.2022.104711.
- [50] A. S. Rao *et al.*, “Real-time monitoring of construction sites: Sensors, methods, and applications,” Apr. 01, 2022, *Elsevier B.V.* doi: 10.1016/j.autcon.2021.104099.
- [51] A. Shapira and B. Lyachin, “Identification and Analysis of Factors Affecting Safety on Construction Sites with Tower Cranes,” *J Constr Eng Manag*, vol. 135, no. 1, pp. 24–33, Jan. 2009, doi: 10.1061/(ASCE)0733-9364(2009)135:1(24)/ASSET/EA0D3A1B-7CFB-4F73-8BAF-2216A34B5A9D/ASSETS/IMAGES/LARGE/1.JPG.
- [52] K. Petroutsatou, I. Ladopoulos, and K. Tsakelidou, “Scientometric Analysis and AHP for Hierarchizing Criteria Affecting Construction Equipment Operators’ Performance,” *Sustainability 2022, Vol. 14, Page 6836*, vol. 14, no. 11, p. 6836, Jun. 2022, doi: 10.3390/SU14116836.
- [53] S. G. Naoum, “Factors influencing labor productivity on construction sites: A state-of-the-art literature review and a survey,” *International Journal of Productivity and Performance Management*, vol. 65, no. 3, pp. 401–421, Mar. 2016, doi: 10.1108/IJPPM-03-2015-0045/FULL/PDF.
- [54] R. Ahmad and S. Kamaruddin, “An overview of time-based and condition-based maintenance in industrial application,” *Comput Ind Eng*, vol. 63, no. 1, pp. 135–149, Aug. 2012, doi: 10.1016/J.CIE.2012.02.002.
- [55] J. P. Kenné, A. Gharbi, and M. Beit, “Age-dependent production planning and maintenance strategies in unreliable manufacturing systems with lost sale,” *Eur J Oper Res*, vol. 178, no. 2, pp. 408–420, Apr. 2007, doi: 10.1016/J.EJOR.2006.01.029.

- [56] A. Ur Rehman and K. Awuah-Offei, "Effect of Bucket Geometry, Machine Variables, and Fragmentation Size on Performance of Rubber-Tired Loaders," *Min Metall Explor*, vol. 39, no. 1, pp. 111–127, Feb. 2022, doi: 10.1007/S42461-021-00497-6/TABLES/10.
- [57] S. Alwi, "Factors influencing construction productivity in the Indonesesian context," *5th EASTS Conference*, 2003.
- [58] S. S. Chandra, S. M. E. Sepasgozar, V. R. P. Kumar, A. K. Singh, L. Krishnaraj, and B. O. Awuzie, "Assessing Factors Affecting Construction Equipment Productivity Using Structural Equation Modeling," *Buildings*, vol. 13, no. 2, pp. 1–15, 2023, doi: 10.3390/buildings13020502.
- [59] A. M. Jarkas and M. Radosavljevic, "Motivational Factors Impacting the Productivity of Construction Master Craftsmen in Kuwait," *Journal of Management in Engineering*, vol. 29, no. 4, pp. 446–454, Oct. 2013, doi: 10.1061/(ASCE)ME.1943-5479.0000160.
- [60] I. Mahamid, "Factors contributing to poor performance in construction projects: studies of Saudi Arabia," *Australian Journal of Multi-Disciplinary Engineering*, vol. 12, no. 1, pp. 27–38, Jan. 2016, doi: 10.1080/14488388.2016.1243034.
- [61] S. Sharma and P. K. Goyal, "Fuzzy assessment of the risk factors causing cost overrun in construction industry," *Evol Intell*, vol. 15, no. 4, pp. 2269–2281, Dec. 2022, doi: 10.1007/S12065-019-00214-9.
- [62] S. Singh, "Factors affecting the productivity of construction operations in the United Arab Emirates," 2010, Accessed: May 27, 2023. [Online]. Available: <https://search.proquest.com/openview/8d2ed13f4cfbf03db23705ac70dfdcc6/1?pq-origsite=gscholar&cbl=18750>
- [63] A. V. Thomas and J. Sudhakumar, "Critical Analysis of the Key Factors Affecting Construction Labour Productivity –An Indian Perspective," <http://dx.doi.org/10.1080/15623599.2013.10878231>, vol. 13, no. 4, pp. 103–125, 2014, doi: 10.1080/15623599.2013.10878231.
- [64] S. Al Hammadi and M. S. Nawab, "Study of Delay Factors in Construction Projects," *IARJSET*, vol. 3, no. 4, pp. 87–93, Apr. 2016, doi: 10.17148/IARJSET.2016.3420.
- [65] A. M. Jarkas and C. G. Bitar, "Factors Affecting Construction Labor Productivity in Kuwait," *J Constr Eng Manag*, vol. 138, no. 7, pp. 811–820, Jul. 2012, doi: 10.1061/(ASCE)CO.1943-7862.0000501.

- [66] A. M. Jarkas, "Factors influencing labour productivity in Bahrain's construction industry," <http://dx.doi.org/10.1080/15623599.2015.1012143>, vol. 15, no. 1, pp. 94–108, Jan. 2015, doi: 10.1080/15623599.2015.1012143.
- [67] G. S. A. Elawi, M. Algahtany, and D. Kashiwagi, "Owners' Perspective of Factors Contributing to Project Delay: Case Studies of Road and Bridge Projects in Saudi Arabia," *Procedia Eng*, vol. 145, no. 480, pp. 1402–1409, 2016, doi: 10.1016/j.proeng.2016.04.176.
- [68] S. Mojahed, F. A.-I. journal of project management, and undefined 2008, "Major factors influencing productivity of water and wastewater treatment plant construction: Evidence from the deep south USA," *Elsevier*, Accessed: May 27, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0263786307001007>
- [69] M. Soham, B. R.-I. J. of E. and, and undefined 2013, "Critical factors affecting labour productivity in construction projects: case study of south Gujarat region of India," *academia.edu*, no. 2, p. 583, 2013, Accessed: May 27, 2023. [Online]. Available: <https://www.academia.edu/download/37484099/D1479042413.pdf>
- [70] R. A. Rivas, J. D. Borcharding, V. González, and L. F. Alarcón, "Analysis of Factors Influencing Productivity Using Craftsmen Questionnaires: Case Study in a Chilean Construction Company," *J Constr Eng Manag*, vol. 137, no. 4, pp. 312–320, Apr. 2011, doi: 10.1061/(ASCE)CO.1943-7862.0000274.
- [71] A. K. Mishra, "Factors Affecting Performance and Time Extension of ongoing Construction Projects under Town Development Fund, Nepal," *Journal of Advanced Research in Construction and Urban Architecture*, vol. 03, no. 04, pp. 7–25, Dec. 2018, doi: 10.24321/2456.9925.201805.
- [72] J. Dai, P. Goodrum, W. M.-C. M. and, and undefined 2007, "Analysis of craft workers' and foremen's perceptions of the factors affecting construction labour productivity," *Taylor & Francis*, vol. 25, no. 11, pp. 1139–1152, 2010, doi: 10.1080/01446190701598681.
- [73] S. G. Naoum, "Factors influencing labor productivity on construction sites: A state-of-the-art literature review and a survey," *International Journal of Productivity and Performance Management*, vol. 65, no. 3, pp. 401–421, Mar. 2016, doi: 10.1108/IJPPM-03-2015-0045/FULL/HTML.
- [74] Y. Alfraidi, M. H. H. Abdelhafez, S. M. Alzahrani, and H. Boussabaine, "Impact of Endogenous Risk Factors on Risk Cost in PPP Projects in Saudi Arabia,"

Engineering, Technology & Applied Science Research, vol. 10, no. 4, pp. 6087–6091, Aug. 2020, doi: 10.48084/ETASR.3462.

- [75] A. Kazaz, E. Manisali, S. U.-J. of civil engineering and, and undefined 2008, “Effect of basic motivational factors on construction workforce productivity in Turkey,” *Taylor & Francis*, vol. 14, no. 2, pp. 95–106, 2008, doi: 10.3846/1392-3730.2008.14.4.
- [76] O. Al Rawi, I. Varouqa, E. A.-T. (IJCIET), and undefined 2021, “Factors influencing the productivity of excavation works for construction projects in Jordan,” *academia.edu* O Al Rawi, IF Varouqa, EAD AmerTechnology (IJCIET), 2021•*academia.edu*, Accessed: May 04, 2024. [Online]. Available: https://www.academia.edu/download/73949822/IJCIET_12_04_005.pdf
- [77] C. J. Bamber, J. M. Sharp, and M. T. Hides, “Factors affecting successful implementation of total productive maintenance: A UK manufacturing case study perspective,” *J Qual Maint Eng*, vol. 5, no. 3, pp. 162–181, 1999, doi: 10.1108/13552519910282601/FULL/XML.
- [78] N. Pradhananga, J. T.-A. in construction, and undefined 2013, “Automatic spatio-temporal analysis of construction site equipment operations using GPS data,” *Elsevier*, Accessed: Mar. 16, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0926580512001537>
- [79] S. R. Wu, G. Shirkey, I. Celik, C. Shao, and J. Chen, “A Review on the Adoption of AI, BC, and IoT in Sustainability Research,” *Sustainability (Switzerland)*, vol. 14, no. 13, 2022, doi: 10.3390/su14137851.
- [80] P. A. de Andrade, A. Martens, and M. Vanhoucke, “Using real project schedule data to compare earned schedule and earned duration management project time forecasting capabilities,” *Autom Constr*, vol. 99, no. November 2018, pp. 68–78, 2019, doi: 10.1016/j.autcon.2018.11.030.
- [81] A. A. A. Abdalgar, “Impact of Modern Construction Equipments Management for Improving Productivity,” no. December, p. 2016, 2017.
- [82] R. Dash, M. McMurtrey, ... C. R.-J. of S., and undefined 2019, “Application of artificial intelligence in automation of supply chain management,” *articlearchives.co*, vol. 14, no. 3, p. 43, 2019, Accessed: May 05, 2024. [Online]. Available: <https://articlearchives.co/index.php/JSIS/article/view/4867>
- [83] A. H. Memon, I. A. Rahman, M. R. Abdullah, A. Asmi, and A. Azis, “Factors Affecting Construction Cost in Mara Large Construction Project: Perspective of Project Management Consultant,” *International Journal of Sustainable*

- Construction Engineering and Technology*, vol. 1, no. 2, pp. 41–54, 2010, Accessed: Nov. 16, 2023. [Online]. Available: <https://penerbit.uthm.edu.my/ojs/index.php/IJSCET/article/view/62>
- [84] S. Dara, “Material and Equipment (M & E) productivity management in the Construction sector,” *Res Sq*, pp. 1–18, 2022.
 - [85] R. Tyagi and K. T. Thomas, “Multiple Safety Equipment’s Detection at Active Construction sites Using Effective Deep Learning Techniques,” in *2022 6th International Conference on Trends in Electronics and Informatics*, Institute of Electrical and Electronics Engineers Inc: ICOEI 2022 - Proceedings, 2022, pp. 1271–1276. doi: 10.1109/ICOEI53556.2022.9777164.
 - [86] B. Sherafat *et al.*, “Automated Methods for Activity Recognition of Construction Workers and Equipment: State-of-the-Art Review,” *J Constr Eng Manag*, vol. 146, no. 6, 2020, doi: 10.1061/(asce)co.1943-7862.0001843.
 - [87] S. G.-T. C. E. Handbook and undefined 2003, “Equipment productivity,” *freeit.free.fr*, Accessed: May 04, 2024. [Online]. Available: <http://freeit.free.fr/The%20Civil%20Engineering%20Handbook,2003/0958%20ch03.pdf>
 - [88] S. Gokhale, “Equipment Productivity,” *The Civil Engineering Handbook*, pp. 85–108, 2020, doi: 10.1201/9781420041217-6.
 - [89] A. Makulsawatudom, ... M. E.-T. journal of, and undefined 2004, “Critical factors influencing construction productivity in Thailand,” *journal.kmutnb.ac.th*, Accessed: May 27, 2023. [Online]. Available: http://www.journal.kmutnb.ac.th/web_old/journal/232910255016131.pdf
 - [90] S. Pindoria, “A Critical Review of Identification of Critical Factors Affecting the Productivity of Construction Equipment.” [Online]. Available: www.ijert.org
 - [91] A. Dimitrov and M. Golparvar-Fard, “Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections,” *Advanced Engineering Informatics*, vol. 28, no. 1, pp. 37–49, 2014, doi: 10.1016/j.aei.2013.11.002.
 - [92] Dr. P. N. Gayatri Mahajan, “Integrating BIM with Digital Technology Trends in the Construction Industry: Implementation Insights for 2023,” *Library Progress International*, vol. 44, no. 3, pp. 20283–20308, Nov. 2024, Accessed: Nov. 18, 2024. [Online]. Available: <https://bpasjournals.com/library-science/index.php/journal/article/view/408>

- [93] T. Omotayo, O. E. Ogunmakinde, T. Egbelakin, and A. Sojobi, "Introduction to Innovations, Disruptions and Future Trends in the Global Construction Industry," *Innovations, Disruptions and Future Trends in the Global Construction Industry*, pp. 1–16, Jun. 2024, doi: 10.1201/9781003372233-1.
- [94] K. K. Naji, M. Gunduz, F. H. Alhenzab, H. Al-Hababi, and A. H. Al-Qahtani, "A Systematic Review of the Digital Transformation of the Building Construction Industry," *IEEE Access*, vol. 12, pp. 31461–31487, 2024, doi: 10.1109/ACCESS.2024.3365934.
- [95] Y. Pan and L. Zhang, "Roles of artificial intelligence in construction engineering and management: A critical review and future trends," *Autom Constr*, vol. 122, p. 103517, Feb. 2021, doi: 10.1016/J.AUTCON.2020.103517.
- [96] S. K. Baduge *et al.*, "Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications," *Autom Constr*, vol. 141, p. 104440, Sep. 2022, doi: 10.1016/J.AUTCON.2022.104440.
- [97] N. Rane, "Integrating Leading-Edge Artificial Intelligence (AI), Internet of Things (IoT), and Big Data Technologies for Smart and Sustainable Architecture, Engineering and Construction (AEC) Industry: Challenges and Future Directions," *SSRN Electronic Journal*, Sep. 2023, doi: 10.2139/SSRN.4616049.
- [98] M. Krzywinski and N. Altman, "Visualizing samples with box plots," *Nat Methods*, vol. 11, no. 2, pp. 119–120, Feb. 2014, doi: 10.1038/NMETH.2813.
- [99] R. McGill, J. W. Tukey, and W. A. Larsen, "Variations of box plots," *American Statistician*, vol. 32, no. 1, pp. 12–16, 1978, doi: 10.1080/00031305.1978.10479236.
- [100] M. Streit and N. Gehlenborg, "Points of View: Bar charts and box plots," *Nat Methods*, vol. 11, no. 2, p. 117, Feb. 2014, doi: 10.1038/NMETH.2807.
- [101] M. Kuhn and K. Johnson, "Data Pre-processing," *Applied Predictive Modeling*, pp. 27–59, 2013, doi: 10.1007/978-1-4614-6849-3_3.
- [102] S. G. K. Patro and K. K. Sahu, "Normalization: A Preprocessing Stage," *IARJSET*, pp. 20–22, Mar. 2015, doi: 10.17148/iarjset.2015.2305.
- [103] C. C. Aggarwal, "Data Preparation," *Data Mining*, pp. 27–62, 2015, doi: 10.1007/978-3-319-14142-8_2.
- [104] M. L. Waskom, "seaborn: statistical data visualization," *J Open Source Softw*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/JOSS.03021.

- [105] W. McKinney, “Data Structures for Statistical Computing in Python,” *scipy*, pp. 56–61, 2010, doi: 10.25080/MAJORA-92BF1922-00A.
- [106] P. Sedgwick, “Pearson’s correlation coefficient,” *BMJ*, vol. 345, no. 7864, Jul. 2012, doi: 10.1136/BMJ.E4483.
- [107] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” *Int J Forecast*, vol. 22, no. 4, pp. 679–688, Oct. 2006, doi: 10.1016/J.IJFORECAST.2006.03.001.
- [108] C. J. Willmott and K. Matsuura, “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,” *Clim Res*, vol. 30, no. 1, pp. 79–82, Dec. 2005, doi: 10.3354/CR030079.
- [109] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature,” *Geosci Model Dev*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, doi: 10.5194/GMD-7-1247-2014.
- [110] A. C. Cameron and F. A. G. Windmeijer, “An R-squared measure of goodness of fit for some common nonlinear regression models,” *J Econom*, vol. 77, no. 2, pp. 329–342, Apr. 1997, doi: 10.1016/S0304-4076(96)01818-0.

Vitae

Habibullah Rahnaward is a dedicated civil engineer and certified PMP professional with a strong focus on construction management and the integration of AI-driven solutions for productivity improvement. His research centers on advancing innovative methodologies in the construction sector to enhance operational efficiency. Through academic and professional efforts, he aims to contribute to sustainable and technologically driven advancements in construction practices.

Publications:

1. *“Machine Learning-Driven Smart Platform for Productivity Estimation of Heavy Equipment in Construction Projects”* (to be submitted to *Automation in Construction Journal*).
2. *“Assessment of Natural Hazards Using Integrated Machine Learning Approaches: A Case Study”*, published in *Natural Hazards Journal*. [Read Article](#)

Name	: Habibullah Rahnaward
Nationality	: Afghan
Date of Birth	: 16/07/1996
Email	: habibullahrahnaward@gmail.com
Address	: Unit 001, Parcel 092, Block 001, 1st District, Herat City, Herat, Afghanistan
Postal Code	: 3002
Academic Background	: Bachelor's Degree in Civil Engineering