B. TECH. PROJECT REPORT

On

MACHINE LEARNING APPROACH TO DETERMINE MECHANICAL PROPERTY OF GEOPOLYMER CONCRETE

By

PITTI PRAVEEN (210004031)

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE,

SIMROL, MADHYA PRADESH, INDIA - 453552

MACHINE LEARNING APPROACH TO DETERMINE MECHANICAL PROPERTY OF GEOPOLYMER CONCRETE

A Project report submitted in partial fulfilment of the Academic requirements for the award of the degrees

of

Bachelor of Technology

In

Civil Engineering

By

PITTI PRAVEEN (210004031)

B-TECH (2021-2025)

Guided by-

Dr. Abhishek Rajput

(Associate Professor, Civil Engineering)

Indian Institute of Technology (IIT) INDORE, SIMROL, MADHYA PRADESH, INDIA-453552

NOVEMBER 2024

PITTI PRAVEEN (210004031)

Date: 20-11-2024

ACKNOWLEDGMENTS

First of all, I would like to give my wholehearted thanks of gratitude to my B.Tech. project supervisor and our Head of the Department, **Dr. Abhishek Rajput,** as well as our Director, **Prof. Suhas Joshi,** who has given me this golden opportunity to work on this wonderful project on a given topic. This project also helps me in doing a lot of research, learning new things, and enhancing the portfolio of my skillset. As a part of this project, I am going through so many new things, and I am really feeling my pleasure to be thankful to them.

1. Secondly, I am thankful to **Mr. Manish Yadav**, Ph.D. scholar in the Department of Civil Engineering, along with my supervisor, **Dr**. **Abhishek Rajput**, who provides a perfect environment for critical thinking and research intelligence. They always remain available for discussions, doubt clearance, and guidance at every part of this project. They constantly encourage us to deal with complexities that occur in the due course of this project and in other prospects of life and career.

Pitti Praveen

(210004031)

B.Tech. IV Year

Discipline of Civil Engineering

Indian Institute of Technology, Indore

CANDIDATES DECLARATION

I hereby declare that the project entitled **"Machine learning approach to determine mechanical property of Geopolymer concrete"** submitted in partial fulfillment for the award of the degree of Bachelor of Technology in 'Civil Engineering' completed under the supervision of **Dr. Abhishek Rajput, Civil Engineering, IIT Indore** is a genuine and authentic work.

Further, I declare that I have not submitted this work for the award of any other degree elsewhere.

P. Praveen (20-11-24)

Pitti Praveen

Signature and name of the student with date

CERTIFICATE BY BTP GUIDE

It is certified that the above statement made by the students is correct to the best of my knowledge.

Dr. Abhishek Rajput, Associate Professor

Signature of BTP Guide(s) with dates and their designation

PREFACE

I feel immensely captivated to present this progress report pertaining to the End-Semester evaluation of my B.Tech. project entitled **“Machine learning approach to determine mechanical property of Geopolymer concrete.”** In this report, I have endeavored to present all the necessary and appropriate things involved in my project.

This report thoroughly outlines the background introduction, novelty, objectives, methodology, and expected outcomes of my project.

Through this report, I have tried to make this project fascinating and easy to understand. Each topic related to the project has been explained enthusiastically in detail to enhance the depth of the reader’s learning experience. I have also included pictures and diagrams related to the project, which is solely my creativity.

It is my privilege that I am working on this project under the guidance of my esteemed supervisor, **Dr. Abhishek Rajput**, whose expertise in this domain is of the utmost importance to me. He also encourages me to deal with complicated topics regarding my project and other prospects in life and career. While doing this project, I am going through many new and interesting things about this domain.

It is sincerely hoped from my side that this project will be potentially helpful to me in enhancing my academic as well as non-academic experiences. I have provided all the information in this report by consulting books, journals, research and review articles, and other useful resources.

Pitti Praveen

(210004031)

B.Tech. IV Year

Discipline of Civil Engineering

Indian Institute of Technology, Indore

ABSTRACT

The production of ordinary Portland cement (OPC), the primary adhesive in conventional concrete, is responsible for approximately 5% of global CO2 emissions. Geopolymer concrete (GPC) has attracted attention due to its capacity to mitigate environmental concerns, particularly global warming. Nevertheless, a significant number of GPC studies do not accurately predict the reduction in global warming. Metakaolin (MK) in geopolymers is a viable substitute for conventional materials. Nevertheless, the mechanical performance of MK-based geopolymer concrete is still being impeded by inconsistent predictive models and research findings. The MK-based GPC cost, CO2 emissions, and compressive strength were estimated using machine learning models that were developed using 1,854 samples in this study. The XGBoost that was optimised outperformed all of the other models that were tested. The training metrics were as follows: RMSE = 2.7128, MAE = 1.6129, MAPE = 5.7491, and R2 = 0.9740. RMSE = 5.9192, MAE = 3.7816, R2 = 0.8775, and MAPE = 14.8391 were the testing metrics. The second-best Gradient Boosting Machine (GBM) had the following metrics: R2 = 0.9758, RMSE = 2.6169, MAE = 1.5617, and MAPE = 5.4596. R2 = 0.8711, RMSE = 6.0717, MAE = 3.9530, and MAPE =14.6354 were attained during the testing process.

The objective function of a multi-objective optimisation framework that employs NSGA-II was the enhanced XGBoost model in order to identify the optimal GPC mix designs. The compressive strength was assigned the greatest weight in the optimisation procedure. The performance of the mixture was substantially influenced by the chemical composition, curing circumstances, coarse-to-fine aggregate ratio (CA/FA), NaOH concentration, and water content, as demonstrated by the SHAP and feature importance analyses.

This investigation demonstrates the potential of machine learning to enhance the accuracy of forecasts, reduce experimental effort, and optimise the compositions of geopolymer concrete based on MK. NSGA-II assists in the identification of cost-effective and sustainable mix designs, thereby enhancing the efficiency of building resources and promoting environmentally responsible construction materials.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 CHAPTER
INTRODUCTION

1.1 OVERVIEW

Cement production contributes 5% to 8% of global emissions, equating to 1.45 ± 0.20 gigatons of CO_2 year (Jiang, 2022). The manufacturing of cement significantly adds to globally CO_2 emissions within the construction sector. Research is being conducted on alternatives to Ordinary Portland Cement (OPC) that are sustainable (Eftekhar Afzali, Shayanfar, Ghanooni-Bagha, Golafshani, & Ngo, 2024). Due to escalating fears around global warming. Geopolymers enhance mechanical qualities like as strength and durability while being environmentally benign. Geopolymers diminish carbon emissions by 22–72%. Civil engineers can now predict and enhance the mechanical properties of novel materials such as geopolymer concrete through machine learning (ML). This has facilitated environmentally sustainable, high-performance construction alternatives.

1.2 BACKGROUND AND MOTIVATION OF STUDY

This study was motivated by the necessity to mitigate the environmental impact of the construction sector, specifically the cement industry, which produces significant CO_2 emissions. Traditional OPC manufacture is energy-consuming and environmentally detrimental. Metakaolin (MK) geopolymers can incorporate industrial waste byproducts and exhibit reduced carbon emissions, rendering them a viable alternative. Geopolymers are resistant to chemical degradation and elevated temperatures. Notwithstanding experimental studies on geopolymer concrete (GPC), a more comprehensive and effective approach is required to improve GPC mix design. Effectively managing the intricate interactions among mixture components is crucial. Eftekhar Afzali, Shayanfar, Ghanooni-Bagha, Golafshani, and Ngo (2024) assert that machine learning can address conventional issues. This instrument facilitates accurate modelling and optimisation.

1.3 OBJECTIVES

Primary Objective: To utilize machine learning to model and predict the compressive strength, CO_2 emissions, and cost of MK-based geopolymer concrete (GPC) using a comprehensive dataset and advanced ML methodologies.

Machine Learning Models: This study employs sophisticated machine learning methodologies, including DT, support vector machines SVM, GBM, compact GBM, random forest models (RF), compact RF, backpropagation neural networks BPNN, and XGBoost, to address the complexities of geopolymer concrete mix design.

Influential Factors: These machine learning models account for various influential factors, including the chemical compositions of the binder and activator, mix design parameters, and curing conditions.

Optimization Goal: To identify key mix design variables and optimize the compressive strength, durability, cost, and environmental impact of MK-based GPC using a multi-objective optimization framework integrated with the XGBoost model.

This research ultimately aims to enhance the understanding and design of sustainable MK-based geopolymer concrete while simplifying experimental efforts and improving resource efficiency.

2 CHAPTER

LITERATURE REVIEW

2.1 OVERVIEW

Several research studies have explored the use of geopolymers as a sustainable substitute for OPC, with a specific focus on MK-based geopolymers. Geopolymers offer several advantages over OPC, including significantly reduced CO_2 emissions, enhanced mechanical properties, and increased resistance to chemical attacks. However, the complexity of geopolymer systems—arising from diverse mix design parameters, curing conditions, and chemical compositions—poses challenges for their widespread adoption. This complexity necessitates the development of robust predictive models capable of addressing the intricate relationships among these variables. Recent studies have leveraged advanced machine learning models to tackle these challenges, demonstrating their potential in optimizing geopolymer concrete formulations for improved compressive strength, environmental sustainability, and cost efficiency. However, further research is required to refine these models and enhance their predictive capabilities for multi-objective optimization tasks in geopolymer concrete design.

2.2 TRADITIONAL EMPIRICAL FORMULA

Empirical models have traditionally been developed to predict the mechanical properties of cement-based materials, including ordinary Portland cement (OPC) and geopolymer concrete. While these models rely on empirical data and straightforward mathematical relationships, they are often inadequate for capturing the complex interactions inherent in advanced materials like geopolymers. Factors such as chemical composition, mix design parameters, curing conditions, and temperature effects introduce intricacies that these models struggle to address effectively. This underscores the need for advanced predictive approaches, such as machine learning, to model and optimize the performance of innovative materials like MK-based geopolymer concrete.

2.3 LIMITATIONS OF TRADITIONAL METHODS

- The limitations of empirical models include their inability to capture the nonlinear interactions between multiple variables in complex systems like geopolymer concrete.

- Empirical methods are both time-consuming and expensive, as they require extensive experimental data to establish reliable relationships.
- The accuracy of these models decreases when applied to materials with novel compositions or unique curing conditions, such as MK-based geopolymers.

2.4 RISE OF AI IN CIVIL ENGINEERING

AI and ML are transforming civil engineering by providing effective tools for modeling and enhancing the characteristics of intricate materials. ML models can manage extensive datasets and capture non-linear relationships between variables, unlike traditional empirical approaches, making them well-suited for predicting the performance of geopolymer concrete. AI has already been utilized in multiple aspects of civil engineering, such as structural health monitoring and material science, and its potential in improving sustainable materials like geopolymers is attracting growing interest.

2.5 WHAT IS GEOPOLYMER CONCRETE

Geopolymer concrete serves as an innovative and sustainable alternative to traditional Ordinary Portland Cement (OPC) concrete by utilizing alumina-silica-rich materials like metakaolin or industrial by-products such as fly ash and ground granulated blast-furnace slag, activated with alkaline solutions. This activation process results in a binder that undergoes a chemical transformation into a three-dimensional network of inorganic polymers, exhibiting exceptional mechanical strength and durability. Geopolymer concrete not only reduces CO_2 emissions associated with cement production but also provides superior resistance to chemical attacks, high temperatures, and environmental degradation. Its use promotes an eco-friendlier approach to construction while ensuring robust and long-lasting structural performance.

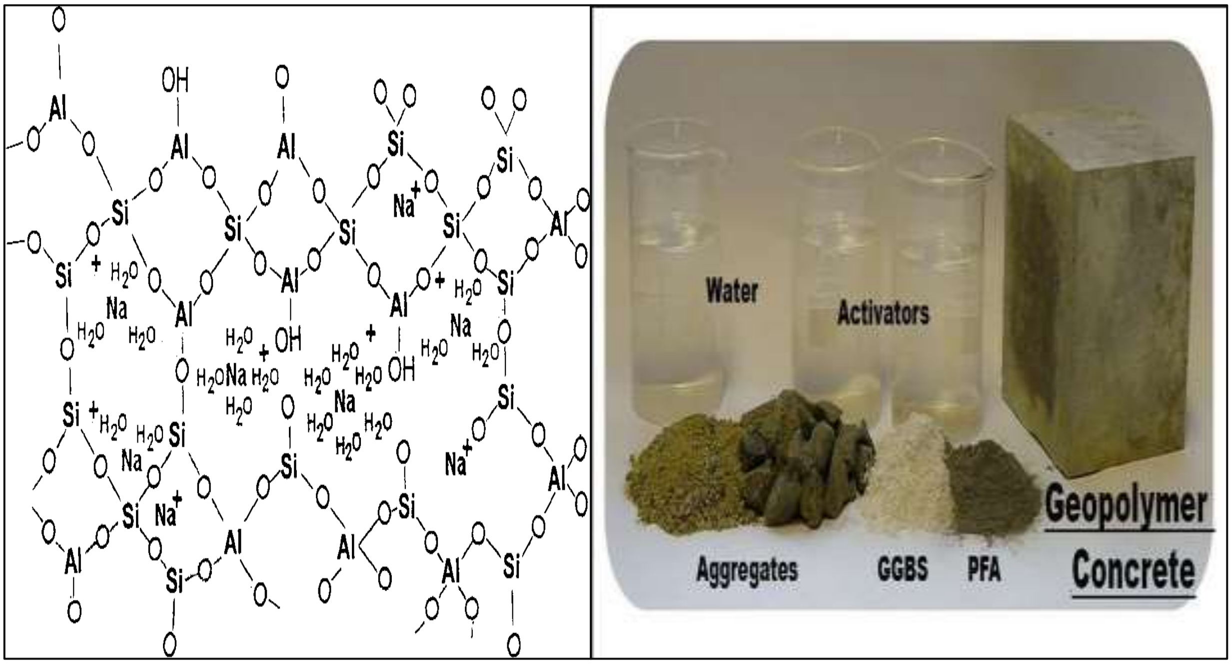


FIGURE 2.1 GEOPOLYMER CONCRETE

2.6 APPLICATION OF AI IN GEOPOLYMER CONCRETE

Numerous studies have leveraged machine learning to predict the mechanical properties of geopolymers. Backpropagation neural networks (BPNNs) have been employed to evaluate the compressive strength of fly ash-based geopolymer concrete, demonstrating superior performance over traditional methods. Additionally, models like random forests (RF) and gradient boosting machines (GBM) have been successfully utilized to predict the compressive strength of alkali-activated materials, relying on input parameters such as the chemical composition of binders and activators. This research investigates the potential of advanced machine learning models, including XGBoost and BPNNs, to enhance the predictive accuracy and optimization of MK-based geopolymer concrete.

2.7 CURRENT GAPS AND FUTURE DIRECTIONS

Despite advancements in AI-driven research on geopolymer concrete, significant challenges remain. Many machine learning models struggle to capture the complex interactions among factors such as chemical composition, mix design parameters, and curing conditions. Furthermore, limited research has been conducted on MK-based geopolymer concrete (GPC), with most studies focusing on fly ash or GGBFS-based systems. This study addresses these gaps by employing advanced machine learning models, including XGBoost, GBM, and BPNN, which account for a comprehensive set of factors influencing MK-based GPC. Additionally, the integration of interpretability tools like Shapley Additive Explanations (SHAP) provides detailed insights into the impact of individual variables on model predictions. This approach aims to deliver more accurate forecasts, optimize mix designs, and promote sustainable construction practices.

3 CHAPTER

MATERIALS AND METHODS

3.1 METHODOLOGY

This study establishes a robust computational framework to accurately predict the compressive strength of metakaolin (MK)-based geopolymer concrete using advanced machine learning models, specifically XGBoost, BPNN, SVM, DT, GBM, RF, Compact GBM, and Compact RF. The dataset, consisting of 1854 data points, includes three output parameters: compressive strength (UCS), CO_2 emissions, and cost. The data is split into training (80%) and testing (20%) subsets, ensuring consistency in model evaluation. Models are trained on the training set and evaluated on the test set using metrics such as R^2, RMSE, MAE, and MAPE. Following evaluation, the models are ranked based on performance, and the best-performing models are selected for further analysis. To optimize the geopolymer mix design, NSGA-II optimization is applied, utilizing the top-performing model, XGBoost, for predictions. Parametric studies are then conducted with the optimized models to gain deeper insights into the relationships between mix design variables and the performance of the geopolymer concrete, including achieving a balance between UCS, CO_2 emissions, and cost.

3.2 DATA COLLECTION

The caliber of the database utilized to predict compressive strength for MK-based geopolymer concrete (GPC) is essential. An extensive experimental database of twenty-five input parameters was created from reliable sources for this study, including parameters such as Fly Ash, GGBFS, RM, SM, AM, and various others related to curing conditions and chemical composition. Only studies featuring comprehensive chemical composition and standardized compression test data were included to ensure database trustworthiness. After eliminating duplicates, superfluous entries, and missing values, the dataset consisted of 1854 valid cases. The incorporation of "pre-curing condition (PCC)" and "curing temperature" addresses inconsistencies in curing methods. Moreover, the chemical composition of MK is contingent upon the ratios of SiO_2 and Na_2O. Table 1 presents the feature statistics. Figure 3 illustrates the distribution patterns of features, highlighting the sporadic application of elevated extra water values and superplasticizer, alongside the predominant testing conducted at temperatures below 25 degrees Celsius. Subsequent research should evaluate specimens exceeding 28 days to enhance predictive models.

TABLE 3.1 SUMMARY STATISTICS

	count	mean	std	min	25%	50%	75%	max
FlyAsh	1854.0	317.86406467098163	153.85637763800926	0.0	276.0	394.29	408.47175	640.0
GGBFS	1854.0	102.14349514563108	170.22153816017686	0.0	0.0	0.0	146.0	560.0
RM	1854.0	0.8745685005393742	0.45209245767459505	0.25	0.6	0.66	1.01	2.08
SM	1854.0	1.6985005393743258	0.6525504909820193	0.82	1.3824999999999998	1.545	1.86	5.43
AM	1854.0	6.233079827400215	5.510502659933104	1.05	2.59	4.23	8.74	44.47
HM	1854.0	0.19661812297734627	0.2520732121430657	0.0	0.03	0.1	0.24	0.84
LM	1854.0	9.031515641855448	11.301567682329585	0.02	1.32	4.71	11.17	38.61
CA	1854.0	1098.4374325782094	244.94935789721404	0.0	1041.0	1155.0	1209.0	1591.0
FA	1854.0	638.0177993527508	160.29065682105048	0.0	561.0	651.0	721.0	923.0
SS	1854.0	111.73394822006472	35.92196362001775	26.3	94.0	104.12	129.3	365.0
SH	1854.0	63.29402373247033	29.97911286895565	13.0	42.0675	55.0	70.0	232.18
SS/SH	1854.0	2.042729234088457	0.8707088099158893	0.18	1.2825	2.49	2.5	14.04
Mol_SH	1854.0	11.804392623117582	2.9550466581072476	3.0	10.0	12.0	14.0	20.0
SS_SiO2/Na2O	1854.0	2.1703128371089533	0.4719608192108726	0.89	2.0	2.06	2.5	3.5
SH_Na2O %	1854.0	13.660922330097087	3.719180715305545	3.58	11.4	14.36	14.73	37.2
SH_SiO2 %	1854.0	28.25351132686084	4.948448449989449	7.68	26.5	29.4	29.93	39.01
SH_H2O %	1854.0	55.210404530744334	8.310849403468012	14.87	55.52	55.9	60.1	65.0
WEff	1854.0	119.19725997842502	41.564141404149815	41.38	89.19	104.55	142.6175	256.8
W/B	1854.0	0.4371952535059331	0.1249380101405908	0.17	0.36	0.42	0.5075000000000001	1.02
Ti (hr.)	1854.0	27.243797195253507	80.47320854084923	0.0	0.0	24.0	24.0	672.0
Cti (°C)	1854.0	53.20550161812298	25.395933315614236	20.0	25.0	60.0	70.0	120.0
Cm	1854.0	0.8387270765911543	0.9288734758336364	0.0	0.0	0.0	2.0	2.0
RH(%)	1854.0	69.17691477885653	8.826625690222864	33.0	70.0	70.0	70.0	100.0
CTf(°C)	1854.0	25.14940668824164	5.720397147278222	20.0	24.0	24.0	25.0	90.0
Age(Days)	1854.0	19.166666666666668	27.326546637039854	1.0	7.0	7.0	28.0	365.0
UCS	1854.0	38.37324703344121	15.822438128448473	1.0	28.0	38.0	49.0	89.0

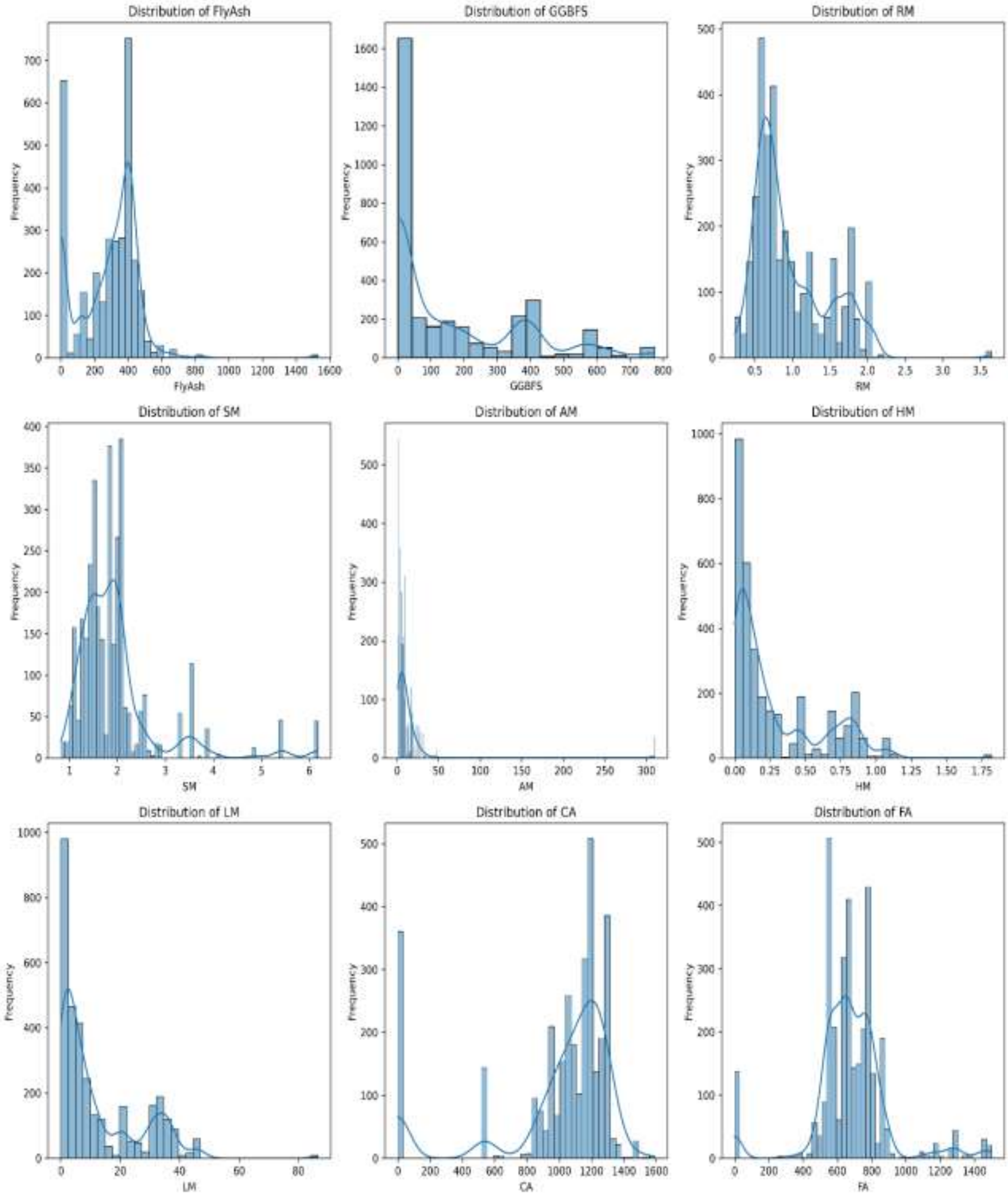


FIGURE 3.1 DISTRIBUTION OF THE INPUT FEATURES.

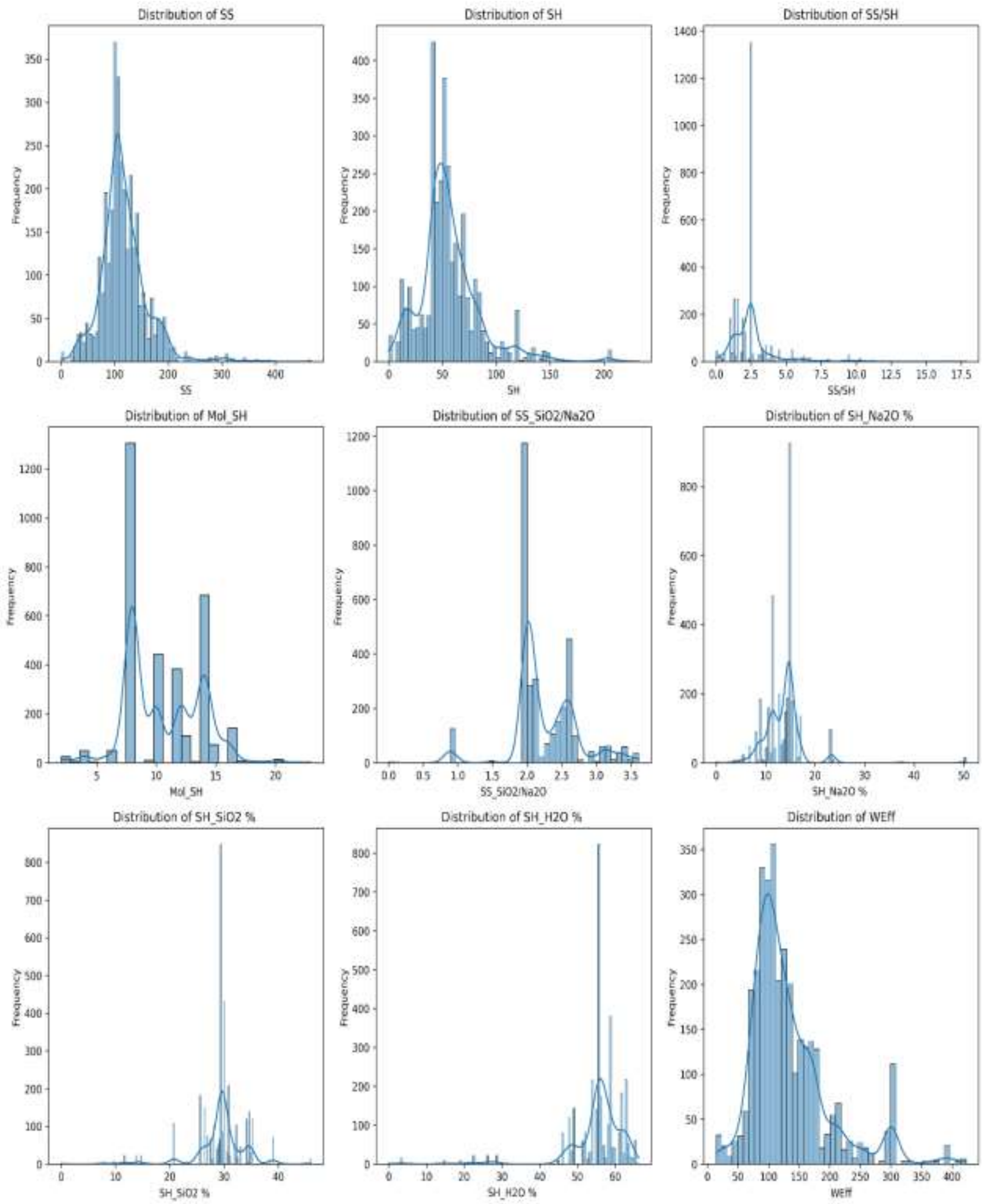


FIGURE 3.2 DISTRIBUTION OF INPUT FEATURES

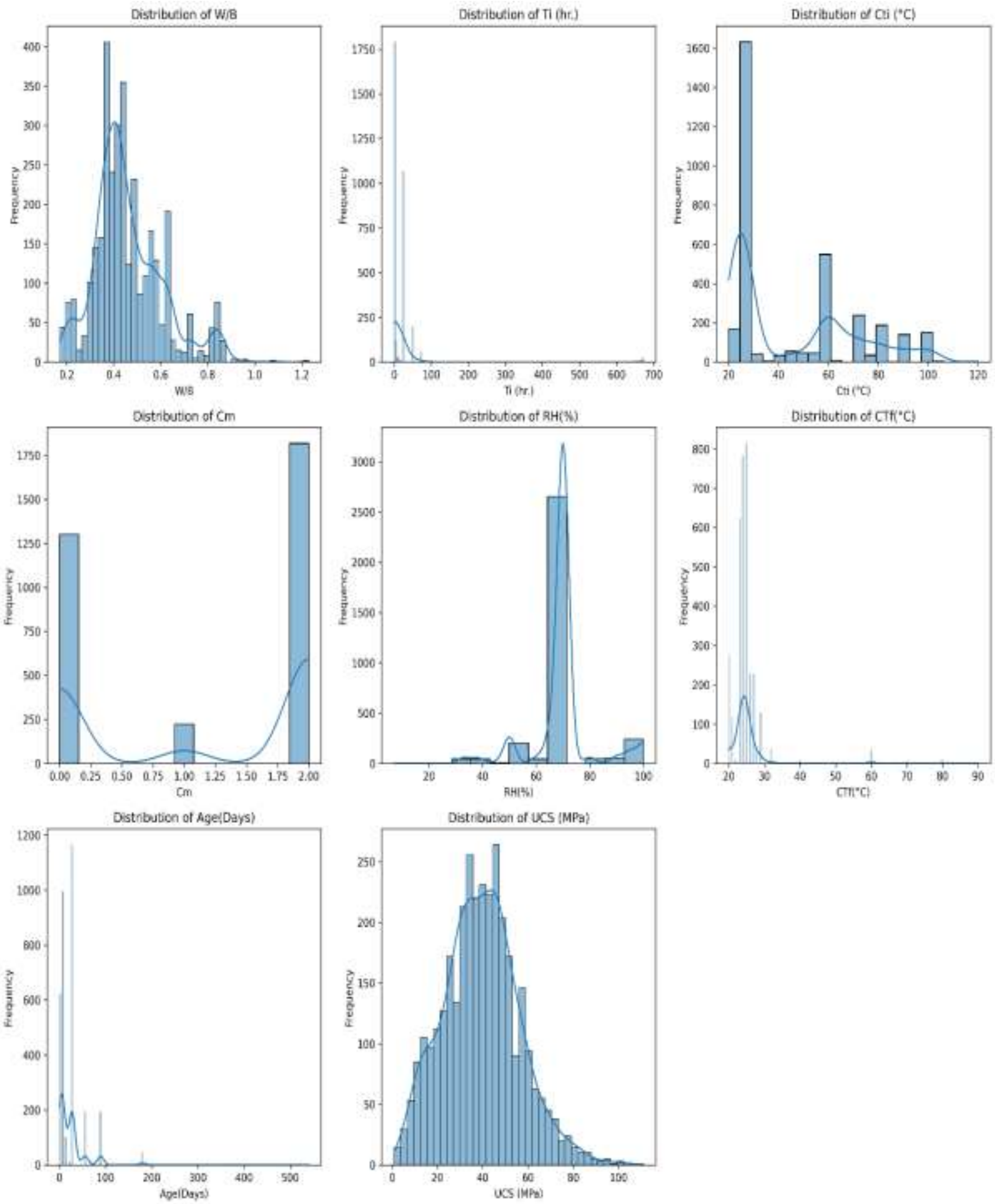


FIGURE 3.3 DISTRIBUTION OF INPUT FEATURES

3.3 HYPER-PARAMETER TUNING

To enhance the performance and reliability of machine learning models, hyperparameter tuning was conducted using Bayesian Optimization coupled with k-fold cross-validation. This systematic approach ensures optimal model configurations while minimizing overfitting and underfitting risks.

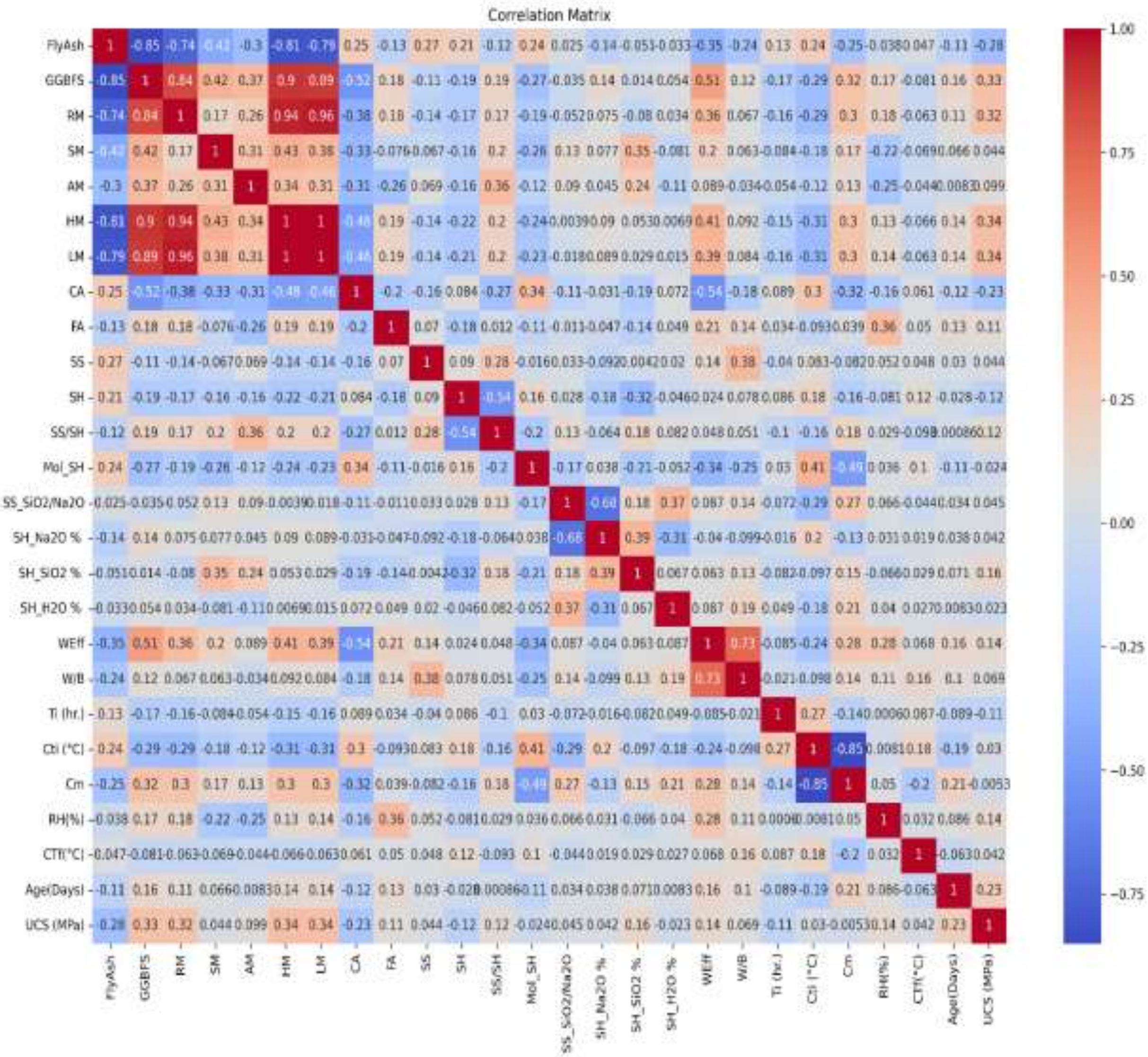


FIGURE 3.4 CORRELATION MATRIX

Bayesian Optimization was employed to identify the best hyperparameters by iteratively refining them based on model performance. The method evaluates a probabilistic model of the objective function, balancing exploration of new hyperparameter spaces and exploitation of known optimal regions. This technique significantly reduces computational effort compared to grid or random search methods, providing efficient parameter optimization.To validate the robustness of the tuned models, k-fold cross-validation was applied. The dataset was split into 5 equal subsets, with 4 folds used for training and one-fold for validation in each iteration. The KFold class from the scikit-learn library was used, ensuring shuffling and consistent random state for reproducibility. A Python function, kfold_cross_val, was implemented to compute cross-validation scores, including training and test R^2 values for each model.

This combined methodology was applied to all models, including XGBoost, Gradient Boosting Machine (GBM), Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and BPNN. The resulting tuned hyperparameters significantly improved performance, as evidenced by lower RMSE and MAE, and higher R2R^2R2 values across training and testing datasets. The framework's robustness and scalability ensure that the models generalize well for predicting compressive strength, cost, and CO_2 emissions of MK-based geopolymer concrete.

3.4 MACHINE LEARNING MODELS

3.4.1 Gradient Boosting Machine

Gradient Boosting Machine is a proficient ensemble learning technique that incrementally constructs decision trees to rectify the faults of preceding trees. A sequential method diminishes the loss function and enhances predictive accuracy at each stage. Gradient Boosting Machines (GBM) are optimal for simulating the compressive strength of geopolymer concrete due to their capability to manage intricate and non-linear relationships. Boosting integrates weak learners, such as decision trees, to enhance predictive accuracy.

Formula: GBM incrementally reduces a differentiable function of loss L (y, F(x)).

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x)$$

Where:

- $F_m(x)$ is the current model at iteration m
- $F_{m-1}(x)$ is the model from the previous iteration
- η is the learning rate
- $h_m(x)$ is the weak learner (decision tree) added at step m

3.4.2 Compact-GBM

Compact-GBM is a kind of Gradient Boosting Machine. This modification decreases model size while maintaining forecast accuracy. Typically, it entails pruning or simplifying trees and compressing the model to enhance computational efficiency. This approach is highly beneficial in resource-limited settings or where minimizing model size is paramount. The boosting concept is analogous to GBM, yet tree pruning or parameter optimization refines the model. This technique eliminates trees without compromising model efficacy.

3.4.3 XGBoost

XGBoost also called Extreme Gradient Boosting is a high-performance, scalable machine learning algorithm for regression and classification tasks. It builds an ensemble of decision trees iteratively, where each tree tries to correct the errors of the previous ones using a gradient descent approach on a specified loss function. At each iteration, XGBoost minimizes the regularized objective function, adding new trees that reduce residual errors while controlling overfitting with regularization (γ, λ). This combination of efficiency and accuracy makes XGBoost a leading choice for structured data.

The key objective function in XGBoost is:

$$\mathbf{Obj}(t) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

Terms in the Formula:

- $\ell(y_i, \hat{y}_i)$: The loss function, measuring the difference between actual (y_i) and predicted values ($\hat{y}_i$). For regression, this is often Mean Squared Error:

$$\ell(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$$

- $\hat{y}_i^{(t-1)}$: Prediction from the previous iteration $t - 1$.
- $f_t(x_i)$: The current decision tree's prediction for data point x_i.
- $\Omega(f_t)$: Regularization term to penalize tree complexity, defined as:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

where T is the number of leaves in the tree, w_j is the weight for leaf j, γ controls tree size regularization, and λ penalizes large leaf weights.

3.4.4 Random Forest (RF)

Random Forest integrates many decision trees for ensemble learning. GBM produces trees in a sequential manner, whereas RF develops them independently and simultaneously. A random subset of data and attributes is utilized to train each tree in the forest. This mitigates overfitting and improves model resilience. Random Forest employs the "wisdom of crowds" to deliver precise and reliable predictions.

Formula: A Random Forest's regression prediction is the average, and its classification prediction is a majority decision of individual trees:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x)$$

Where:

- T is the number of trees
- $h_t(x)$ is the prediction from tree t

3.4.5 Compact-RF

Compact-RF is a streamlined version of Random Forest where the number of trees is reduced, or the trees themselves are pruned, leading to a more computationally efficient model. This version sacrifices some flexibility to create a more portable and resource-efficient model without a significant drop in accuracy. Formula **is s**imilar to RF, but with tree pruning and reduced model complexity to shrink the model size.

3.4.6 Decision Tree (DT)

Decision trees (DTs) are a straightforward yet effective paradigm for data classification and regression. A decision tree partitions data based on feature values, resembling a tree structure. The terminal nodes signify the output (class label or regression value), and the internal nodes denote feature-based evaluations. Decision trees systematically partition data to minimise impurity and variation.

Formula: For regression, a decision tree minimizes the sum of squared errors (SSE):

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Where:

- y_i is the actual value
- $\hat{y}_i$ is the predicted value at leaf nodes
- n is the number of samples

3.4.7 Backpropagation Neural Network (BPNN):

A **Backpropagation Neural Network (BPNN)** is a type of artificial neural network that uses the backpropagation algorithm to train its weights. It consists of an input layer, one or more hidden layers, and an output layer. During training, the network uses forward propagation to calculate outputs and backward propagation to adjust weights to minimize error. The backpropagation algorithm uses the chain rule to compute gradients efficiently, propagating errors backward from the output to the input layer to optimize weights for better predictions.

Formula: The weight update formula in backpropagation is:

$$w_{ij} \leftarrow w_{ij} - \eta \frac{\partial E}{\partial w_{ij}}$$

Here:

- w_{ij}: Weight between neuron i in the previous layer and neuron j in the current layer.
- η: Learning rate, controlling the step size for weight updates.
- $\frac{\partial E}{\partial w_{ij}}$: Gradient of the error E with respect to the weight w_{ij}.

The error E is typically calculated using Mean Squared Error (MSE):

$$E = \frac{1}{N} \sum_{k=1}^{N} (y_k - \hat{y}_k)^2$$

Where:

- y_k: Actual target output.
- $\hat{y}_k$: Predicted output by the network.
- N: Total number of samples.

3.4.8 Support Vector Machine (SVM):

Support Vector Machines (SVMs) are supervised learning algorithms employed for classification and regression tasks. In multi-dimensional feature space, support vector machines (SVM) seek the optimal hyperplane to distinguish data points. Regression employs epsilon, a margin of tolerance, to forecast continuous values in proximity to the hyperplane. SVM optimizes the margin between the hyperplane and data points.

Formula: The objective of support vector machine regression, which is also referred to as SVR, is to minimize the amount of error that is subjected to an epsilon-insensitive loss function:

$$L_\epsilon(y, f(x)) = \max(0, |y - f(x)| - \epsilon)$$

Where:

- y is the actual value
- $f(x)$ is the predicted value
- ϵ defines the margin of tolerance

3.5 SHAP ANALYSIS

In this study, feature importance and interpretability of machine learning models were extensively analyzed using Gini importance, permutation importance, and SHAP (SHapley Additive exPlanations) values. Gini importance, derived from tree-based models like Decision Tree, Random Forest, Gradient Boosting, and XGBoost, provided an initial assessment of feature contributions by quantifying their influence on reducing model impurity. The permutation importance, calculated using out-of-bag (OOB) samples from the Random Forest model, further validated the critical features by measuring the decrease in model performance when specific features were shuffled. These analyses revealed the most influential input parameters affecting compressive strength predictions.

To complement these methods, SHAP analysis was employed to gain deeper insights into feature impact at both the global and local levels. SHAP values, computed specifically for the XGBoost model, allowed for the quantification of each feature's contribution to individual predictions. The mean absolute SHAP values highlighted the overall importance of features, while SHAP summary plots provided a detailed visualization of their magnitude, direction, and interaction effects. This approach not only explained the model's decision-making process but also revealed non-linear relationships and dependencies among the features. The combined use of Gini importance, permutation importance, and SHAP analysis ensured a comprehensive understanding of feature relevance, enhancing the interpretability and reliability of the models. Visualizations such as bar plots, summary plots, and permutation importance graphs further strengthened the transparency and effectiveness of the results, which are critical for guiding practical applications and future research in geopolymer concrete mix optimization.

3.6 MULTI OBJECTIVE OPTIMIZATION

The multi-objective optimization process employed in this study aims to optimize the mix design of MK-based geopolymer concrete by minimizing carbon dioxide (CO_2) emissions and cost while ensuring that the compressive strength (UCS) remains close to a target value. Using **NSGA-II** (Non-dominated Sorting Genetic Algorithm II), the problem was formulated with three **objectives:** minimize CO_2 emissions, minimize cost, and penalize deviations from the target UCS. The optimization included 26 input features, such as Fly Ash, GGBFS, molar ratios, and curing conditions, with specified bounds reflecting practical constraints in mix design.

The algorithm did not enforce hard constraints directly but penalized UCS deviations to guide solutions toward meeting target requirements. **NSGA-II** utilized Latin Hypercube Sampling (LHS) for initializing the population, Simulated Binary Crossover (SBX) for generating new solutions, and Polynomial Mutation (PM) for maintaining diversity. A population size of 100 with 100 generations was used, and the Pareto fraction was set to 0.3 to ensure diverse Pareto-optimal solutions. Separate XGBoost models trained for UCS, CO_2 emissions, and cost predictions were incorporated, allowing precise evaluation of each candidate mix design within the bounds defined.

The optimization results provided a Pareto-optimal set, representing trade-offs between CO_2 emissions, cost, and UCS. Solutions were analyzed for predicted CO_2 emissions, cost, and UCS values, with deviations from the target UCS penalized using an absolute penalty approach. This method integrated advanced predictive modeling with robust multi-objective optimization under practical **constraints**, offering a sustainable framework for geopolymer concrete mix design.

4 CHAPTER
RESULTS AND DISCUSSIONS

This section evaluates the performance of machine learning models applied to the regression problem of predicting the compressive strength, cost, and CO_2 emissions of MK-based geopolymer concrete, integrating these predictions into the NSGA-II optimization framework. The assessment uses metrics such as R^2, RMSE, MAE, MAPE, and the α (RMSE/MAE) error ratio to analyze the models' accuracy, consistency, and error distribution. The results demonstrate that the Optimized XGBoost model delivers exceptional performance for cost and CO_2 predictions, with R^2 values of 1.0000 (train) and 0.9934 (test) for cost, and 0.9994 (train) and 0.9954 (test) for CO_2. The corresponding MAE values are 0.2232 for cost and 2.1485 for CO_2, showcasing its precision. Similarly, for compressive strength prediction, XGBoost achieves R^2 values of 0.9740 (train) and 0.8775 (test), alongside low RMSE, MAE, and MAPE scores. Other models, such as Gradient Boosting Machine (GBM) and Random Forest (RF), exhibit strong predictive capabilities, albeit with slightly higher testing errors.

The NSGA-II optimization framework utilizes these predictive models to identify optimal mix designs by minimizing cost and CO_2 emissions while maintaining compressive strength within target ranges. Feature importance analyses using Gini importance and SHAP (SHapley Additive exPlanations) indicate that key features, such as the CA-to-FA ratio, H_2O-to-Na_2O molar ratio, sodium hydroxide concentration, and added water content, play significant roles in achieving desired outcomes. This integration of machine learning and multi-objective optimization highlights a robust, interpretable approach to sustainable geopolymer concrete design.

Table 4.1 SCORES OF DIFFERENT MODELS

Metric	GBM	RF	DT	SVM	Optimized XGBoost	Optimized Compact GBM	Optimized Compact RF	Optimized BPNN
R^2 (Train)	0.9758	0.9706	0.9594	0.8885	0.9740	0.9431	0.9131	0.6865
R^2 (Test)	0.8711	0.8316	0.7686	0.8044	0.8775	0.8414	0.7853	0.6110
RMSE (Train)	2.6169	2.8831	3.3880	5.6170	2.7128	4.0136	4.9578	9.4184
RMSE (Test)	6.0717	6.9381	8.1341	7.4785	5.9192	6.7340	7.8355	10.5466
MAE (Train)	1.5617	1.6830	2.0015	3.4525	1.6129	2.7998	3.5300	7.1478
MAE (Test)	3.9530	4.6326	5.4637	5.1231	3.7816	4.6947	5.5693	8.1502
MAPE (Train)	5.4596	6.2691	7.1092	14.5230	5.7491	9.8760	11.8379	29.4217
MAPE (Test)	14.6354	18.3704	21.0355	18.6698	14.8391	17.2931	20.9632	30.5404
α (Train RMSE/MAE)	1.6757	1.7131	1.6928	1.6269	1.6820	1.4335	1.4045	1.3177
α (Test RMSE/MAE)	1.5360	1.4977	1.4888	1.4598	1.5653	1.4344	1.4069	1.2940

4.1 MODEL PERFORMANCE OVERVIEW

The results indicate significant variation in the predictive performance of different regression models, as summarized in Table. The following subsections delve into the individual models and their performance metrics.

4.1.1 Gradient Boosting Machines (GBM)

The Gradient Boosting Machine (GBM) demonstrated strong performance in both training and testing phases, with an R^2 score of 0.9758 during training, indicating excellent predictive accuracy. On testing, the R^2 score slightly reduced to 0.8711, reflecting a reasonable drop in performance but still maintaining a good level of prediction accuracy. This drop in R^2 could be attributed to the model's generalization capability on unseen data. The high training R^2 score suggests that the model is well-optimized for the given data, although there is a moderate performance gap on testing data, which is common in machine learning models when exposed to new or diverse test sets.

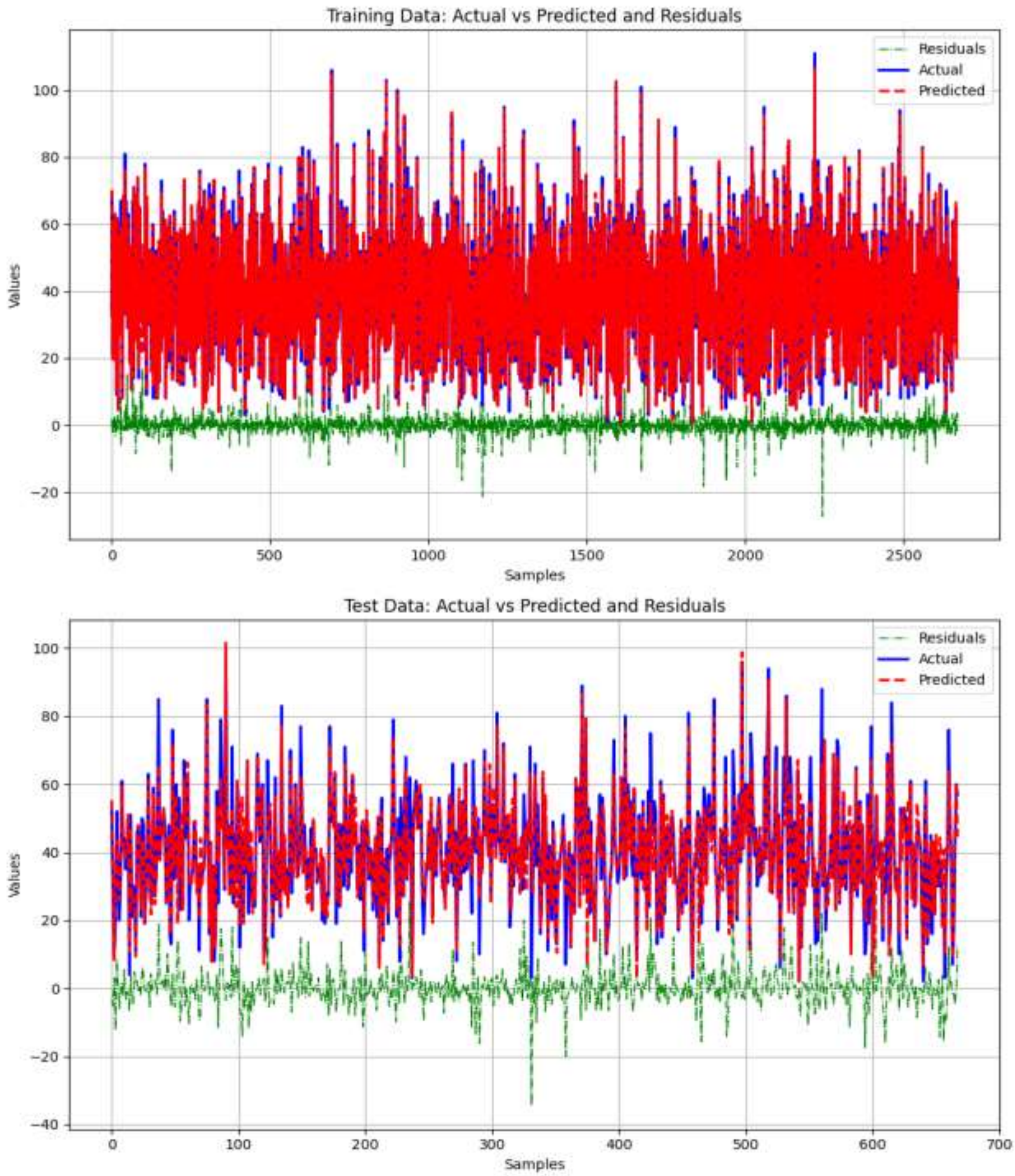


FIGURE 4.1 RESIDUAL PLOTS FOR GRADIENT BOOSTING MACHINE

4.1.2 Compact Gradient Boosting Machines (Compact GBM):

The Optimized Compact Gradient Boosting Machine (GBM) achieved an R^2 score of 0.9431 on the training data, demonstrating a strong ability to fit the model to the training set. On the testing data, the R^2 score decreased to 0.8414, suggesting a moderate decline in predictive performance when applied to new, unseen data. Despite this, the model still exhibits solid generalization capability, with a reasonable performance drop. The performance difference between training and testing R^2 scores is common in machine learning models, indicating that the optimized compact GBM model effectively balances between overfitting and underfitting.

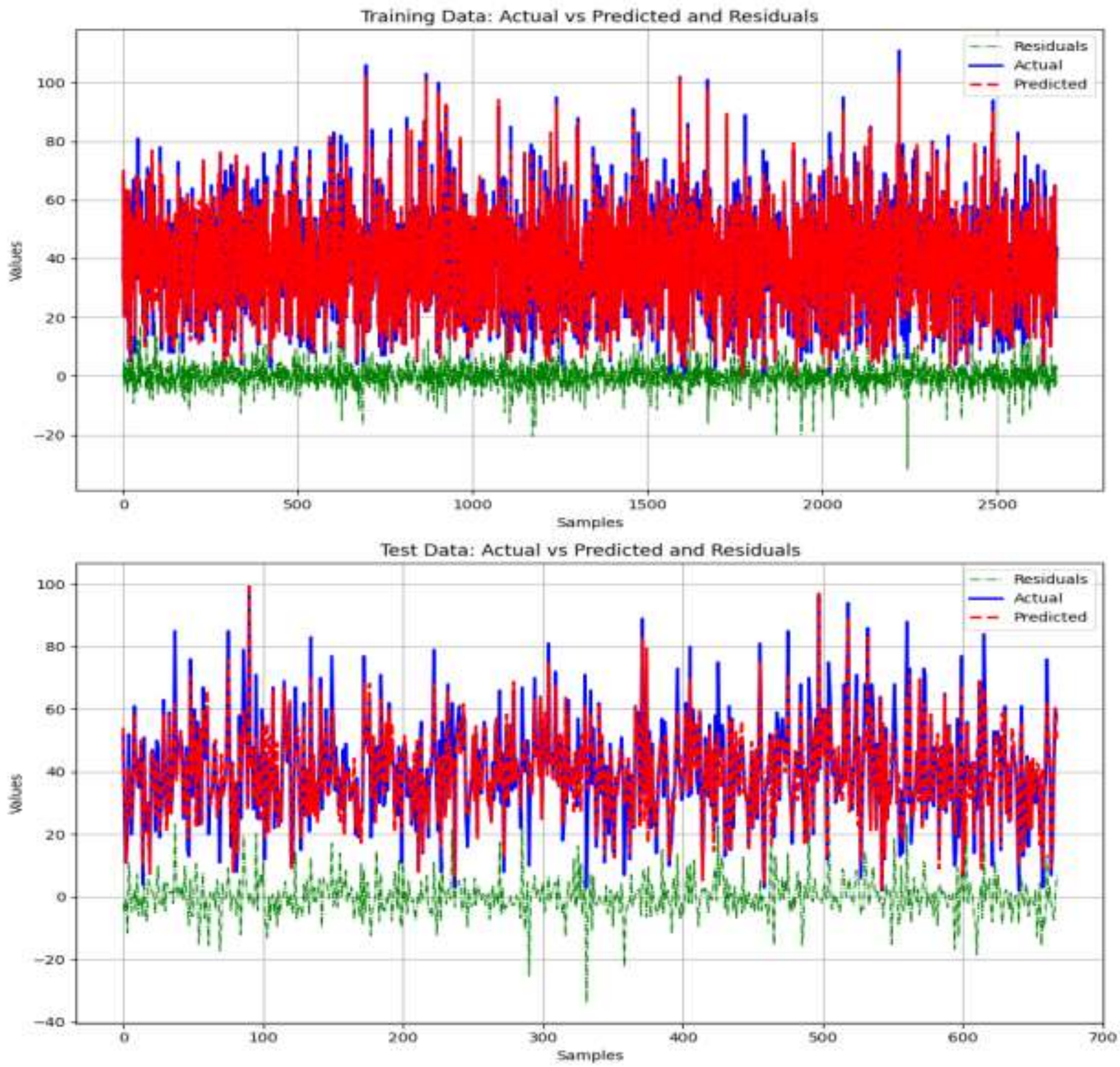


FIGURE 4.2 RESIDUAL PLOTS FOR COMPACT GBM

4.1.3 Extreme Gradient Boosting (XGBoost):

The Optimized XGBoost model achieved a high R^2 score of 0.9740 on the training data, indicating excellent fit and predictive power for the training set. However, when tested on unseen data, the R^2 score dropped to 0.8775, suggesting a slight reduction in performance. Despite this decrease, the model continues to perform well. The relatively small difference between training and testing R^2 scores indicates that the model strikes a good balance between fitting the data and avoiding overfitting, making it a robust choice for prediction tasks.

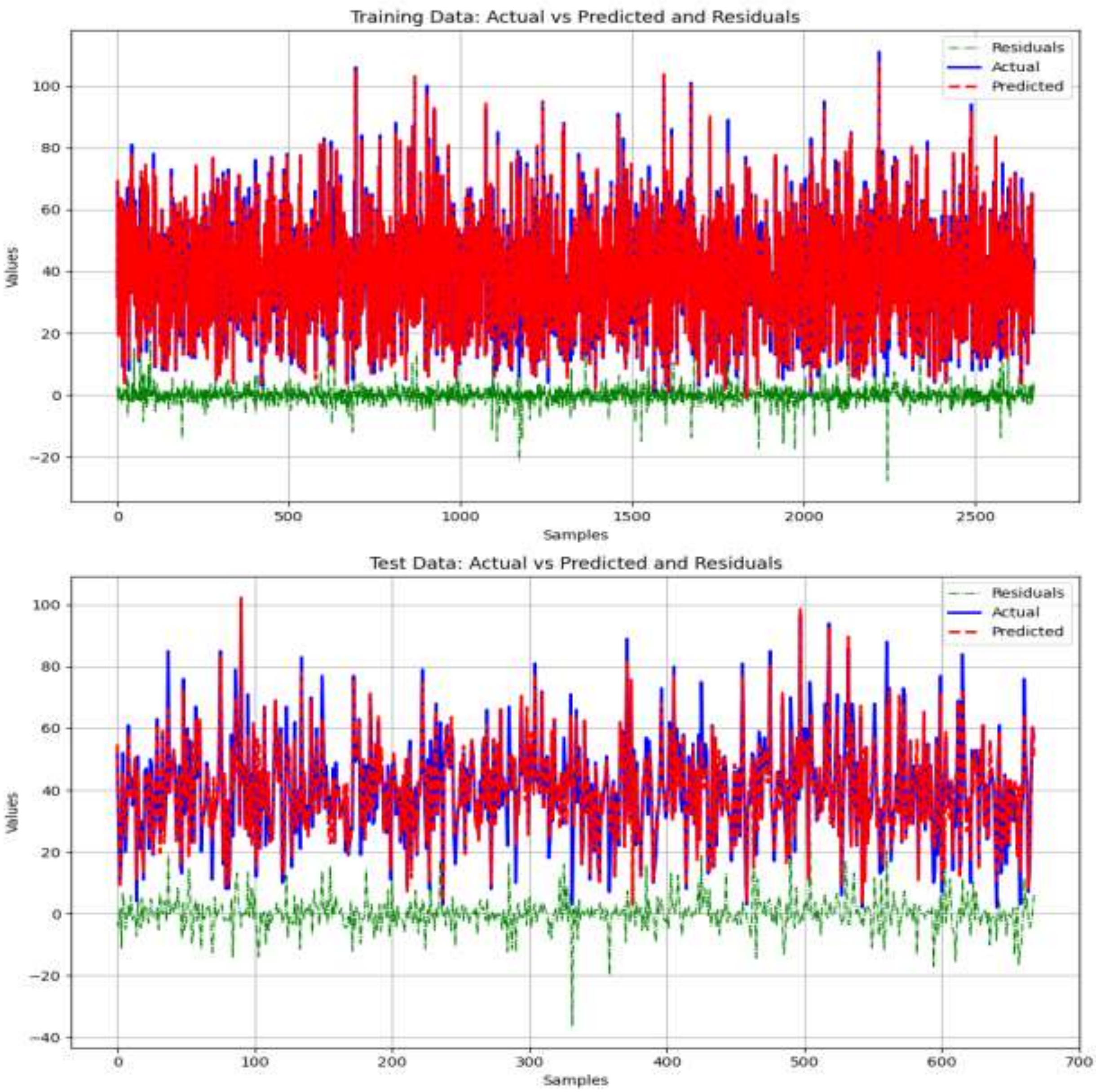


FIGURE 4.3 RESIDUAL PLOT FOR OPTIMIZED XGBOOST

4.1.4 <u>Random Forest (RF):</u>

The Random Forest model demonstrated strong performance during training with an R^2 score of 0.9706, indicating a good fit to the training data. However, when evaluated on the testing data, the R^2 score dropped to 0.8316, reflecting a decline in predictive accuracy. This suggests that model is more suited to the training data and may have some difficulty generalizing to new, unseen data.

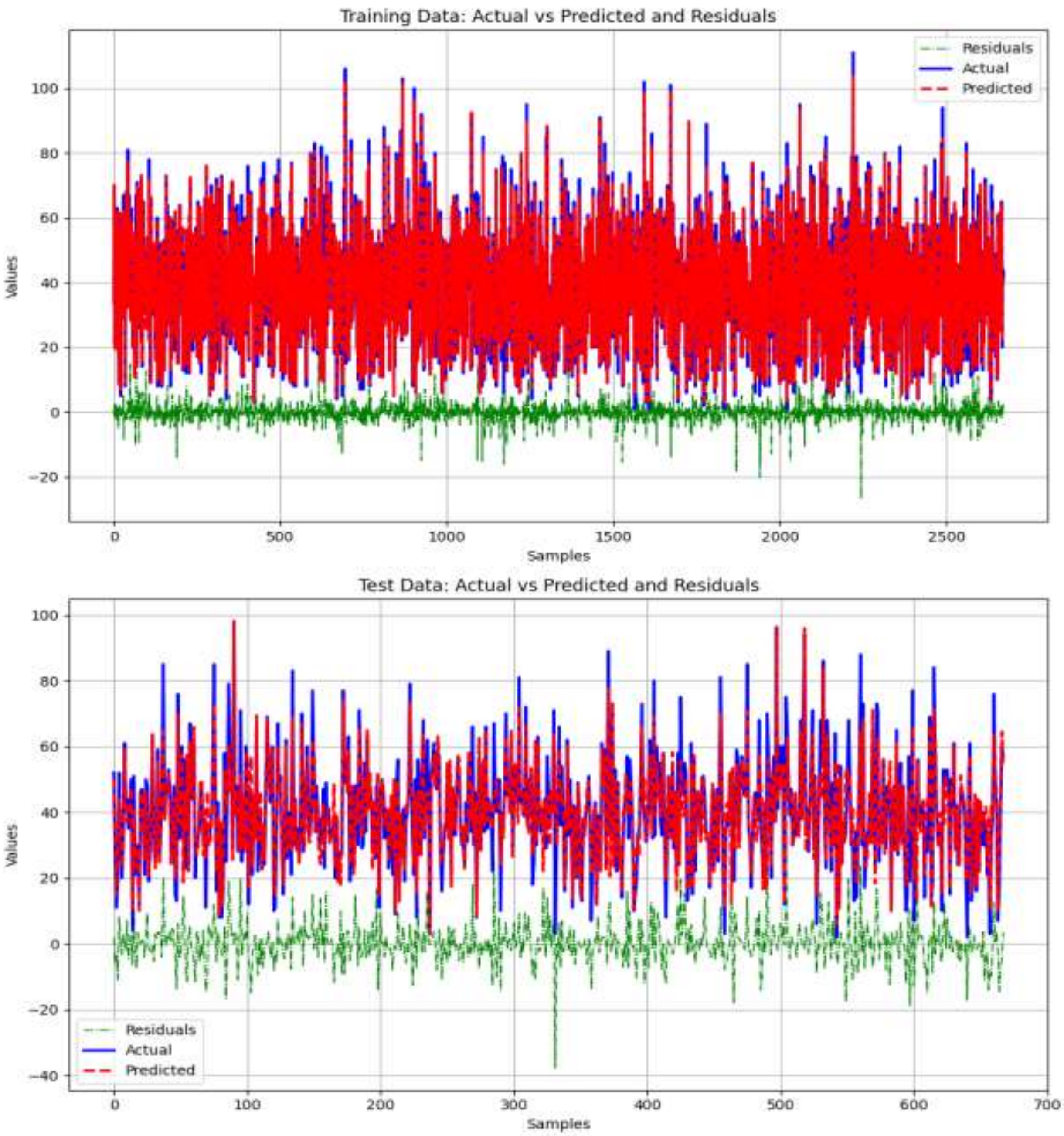


FIGURE 4.4 RESIDUAL PLOTS FOR *RANDOM FOREST*

4.1.5 <u>Compact Random Forest (Compact RF)</u>

The Compact RF model exhibited an R^2 score of 0.9131 on the training data, reflecting a strong fit to the training set. However, the performance on the testing data was slightly lower, with an R^2 score of 0.7853, indicating a decrease in the model's generalization ability. While the model still performs reasonably well, the decline in R^2 between training and testing suggests it may be overfitting to the training data.

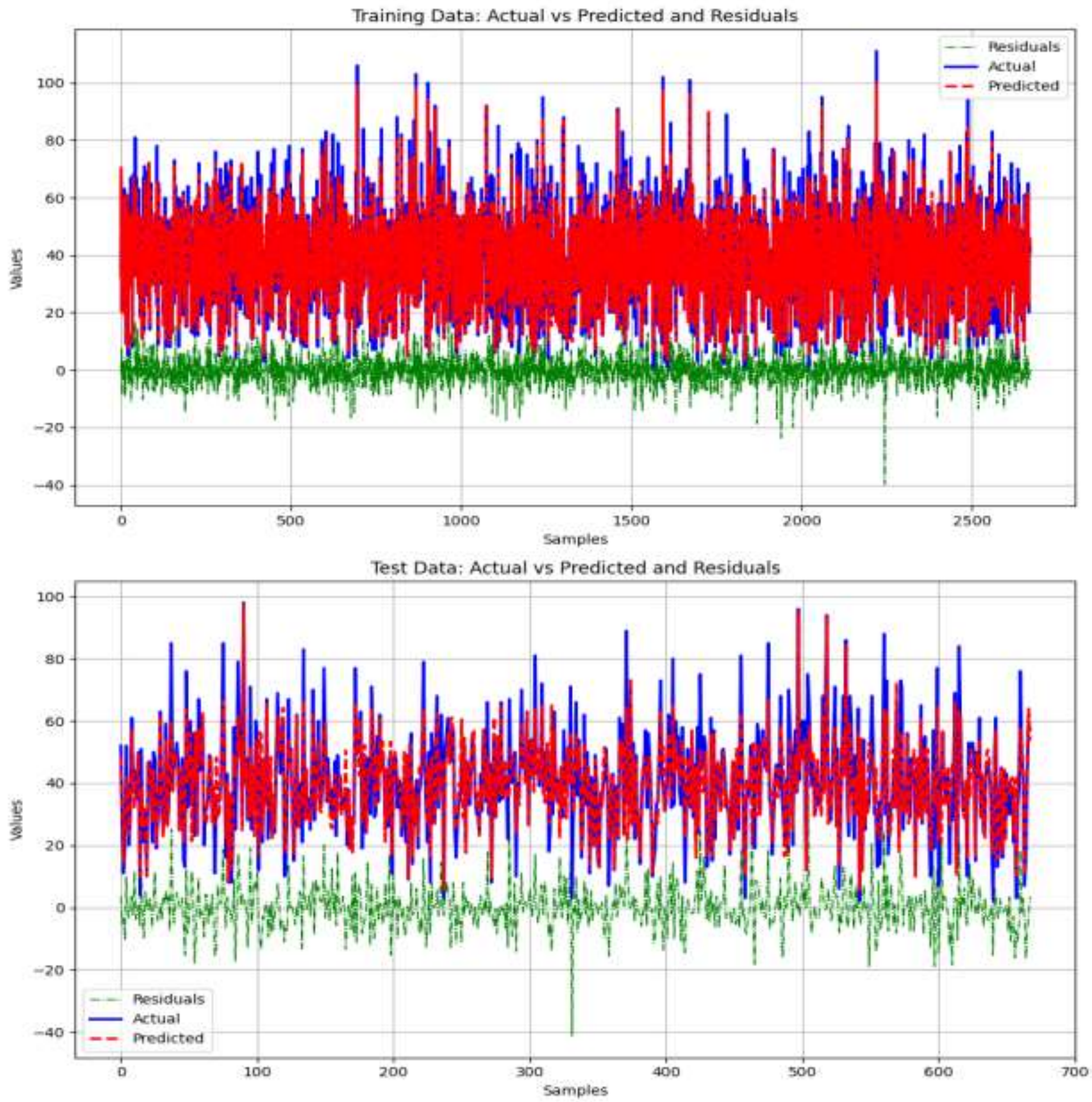


FIGURE 4.5 RESIDUAL PLOTS FOR *COMPACT RF*

4.1.6 **Decision Tree (DT):**

The Decision Tree model achieved an R^2 score of 0.9594 on the training set, indicating a very good fit to the data. However, its performance decreased on the testing set, with an R^2 score of 0.7686. This suggests that the model is overfitting the training data, as it performs well during training but struggles to generalize on unseen data.

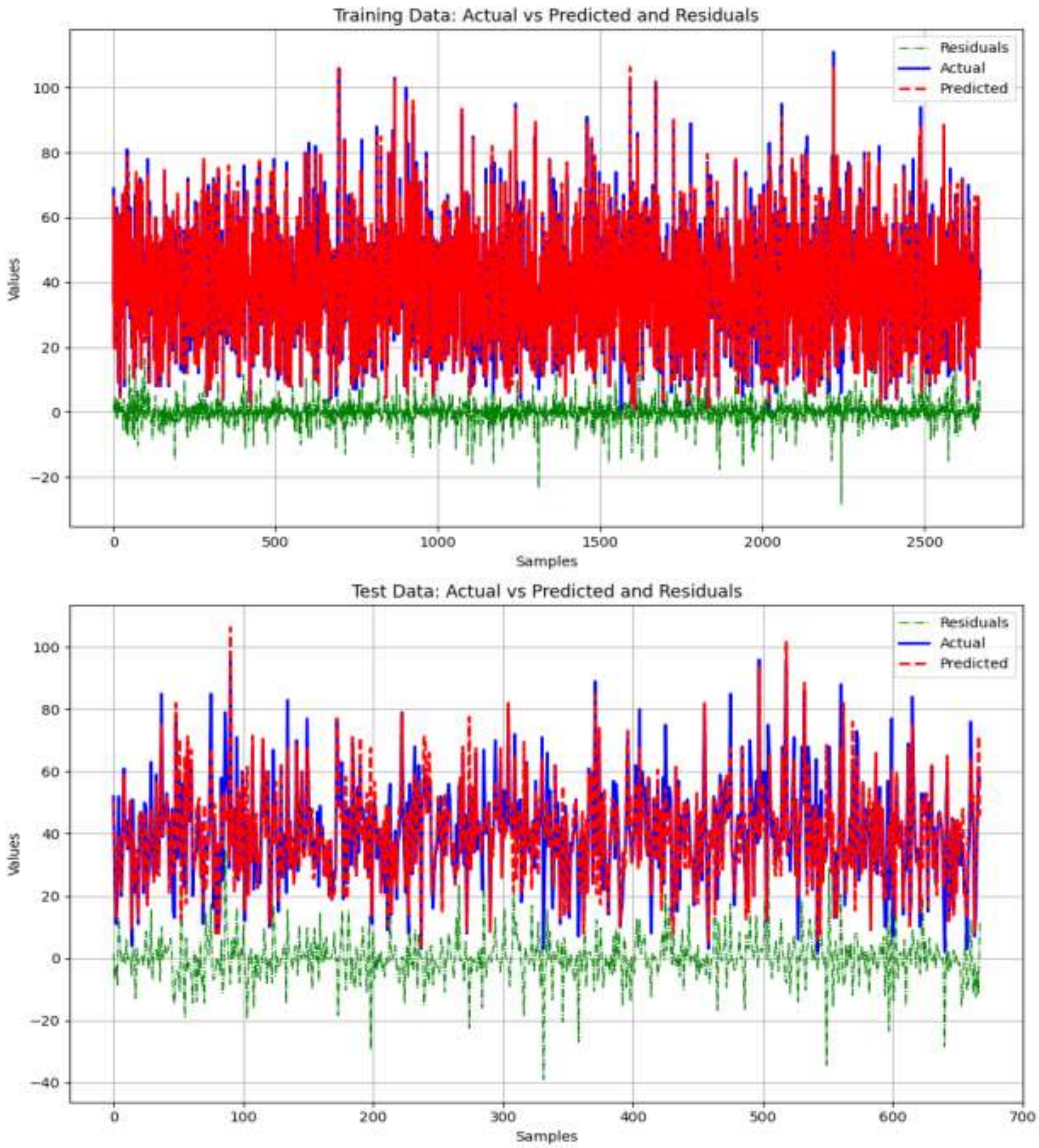


FIGURE 4.6 RESIDUAL PLOT FOR DECISION TREE

4.1.7 <u>Backpropagation Neural Network (BPNN):</u>

The Optimized BPNN model showed relatively lower performance compared to other models, with an R^2 score of 0.6865 on the training data and 0.6110 on the testing data. This indicates that the model struggles to capture the underlying relationships between the features and the target variables.

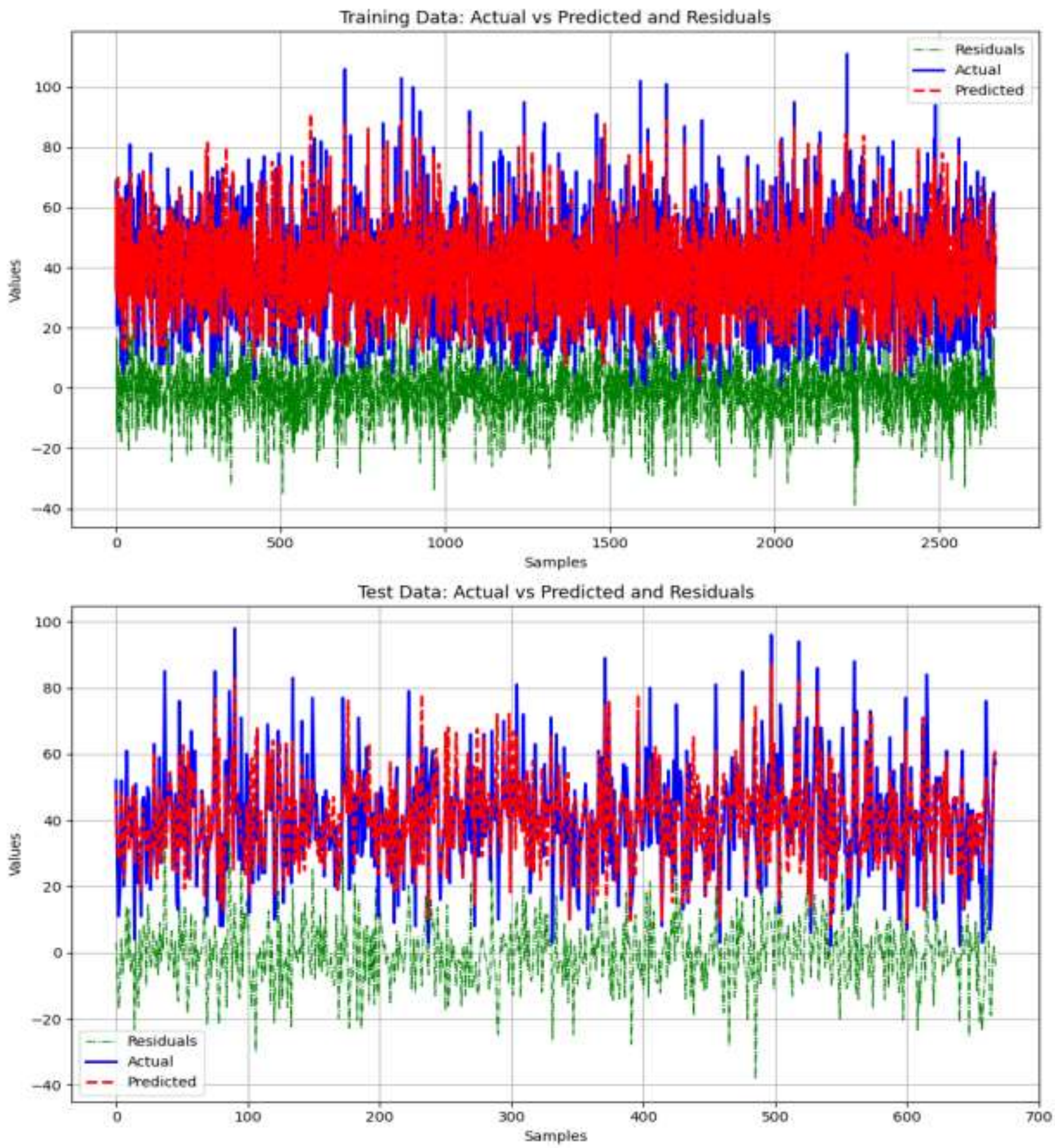


FIGURE 4.7 RESIDUAL PLOTS FOR BPNN

4.1.8 Support Vector Machine (SVM):

The Support Vector Machine model demonstrated an R^2 score of 0.8885 on the training set, reflecting good accuracy in fitting the data. On the testing set, the R^2 score dropped slightly to 0.8044, indicating a modest decline in performance. While SVM is still effective in predicting the outputs, the decrease in performance suggests it may not be capturing all the underlying patterns in the data as well as some other models.

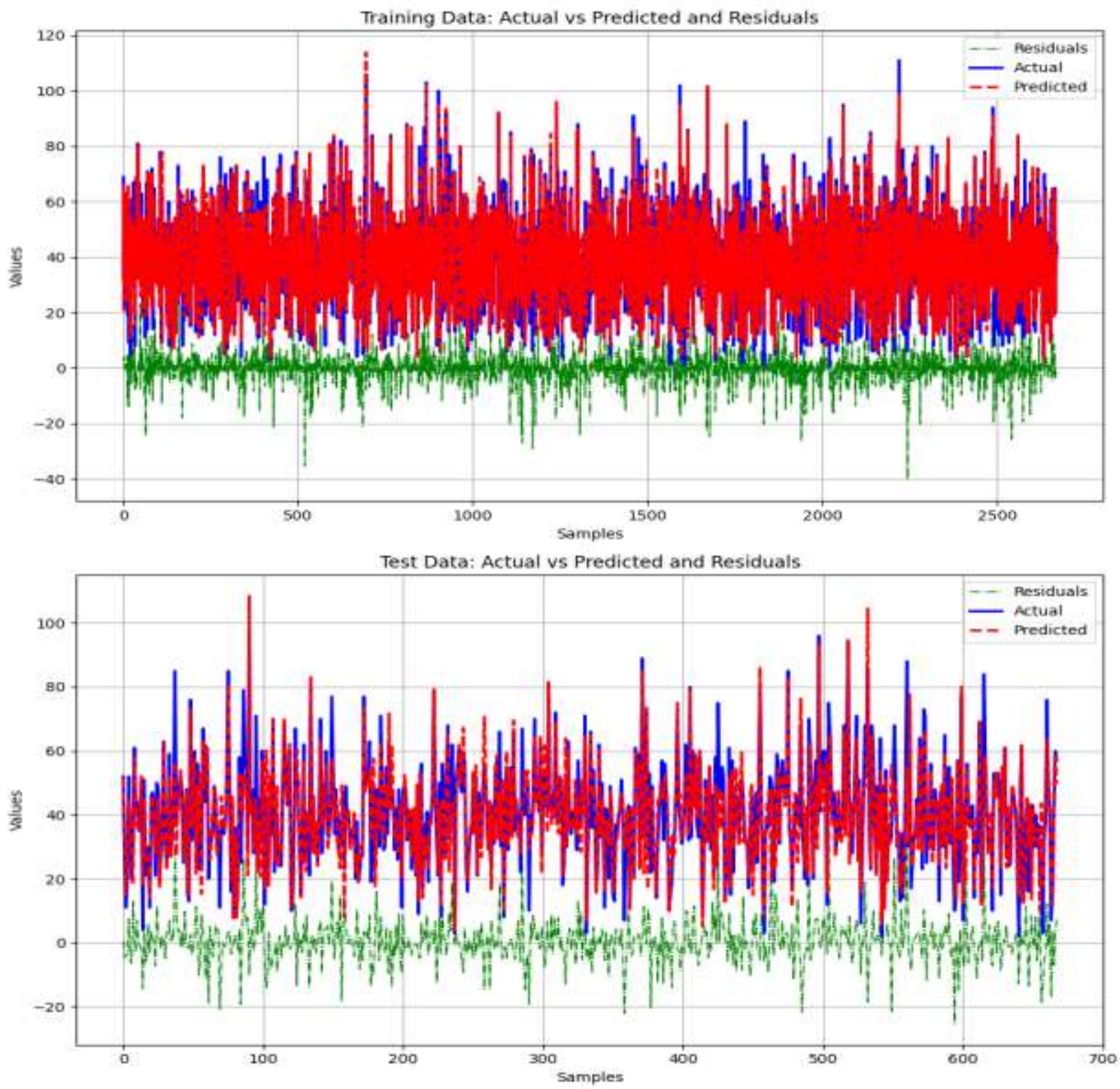
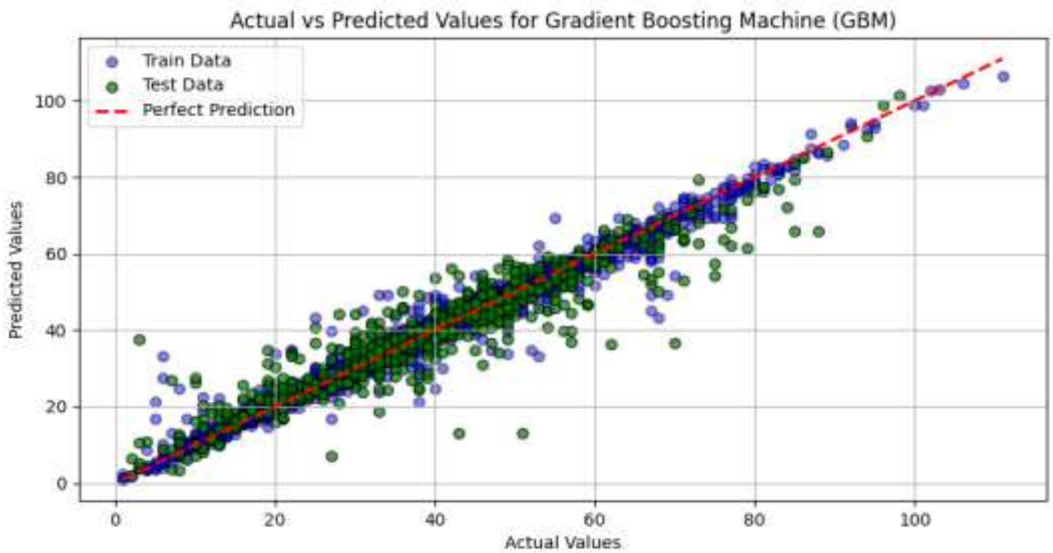
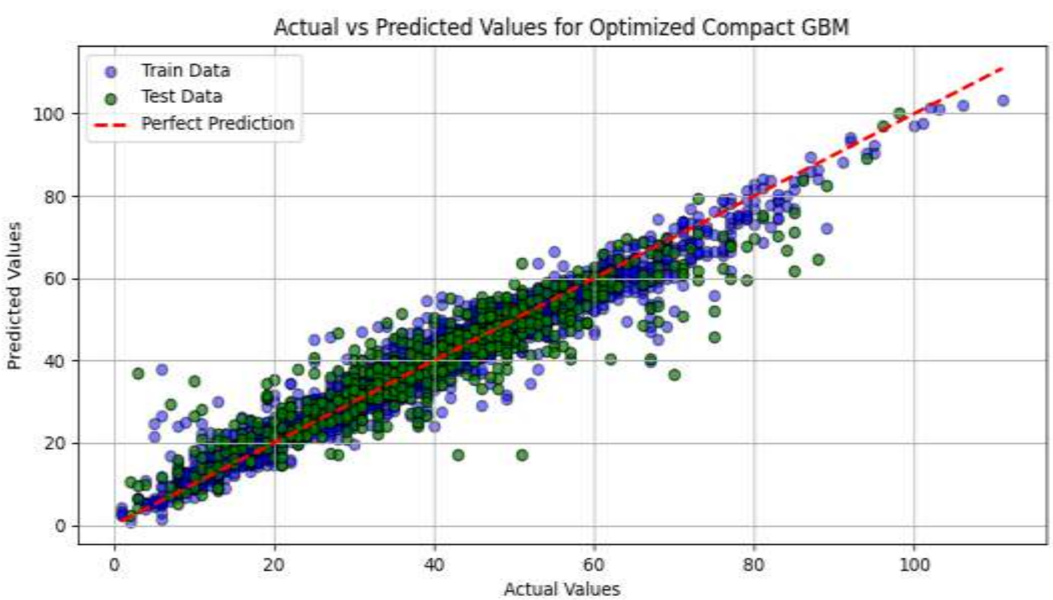


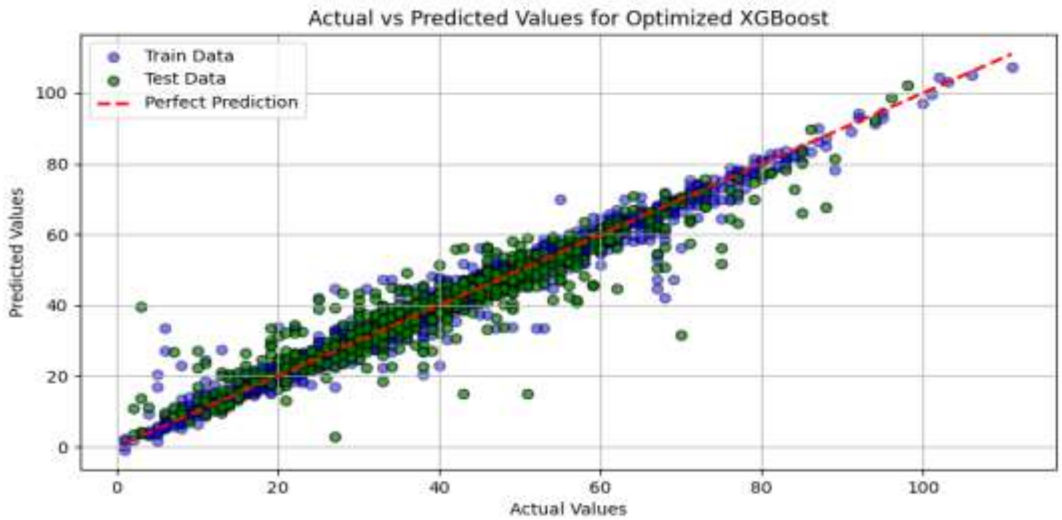
FIGURE 4.8 RESIDUAL PLOTS FOR SUPPORT VECTOR MACHINE

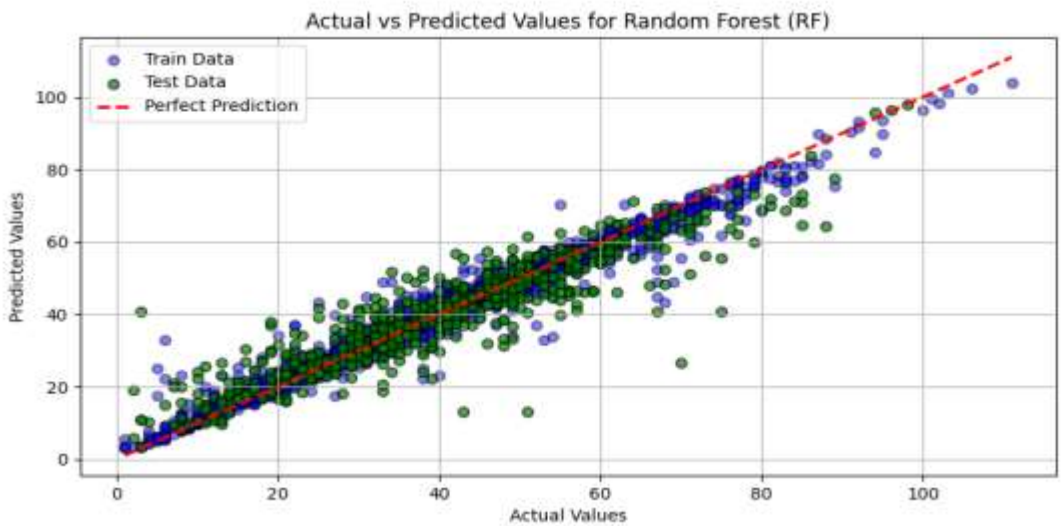
4.2 ACTUAL VS. PREDICTED PLOTS FOR ML MODELS

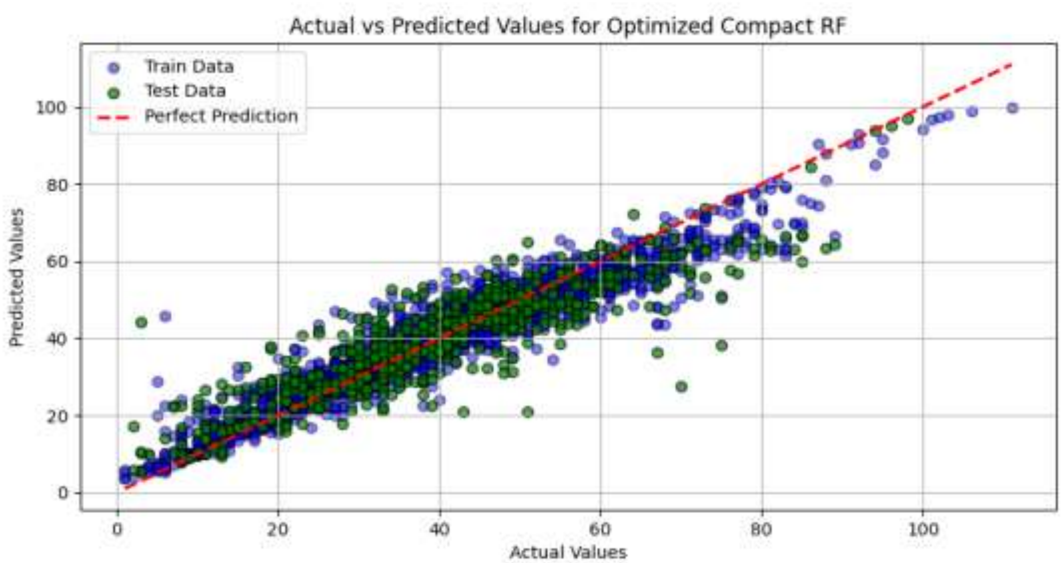
Actual versus predicted plots visually compare true UCS values to model predictions, with points ideally clustering along the diagonal for perfect accuracy. High-performing models like GBM and Optimized XGBoost show tightly clustered points, reflecting minimal error and strong predictive accuracy. In contrast, models like Optimized BPNN exhibit greater scatter and deviations, indicating lower R^2 scores and less reliable predictions. These plots are essential for assessing model accuracy and identifying biases or errors across the UCS range.

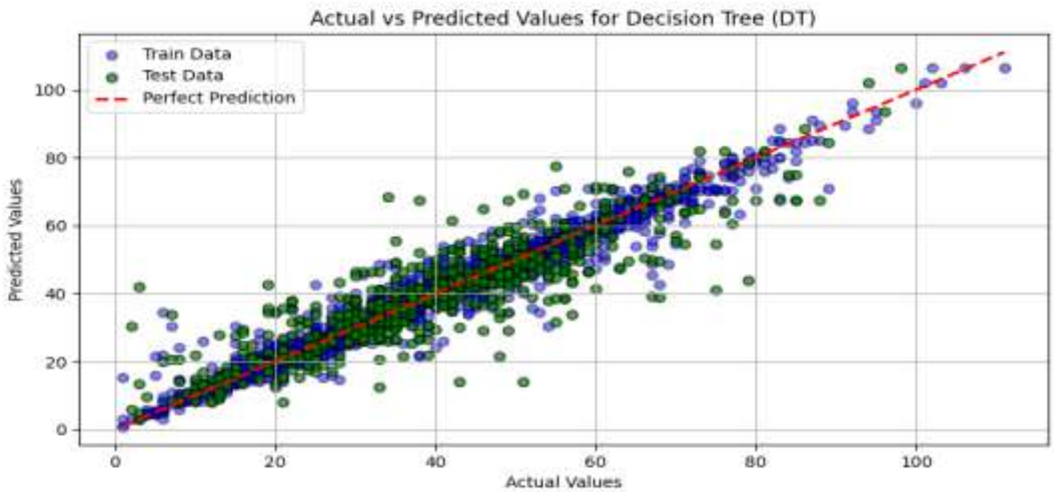


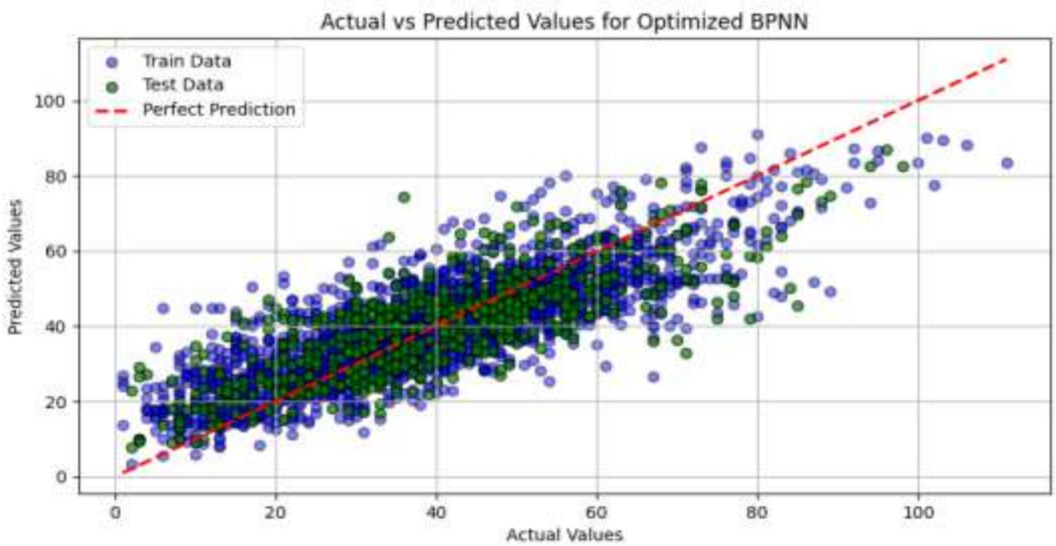


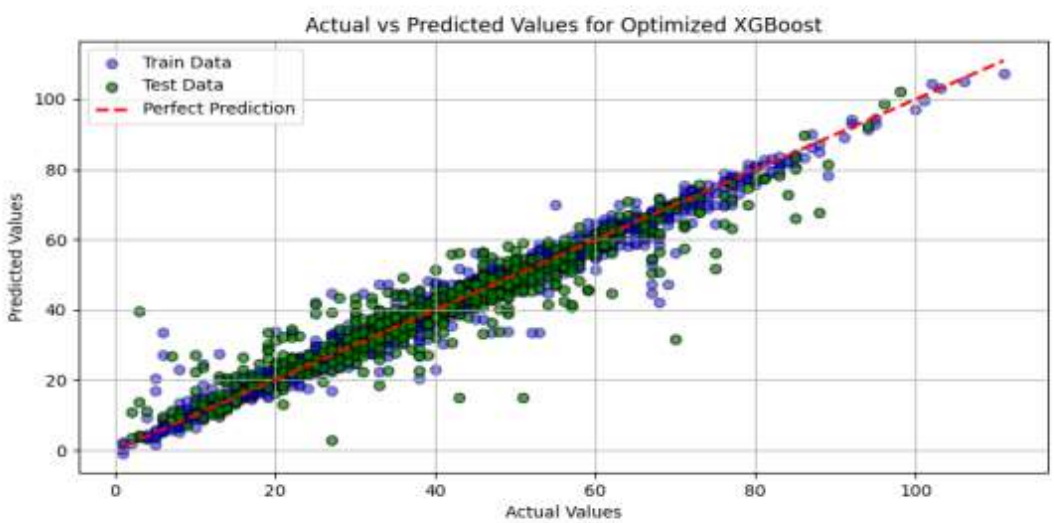

Actual vs Predicted Values for Optimized XGBoost
Train Data
Test Data
Perfect Prediction
Predicted Values
Actual Values
0
20
40
60
80
100


Actual vs Predicted Values for Random Forest (RF)
Train Data
Test Data
Perfect Prediction
Predicted Values
Actual Values
0
20
40
60
80
100


Actual vs Predicted Values for Optimized Compact RF
Train Data
Test Data
Perfect Prediction
Predicted Values
Actual Values
0
20
40
60
80
100

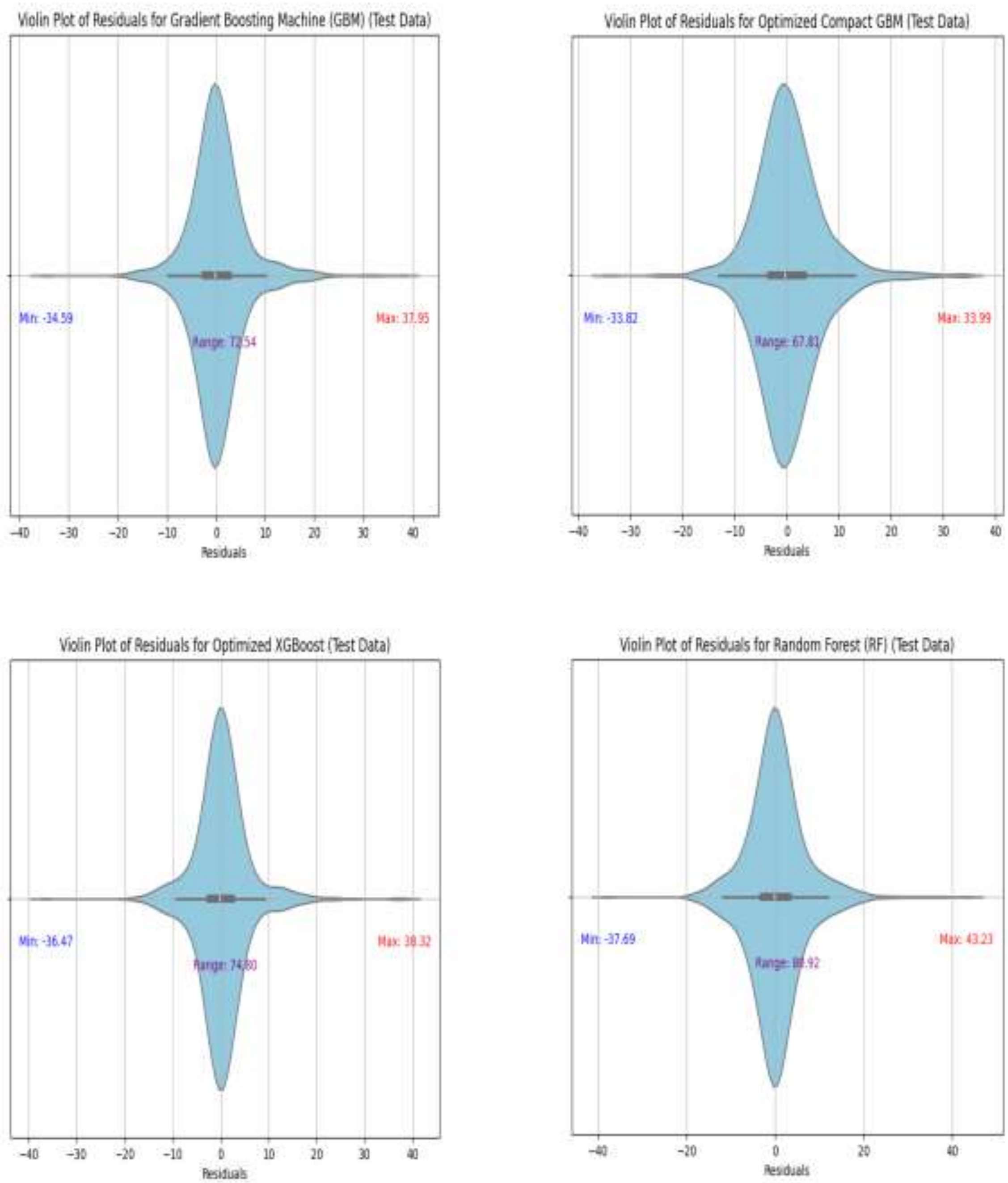

Actual vs Predicted Values for Decision Tree (DT)
Train Data
Test Data
Perfect Prediction
Predicted Values
Actual Values


Actual vs Predicted Values for Optimized BPNN
Train Data
Test Data
Perfect Prediction
Predicted Values
Actual Values


Actual vs Predicted Values for Optimized XGBoost
Train Data
Test Data
Perfect Prediction
Predicted Values
Actual Values

4.3 VIOLIN PLOT ANALYSIS OF ML MODEL PREDICTIONS

Violin plots effectively illustrate the distribution and variability of actual and predicted UCS values across machine learning models. Models like XGBoost and Gradient Boosting Machine (GBM) closely align predicted distributions with actual values, showcasing their superior accuracy. Narrow violins indicate limited flexibility, while wider ones reflect greater variability, which may suggest robust modeling or overfitting. In contrast, models with lower R^2 scores, such as the



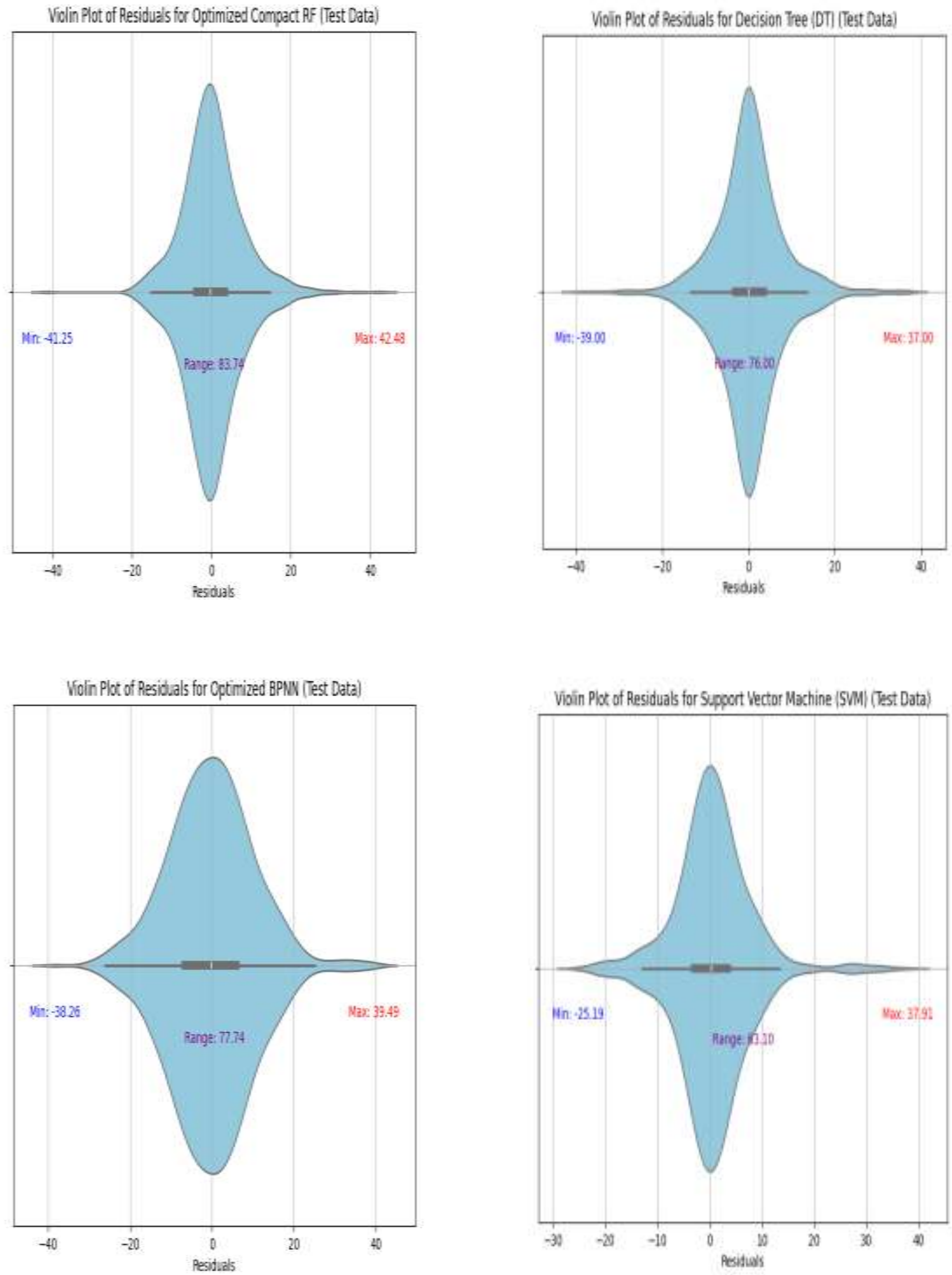


FIGURE 4.9 VIOLIN PLOTS OF ML MODELS

4.4 MODEL COMPARISON AND METRICS ANALYSIS

To gain a comprehensive view of model performance, various error metrics across all models were compared, as summarized in Table 3. Notably, the Gradient Boosting Machines (GBM) and Optimized XGBoost models consistently outperformed others in terms of R^2, achieving scores of 0.9758 and 0.9740 during training and 0.8711 and 0.8775 during testing, respectively. Their relatively low MAPE values (below 15%) indicate fewer significant prediction errors in percentage terms. The α ratios for these models, around 1.6, suggest a good balance between residual error and prediction magnitude. Conversely, models such as Optimized BPNN and Decision Tree showed lower R^2 scores (0.6865 and 0.9594 during training, 0.6110 and 0.7686 during testing), with higher MAPE values exceeding 21%, indicating larger fluctuations and less accurate predictions. These discrepancies are reflected in their α ratios, which were generally above 1.7, highlighting room for improvement in their predictive performance.

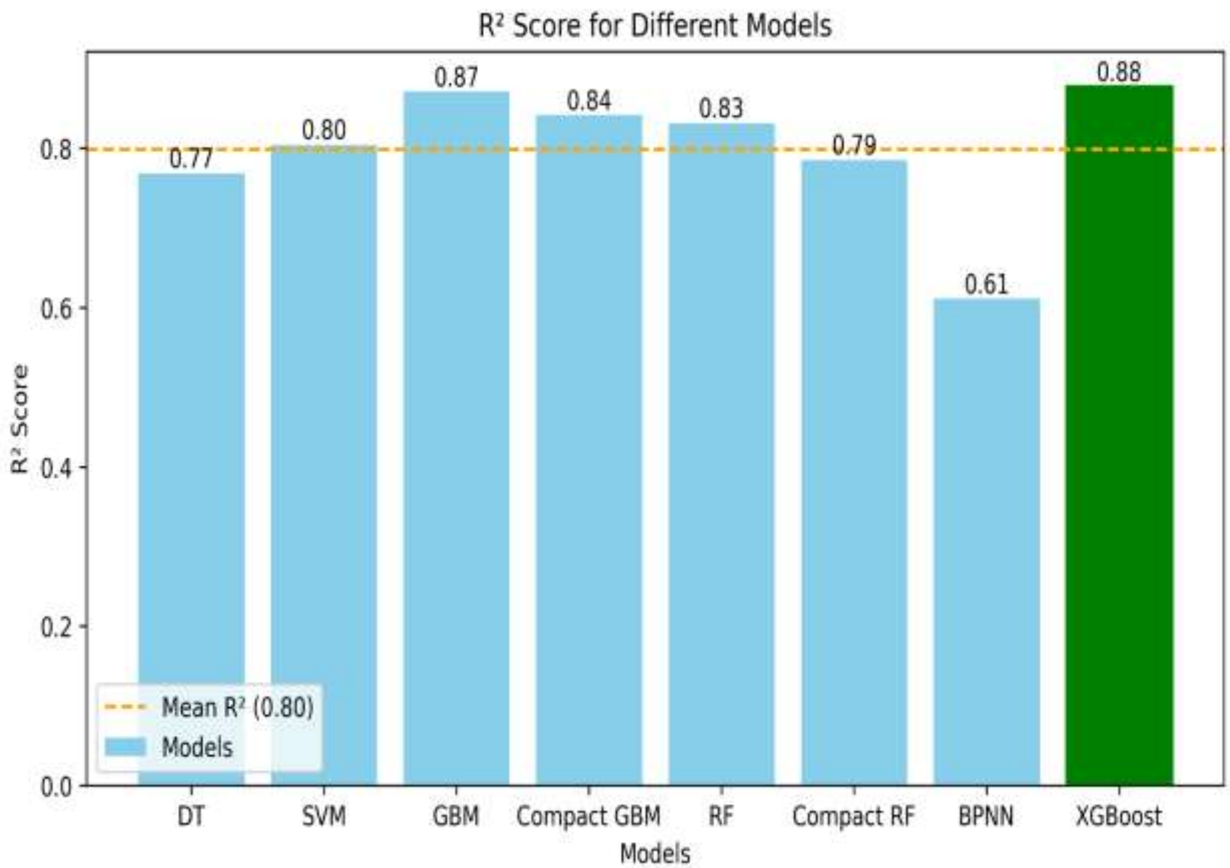


FIGURE 4.10 R^2 SCORES OF DIFFERENT MODELS

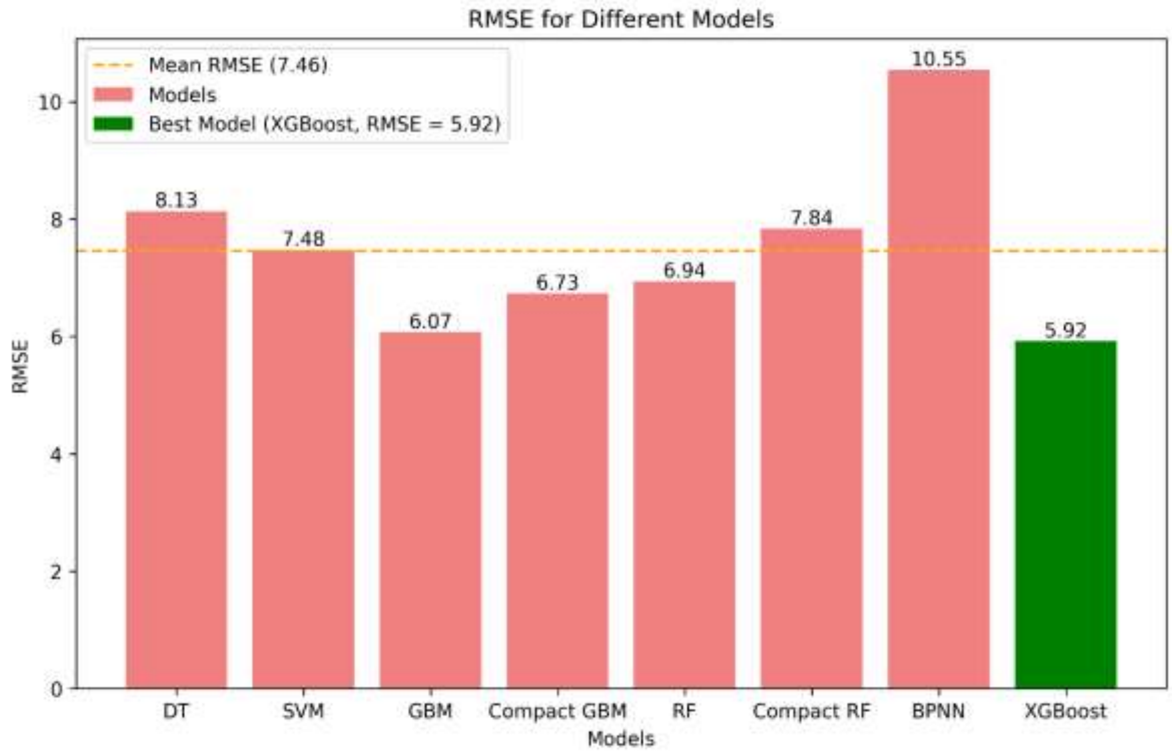


FIGURE 4.12 RMSE FOR DIFFERENT MODELS

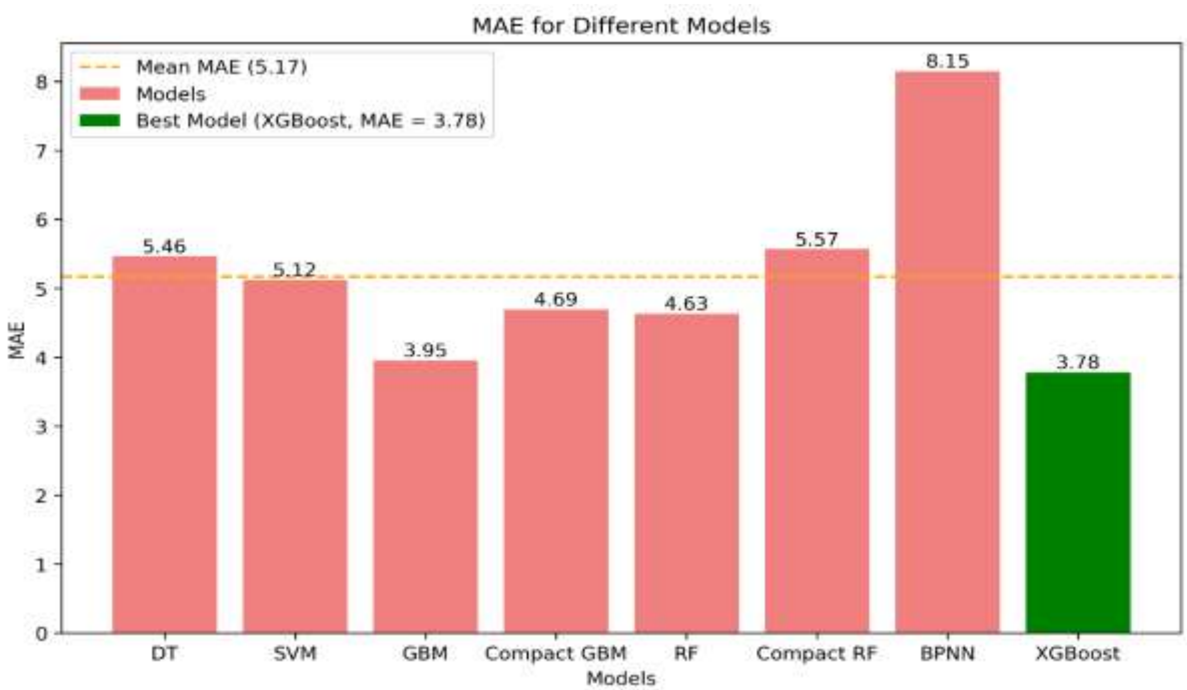


FIGURE 4.11 MAE FOR DIFFERENT MODELS

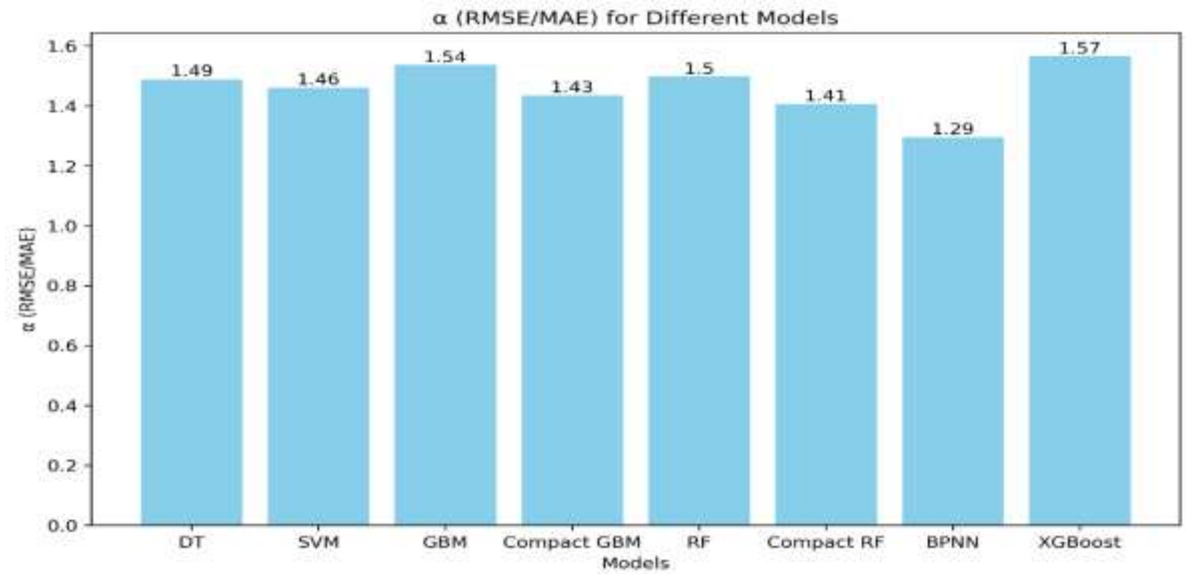


FIGURE 4.13 ALPHA FOR DIFFERENT MODELS

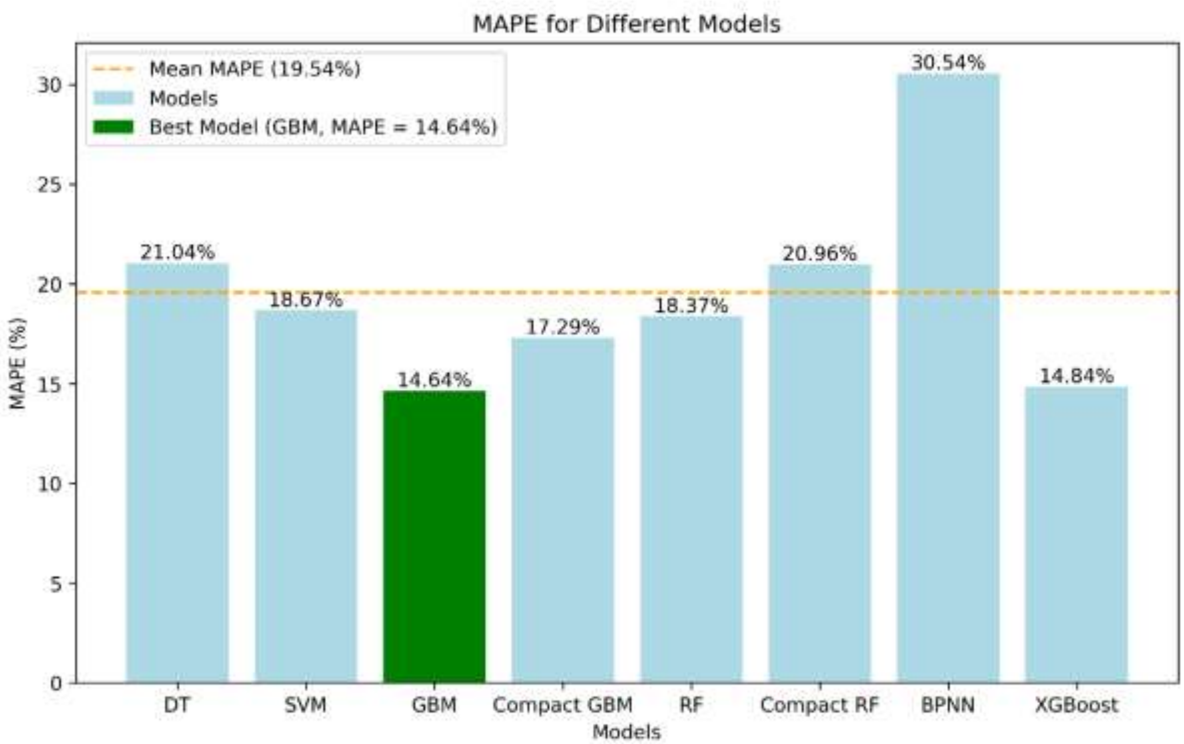


FIGURE 4.14 MAPE FOR DIFFERENT MODELS

4.5 FEATURE IMPORTANCE ANALYSIS

The feature importance analysis for the Optimized XGBoost model highlights the critical factors influencing the prediction of UCS values. Out-of-bag (OOB) analysis resulted in a score of 0.8616, indicating robust model performance. Among the features, GGBFS emerged as the most influential variable, with a permuted importance of 0.8909, significantly surpassing others. Age (Days) and Cti (°C) followed as the second and third most important features, with respective importances of 0.4597 and 0.4526, underscoring their pivotal roles in determining compressive strength. Secondary contributors included **SS**, **Mol_SH**, and **FA**, with relatively lower permuted importance values but still contributing meaningfully to model predictions.

The mean SHAP (SHapley Additive exPlanations) values provided further insights into feature impact, aligning with the importance rankings from OOB analysis. SHAP plots revealed that features like GGBFS and Age (Days) had the largest individual contributions to UCS predictions, with higher values generally associated with increased strength. Conversely, features such as Fly Ash, RH (%), and SH_H2O % showed minimal influence, as evidenced by their low permuted importance and SHAP contributions. The SHAP summary plot also demonstrated the distribution of feature impacts, illustrating both positive and negative contributions across the dataset. These results emphasize the importance of optimizing key mix design and curing parameters to enhance UCS predictions in geopolymer concrete formulations.

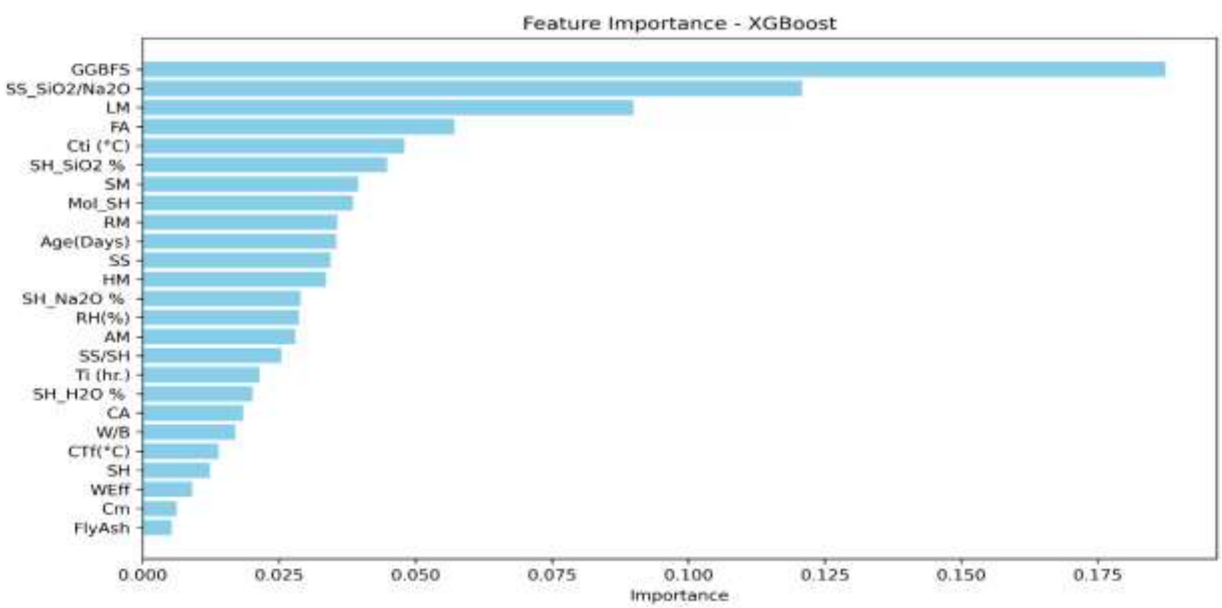


FIGURE 4.15 XGBOOST FEATURE IMPORTANCE

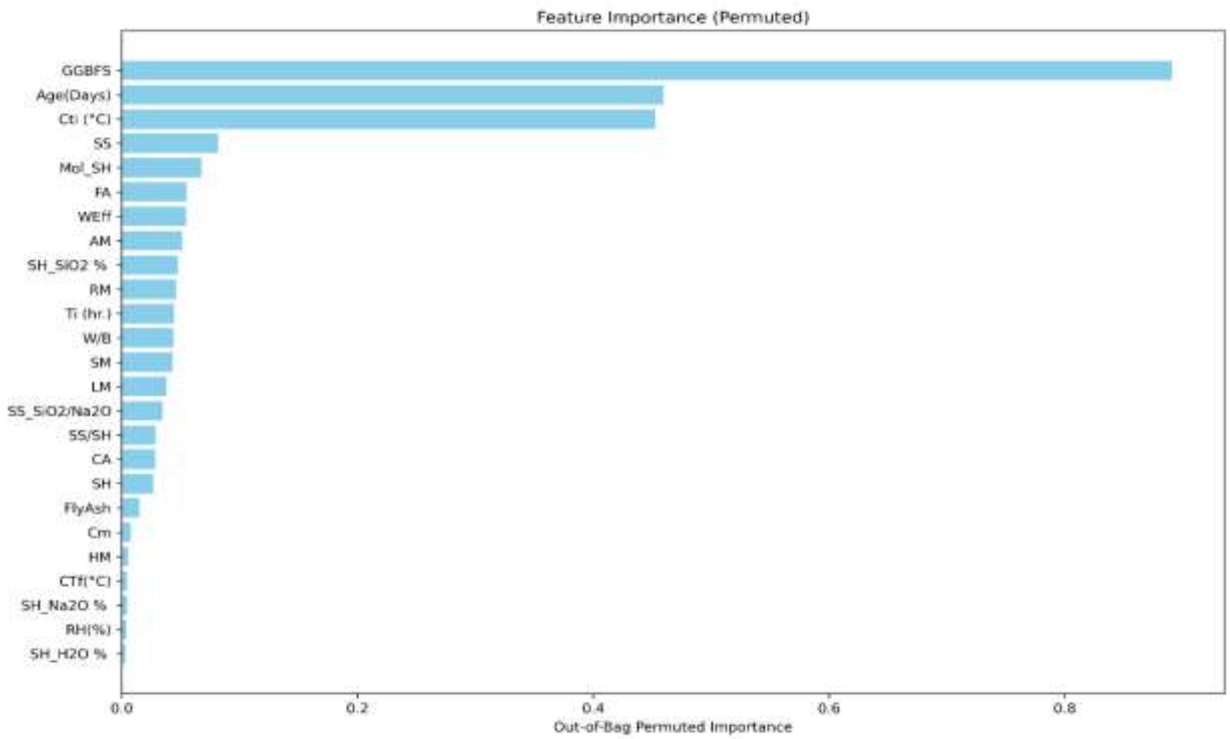


FIGURE 4.17 FEATURE IMPORTANCE (PERMUTED)

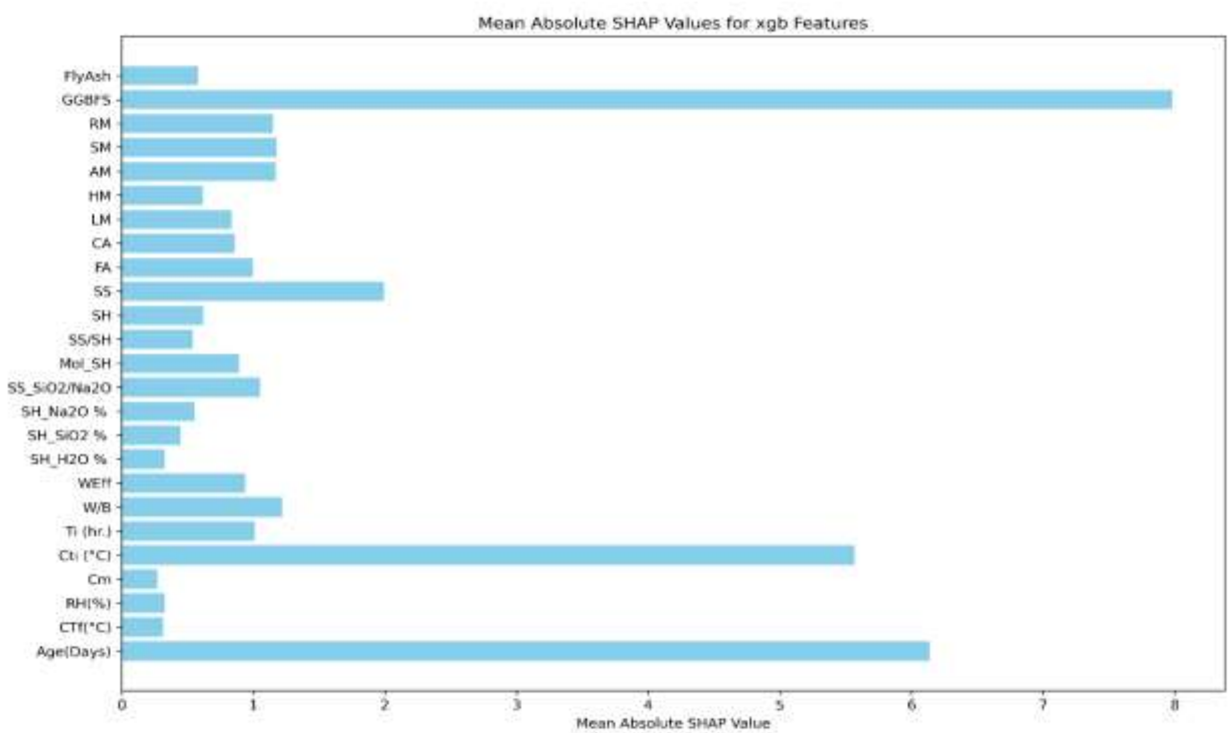


FIGURE 4.16 MEAN ABSOLUTE SHAP VALUE

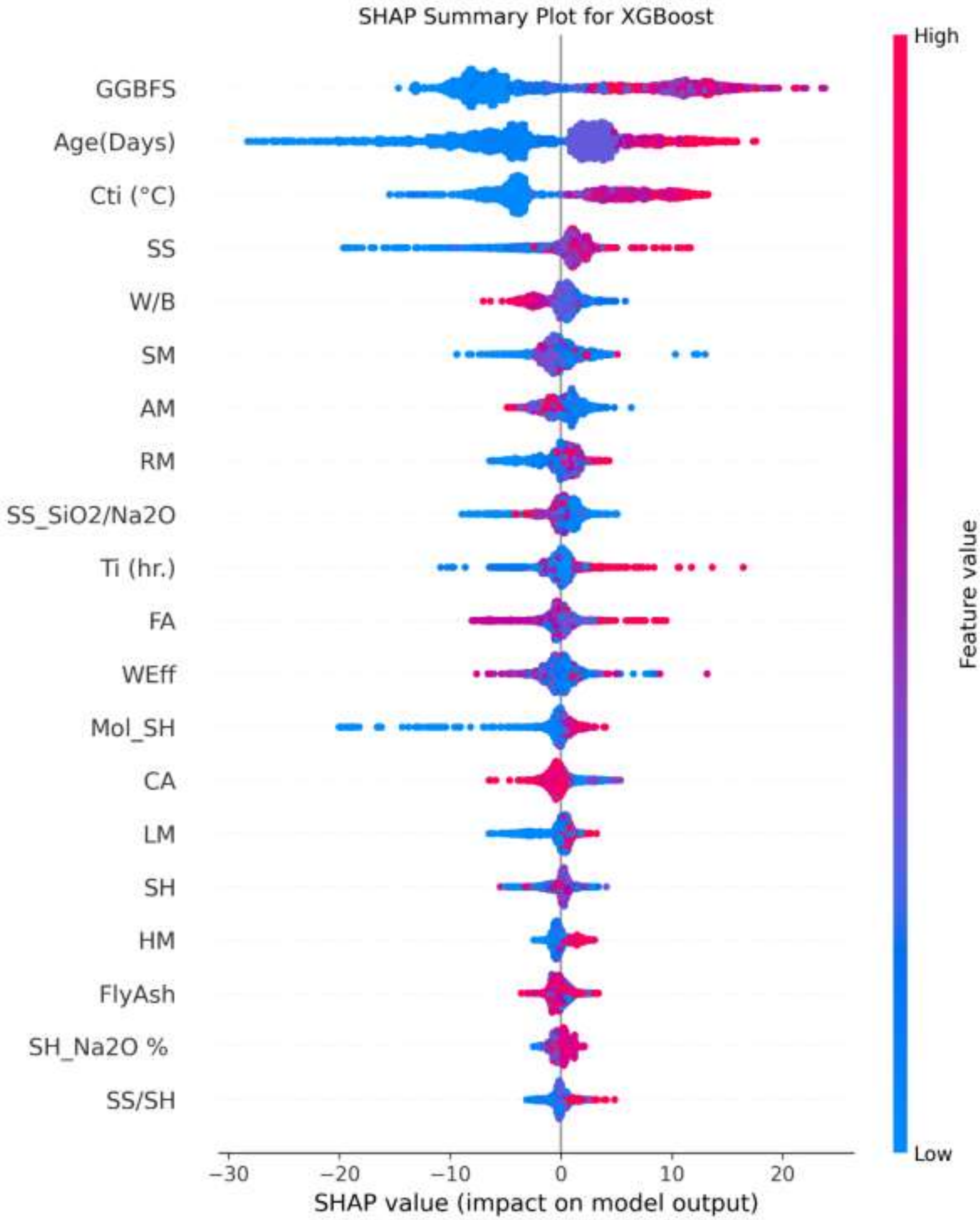


FIGURE 4.18 SHAP SUMMARY PLOT FOR XGBOOST

4.6 MULTI-OBJECTIVE OPTIMIZATION

The multi-objective optimization process successfully identified a geopolymer concrete mix design that balances compressive strength (UCS), CO_2 emissions, and cost. Using a weighted scoring approach, the optimization targeted a UCS of 50 MPa while minimizing cost and CO_2 emissions. Among the evaluated solutions, the best mix design (solution index: 85) achieved a UCS of **50.011 MPa**, CO_2 emissions of **231.88 kg/m³**, and cost of **$56.89**, resulting in a total score of **288.82**. This mix design emphasizes the critical trade-offs among objectives, with UCS deviation weighted five times more than cost and CO_2, ensuring strength remains the top priority.

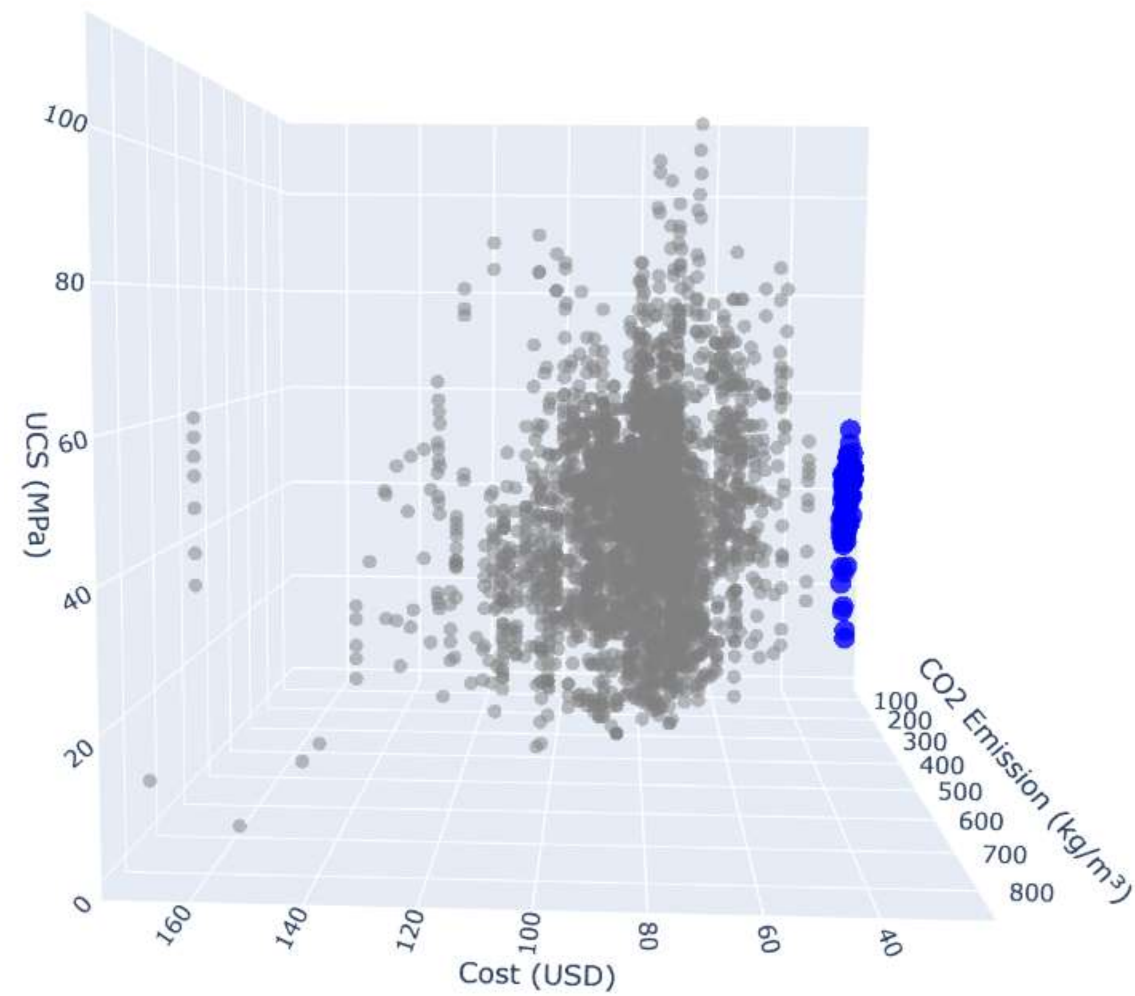


FIGURE 4.19. 3D PARETO FRONT FOR UCS, COST AND CO2 EMISSION

A 3D Pareto front visualization further highlights the trade-offs and interactions between UCS, cost, and CO_2 emissions across the set of Pareto-optimal solutions. The scatter plot illustrates the non-dominated solutions, forming a smooth surface that can be explored interactively. Each point on the Pareto front represents a viable design configuration, enabling decision-makers to choose a solution based on their specific priorities. For instance, solutions at the lower-cost end might slightly compromise UCS or CO_2, whereas high-UCS designs may incur greater cost or emissions.

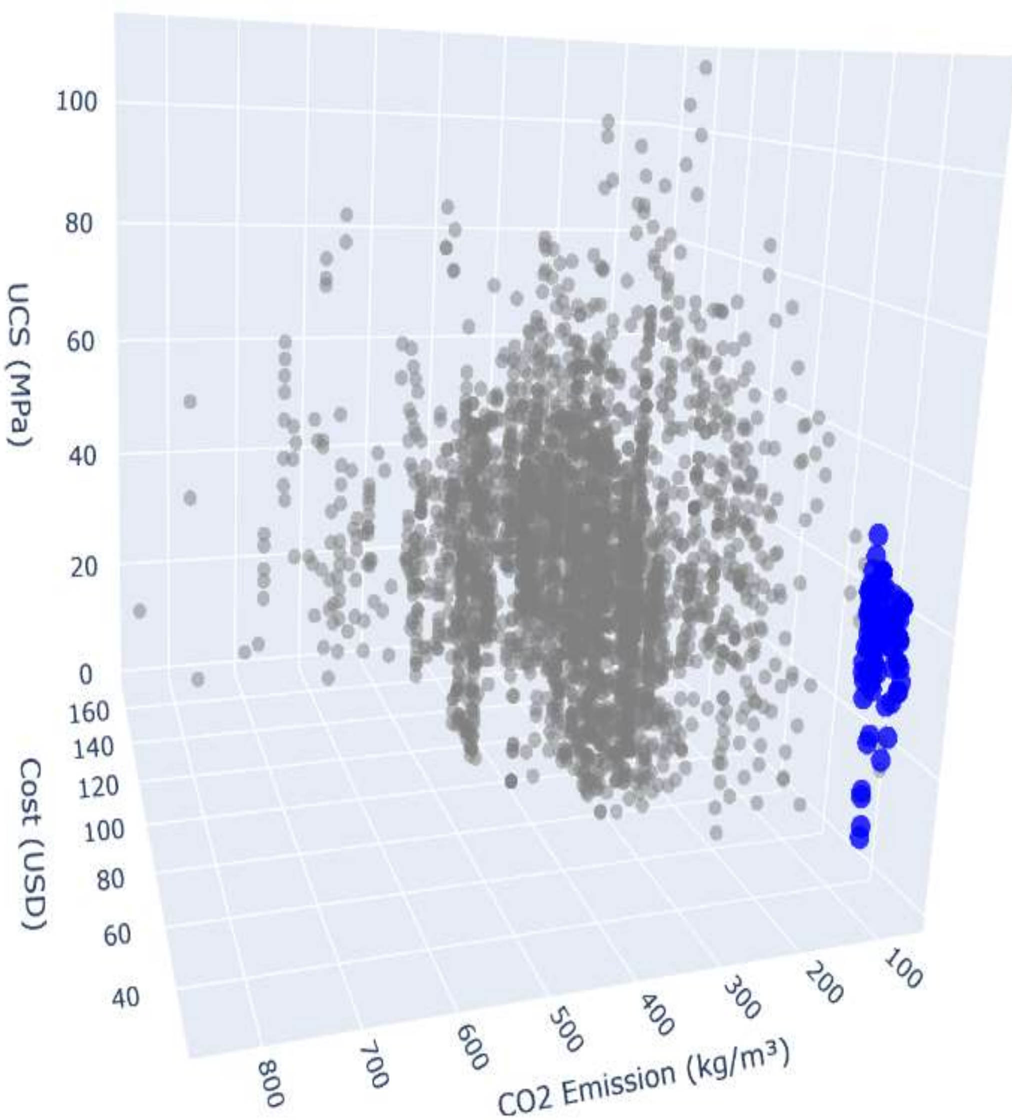


FIGURE 4.20.3D PARETO FRONT FOR UCS, COST AND CO2 EMISSION

The plot also includes a comparison with the broader dataset, showing how Pareto-optimal solutions outperform the dataset points in balancing the three objectives. Notably, UCS values were negated during the optimization process to align with minimization objectives and later reverted to positive values for accurate visualization. The use of Pareto front analysis provides insights into the efficiency of NSGA-II in generating diverse, high-performing mix designs, demonstrating its effectiveness in sustainable geopolymer concrete optimization.

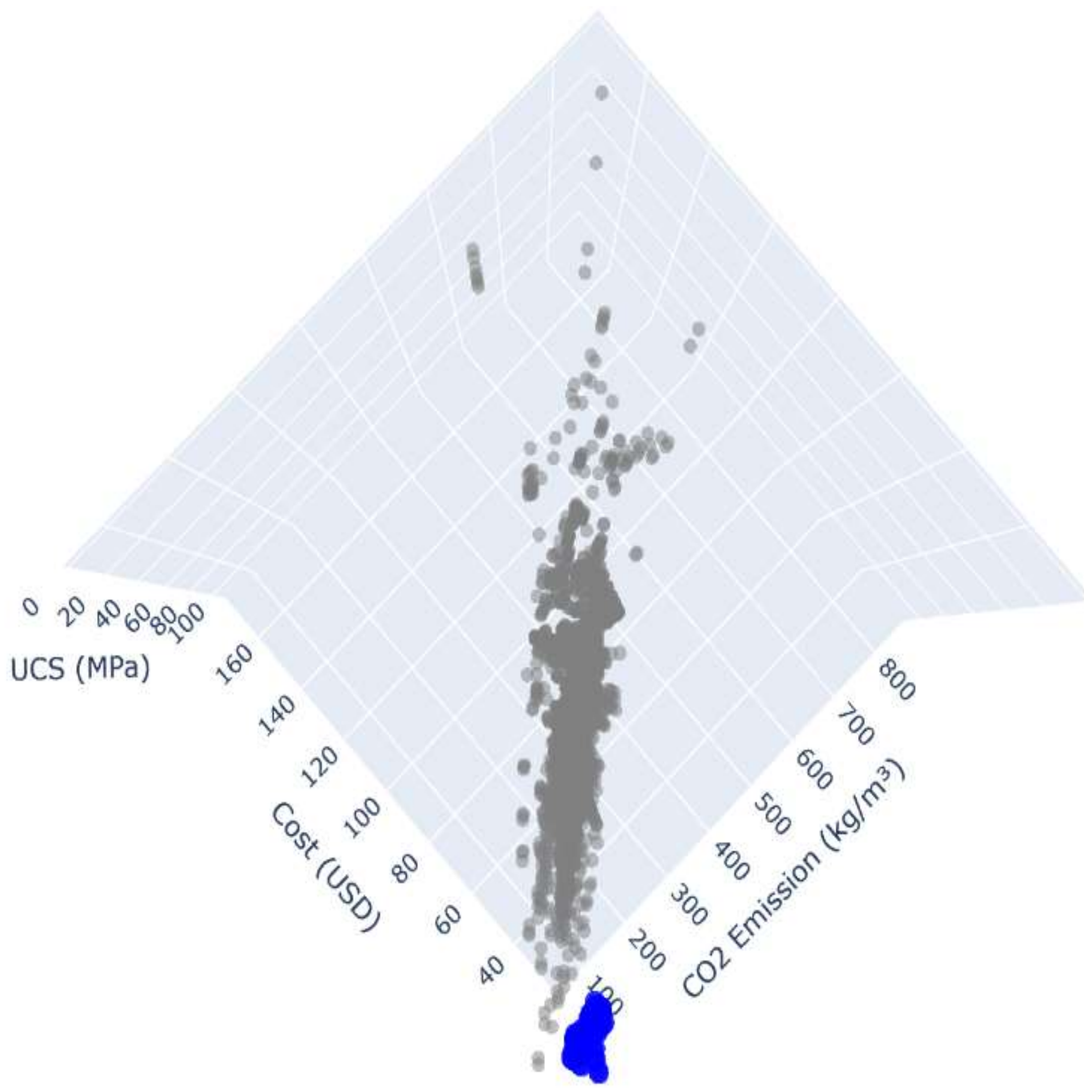


FIGURE 4.21.3D PARETO FRONT FOR UCS, COST AND CO2 EMISSION

5 CHAPTER
CONCLUSION

The study successfully developed a computational framework to predict and optimize the compressive strength (UCS) of metakaolin (MK)-based geopolymer concrete (GPC) using advanced machine learning (ML) models. A comprehensive dataset comprising 1854 data points, including 25 input features such as chemical composition, mix design parameters, and curing conditions, was curated for this purpose. Ensemble models, particularly XGBoost and GBM, outperformed individual models like RF, BPNN, SVM, and DT in predictive accuracy. XGBoost emerged as the best-performing model for UCS prediction, with metrics such as $R^2 = 0.9121$, RMSE = 4.1408 MPa, MAE = 2.8630 MPa, and MAPE = 10.3295%.

Feature importance analysis using Gini importance and SHapley Additive exPlanations (SHAP) identified critical parameters influencing UCS, including the ratios of coarse aggregate (CA) to fine aggregate (FA), H_2O to Na_2O molar ratios, sodium hydroxide (SH) concentration, and the volume of added water. Partial Dependency Plot (PDP) analysis revealed that CA-to-FA ratios exceeding 1, H_2O/Na_2O molar ratios below 10, minimal water addition, and SH concentrations above 10 significantly enhance UCS. Furthermore, the analysis highlighted strong interactions between features, notably the SiO_2/NaO ratios and low H_2O/Na_2O ratios, as key determinants of UCS.

For NSGA-II optimization, XGBoost models were employed to predict cost and CO_2 emissions alongside UCS. These models demonstrated exceptional performance, with the cost model achieving R^2 values of 1.0000 (train) and 0.9934 (test), and a test MAE of 0.2232. The CO_2 model similarly performed well, with R^2 values of 0.9994 (train) and 0.9954 (test), and a test MAE of 2.1485. These results underscore the robustness of XGBoost for multi-objective optimization tasks, enabling effective trade-offs between compressive strength, cost, and CO_2 emissions.

This study highlights the potential of integrating machine learning and optimization techniques, such as NSGA-II, to design sustainable and cost-effective GPC. Parametric studies further validated the predictive and interpretive capabilities of the ML models, offering actionable insights into mix design strategies for advancing GPC production. Ensemble ML models, coupled with interpretive tools like SHAP and PDP, pave the way for efficient, sustainable concrete mix designs.

6 REFERENCES

 Eftekhar Afzali, S. A., Shayanfar, M. A., Ghanooni-Bagha, M., Golafshani, E., & Ngo, T. (2024). The use of machine learning techniques to investigate the properties of metakaolin-based geopolymer concrete. *Journal of Cleaner Production*.

Shobeiri, V., Bennett, B., Xie, T., & Visintin, P. (2021). A comprehensive assessment of the global warming potential of geopolymer concrete. *Journal of Cleaner Production, 290,* 125667.

Abbas, R., Khereby, M.A., Ghorab, H.Y., Elkhoshkhany, N., 2020. Preparation of geopolymer concrete using Egyptian kaolin clay and the study of its environmental effects and economic cost. Clean Technol. Environ. Policy 22, 669–687.

Ahmad, A., Ahmad, W., Aslam, F., Joyklad, P., 2022a. Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques. Case Stud. Constr. Mater. 16, e00840.

Ahmed, H.U., Abdalla, A.A., Mohammed, A.S., Mohammed, A.A., 2022b. Mathematical modeling techniques to predict the compressive strength of high-strength concrete incorporated metakaolin with multiple mix proportions. Cleaner Materials 5, 100132.

Ahmed, H.U., Mohammed, A.A., Mohammed, A., 2022c. Soft computing models to predict the compressive strength of GGBS/FA-geopolymer concrete. PLoS One 17 (5), e0265846.

Albidah, A., Alghannam, M., Abbas, H., Almusallam, T., Al-Salloum, Y., 2021. Characteristics of metakaolin-based geopolymer concrete for different mix design parameters. J. Mater. Res. Technol. 10, 84–98.

Albidah, A., Alqarni, A.S., Abbas, H., Almusallam, T., Al-Salloum, Y., 2022a. Behavior of Metakaolin-Based geopolymer concrete at ambient and elevated temperatures. Construct. Build. Mater. 317, 125910.

Albidah, A., Alsaif, A., Abadel, A., Abbas, H., Al-Salloum, Y., 2022b. Role of recycled vehicle tires quantity and size on the properties of metakaolin-based geopolymer rubberized concrete. J. Mater. Res. Technol. 18, 2593–2607.

Albidah, A., Altheeb, A., Alrshoudi, F., Abadel, A., Abbas, H., Al-Salloum, Y., 2020. Bond Performance of GFRP and Steel Rebars Embedded in Metakaolin Based Geopolymer Concrete, Structures. Elsevier, pp. 1582–1593.

Alghannam, M., Albidah, A., Abbas, H., Al-Salloum, Y., 2021. Influence of critical parameters of mix proportions on properties of MK-based geopolymer concrete. Arabian J. Sci. Eng. 46 (5), 4399–4408.

Asteris, P.G., Lourenço, P.B., Roussis, P.C., Adami, C.E., Armaghani, D.J., Cavaleri, L., Chalioris, C.E., Hajihassani, M., Lemonis, M.E., Mohammed, A.S., 2022. Revealing the nature of metakaolin-based concrete materials using artificial intelligence techniques. Construct. Build. Mater. 322, 126500.

Ayeni, O., Onwualu, A.P., Boakye, E., 2021. Characterization and mechanical performance of metakaolin-based geopolymer for sustainable building applications. Construct. Build. Mater. 272, 121938.

Bature, A., Khorami, M., Ganjian, E., Tyrer, M., 2021. Influence of alkali activator type and proportion on strength performance of calcined clay geopolymer mortar. Construct. Build. Mater. 267, 120446.

Chaabene, W.B., Flah, M., Nehdi, M.L., 2020. Machine learning prediction of mechanical properties of concrete: critical review. Construct. Build. Mater. 260, 119889.

da Silva Rocha, T., Dias, D.P., França, F.C.C., de Salles Guerra, R.R., de Oliveira, L.R.d.C., 2018. Metakaolin-based geopolymer mortars with different alkaline activators (Na+ and K+). Construct. Build. Mater. 178, 453–461.

Degefu, D.M., Liao, Z., Berardi, U., Labb´ e, G., 2022. The dependence of thermophysical and hygroscopic properties of macro-porous geopolymers on Si/Al. J. Non-Cryst. Solids 582, 121432.

Duxson, P., Mallicoat, S.W., Lukey, G.C., Kriven, W.M., Van Deventer, J.S., 2007. The effect of alkali and Si/Al ratio on the development of mechanical properties of metakaolin-based geopolymers. Colloids Surf. A Physicochem. Eng. Asp. 292 (1), 8–20.

Environment, U., Scrivener, K.L., John, V.M., Gartner, E.M., 2018. Eco-efficient cements: potential economically viable solutions for a low-CO2 cement-based materials industry. Cement Concr. Res. 114, 2–26.

Gomaa, E., Han, T., ElGawady, M., Huang, J., Kumar, A., 2021. Machine learning to predict properties of fresh and hardened alkali-activated concrete. Cement Concr. Compos. 115, 103863.

Heath, A., Paine, K., McManus, M., 2014. Minimising the global warming potential of clay based geopolymers. J. Clean. Prod. 78, 75–83.

Jiang, X., Xiao, R., Bai, Y., Huang, B., Ma, Y., 2022. Influence of waste glass powder as a supplementary cementitious material (SCM) on physical and mechanical properties of cement paste under high temperatures. J. Clean. Prod. 340, 130778.

Jiang, X., Xiao, R., Ma, Y., Zhang, M., Bai, Y., Huang, B., 2020. Influence of waste glass powder on the physico-mechanical properties and microstructures of fly ash-based geopolymer paste after exposure to high temperatures. Construct. Build. Mater. 262, 120579.

Jiang, X., Zhang, Y., Zhang, Y., Ma, J., Xiao, R., Guo, F., Bai, Y., Huang, B., 2023. Influence of size effect on the properties of slag and waste glass-based geopolymer paste. J. Clean. Prod. 383, 135428.

Lahoti, M., Narang, P., Tan, K.H., Yang, E.-H., 2017. Mix design factors and strength prediction of metakaolin-based geopolymer. Ceram. Int. 43 (14), 11433–11441.

Moradikhou, A.B., Esparham, A., Avanaki, M.J., 2020. Physical & mechanical properties of fiber reinforced metakaolin-based geopolymer concrete. Construct. Build. Mater. 251, 118965.

Nguyen, K.T., Nguyen, Q.D., Le, T.A., Shin, J., Lee, K., 2020. Analyzing the compressive strength of green fly ash based geopolymer concrete using experiment and machine learning approaches. Construct. Build. Mater. 247, 118581.

Perez-Cortes, P., Escalante-Garcia, J.I., 2020. Alkali activated metakaolin with high limestone contents–Statistical modeling of strength and environmental and cost analyses. Cement Concr. Compos. 106, 103450.

Pouhet, R., Cyr, M., 2016. Formulation and performance of flash metakaolin geopolymer concretes. Construct. Build. Mater. 120, 150–160.

Provis, J.L., Arbi, K., Bernal, S.A., Bondar, D., Buchwald, A., Castel, A., Chithiraputhiran, S., Cyr, M., Dehghan, A., Dombrowski-Daube, K., 2019. RILEM TC 247-DTA round robin test: mix design and reproducibility of compressive strength of alkali-activated concretes. Mater. Struct. 52, 1–13.

Ribeiro, M.G.S., Ribeiro, M.G.S., Keane, P.F., Sardela, M.R., Kriven, W.M., Ribeiro, R.A. S., 2021. Acid resistance of metakaolin-based, bamboo fiber geopolymer composites. Construct. Build. Mater. 302, 124194.

Rovnaník, P., 2010. Effect of curing temperature on the development of hard structure of metakaolin-based geopolymer. Construct. Build. Mater. 24 (7), 1176–1183.

Rowles, M., O'connor, B., 2003. Chemical optimisation of the compressive strength of aluminosilicate geopolymers synthesised by sodium silicate activation of metakaolinite. J. Mater. Chem. 13 (5), 1161–1165.

Shamsabadi, E.A., Roshan, N., Hadigheh, S.A., Nehdi, M.L., Khodabakhshian, A., Ghalehnovi, M., 2022. Machine learning-based compressive strength modelling of concrete incorporating waste marble powder. Construct. Build. Mater. 324, 126592.

Shehata, N., Mohamed, O., Sayed, E.T., Abdelkareem, M.A., Olabi, A., 2022. Geopolymer concrete as green building materials: recent applications, sustainable development and circular economy potentials. Sci. Total Environ. 836, 155577.

Singh, B., Ishwarya, G., Gupta, M., Bhattacharyya, S., 2015. Geopolymer concrete: a review of some recent developments. Construct. Build. Mater. 85, 78–90.

Valencia-Saavedra, W.G., de Guti´ errez, R.M., Puertas, F., 2020. Performance of FA-based geopolymer concretes exposed to acetic and sulfuric acids. Construct. Build. Mater. 257, 119503.

Yunsheng, Z., Wei, S., Zongjin, L., 2010. Composition design and microstructural characterization of calcined kaolin-based geopolymer cement. Appl. Clay Sci. 47 (3–4), 271–275.

Zhang, L.V., Marani, A., Nehdi, M.L., 2022. Chemistry-informed machine learning prediction of compressive strength for alkali-activated materials. Construct. Build. Mater. 316, 126103.

Zhang, M., Zhang, C., Zhang, J., Wang, L., Wang, F., 2023. Effect of composition and curing on alkali activated fly ash-slag binders: machine learning prediction with a random forest-genetic algorithm hybrid model. Construct. Build. Mater. 366, 129940.

Zhu, X., Li, W., Du, Z., Zhou, S., Zhang, Y., Li, F., 2021. Recycling and utilization assessment of steel slag in metakaolin based geopolymer from steel slag by-product to green geopolymer. Construct. Build. Mater. 305, 124654.

7 BIBLIOGRAPHY

al., S. e. (2022). Geopolymer concrete as green building materials: recent applications, sustainable developmentand circular economy potentials.

Eftekhar Afzali, S. A., Shayanfar, M. A., Ghanooni-Bagha, M., Golafshani, E., & Ngo, T. (2024). The use of machine learning techniques to investigate the properties of metakaolin-based geopolymer concrete. *Journal of Cleaner Production*.

Jiang. (n.d.). Influence of waste glass powder as a supplementary cementitious material (SCM) on physical and mechanical properties of cement paste under high temperatures.