

**MACHINE LEARNING-AIDED OPTIMIZATION OF
ULTRA-HIGH PERFORMANCE CONCRETE
COMPRESSIVE STRENGTH**

MUHAMMED OLAWALE BALOGUN

CIVIL ENGINEERING

DECEMBER 2024

**MACHINE LEARNING-AIDED OPTIMIZATION
OF ULTRA-HIGH PERFORMANCE CONCRETE
COMPRESSIVE STRENGTH**

BY

MUHAMMED OLAWALE BALOGUN

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

CIVIL ENGINEERING

DECEMBER, 2024

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **MUHAMMED OLAWALE BALOGUN** 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 CIVIL ENGINEERING**.



Dr. Shamshad Ahmad
Department Chairman



Dr. Suliman S. Al-Homidan
Dean of Graduate Studies

Date



Dr. Saheed K. Adekunle
(Advisor)



Dr. Sunday O. Olusanya
(Member)



Dr. Zaher M. Yaseen
(Member)
Zaher M. Yaseen



© Muhammed Olawale Balogun

2024

[This work is dedicated to my mother (Mrs Amina Sadiq Balogun) and my late father (Mr
Moshood Kolawole Balogun) |

ACKNOWLEDGMENTS

First and foremost, all praises and thanks are due to Allah (Subhanahu wa Ta'ala) for His guidance and protection throughout my MSc journey. I am deeply grateful to Allah for seeing me through this program. May the peace and blessings of Allah be upon the Messenger (Sallallahu alayhi wasallam), his household, companions, and all who follow his path until the Day of Judgment.

I am profoundly thankful to my mother for her unwavering support and prayers throughout this journey. My heartfelt gratitude also goes to my wife and children for their patience and support during my master's degree.

I sincerely appreciate the government of Saudi Arabia for creating a conducive environment for my studies and the institution for awarding me this scholarship. My deepest thanks go to my thesis advisor, Dr. Saheed Adekunle, for his invaluable support and guidance throughout my program. May Allah reward you abundantly. I am also grateful to Dr. Sunday Olusanya and Dr. Zaher Yaseen for their insights and encouragement during my research. I thank Dr. Shamsad Ahmad, head of the Civil Engineering Department, and the entire departmental staff for their support.

Lastly, my sincere gratitude goes to the Nigerian community for their steadfast support from the first day of my arrival until the completion of my program. Jazakumullahu Khairan. |

TABLE OF CONTENTS

ACKNOWLEDGMENTS	V
TABLE OF CONTENTS	VI
LIST OF TABLES	X
LIST OF FIGURES	XI
LIST OF ABBREVIATIONS	XII
ABSTRACT	XIV
ملخص الرسالة	XV
CHAPTER 1 INTRODUCTION	1
1.1 Background of Study	1
1.2 Research Significance	3
1.3 Research Objectives	3
1.4 Scope of the study	4
1.5 Thesis Organization	5
CHAPTER 2 LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Data Collection	7
2.2.1 Experimental Data from Published Journals	8
2.2.2 Combining Literature Data with Experimental Results	8
2.2.3 Size of Datasets	9
2.3 Data Preparation and Preprocessing	10
2.3.1 Handling Missing Data	11

2.3.2	Outlier Removal	11
2.3.3	Standardization and Normalization	11
2.3.4	Data Partitioning.....	12
2.4	Prediction Variables for UHPC Compressive Strength	12
2.4.1	Number of Features for UHPC Compressive Strength Prediction.....	19
2.5	Machine Learning Algorithms.....	20
2.5.1	Artificial Neural Network (ANN)	20
2.5.2	Deep Neural Network (DNN)	21
2.5.3	Support Vector Regressor (SVR).....	21
2.5.4	Decision Tree Regressor (DTR)	22
2.5.5	Random Forest Regressor (RFR)	23
2.5.6	Extra Trees Regressor	23
2.5.7	Gradient Boosting Regressor (GBR)	24
2.5.8	K-Nearest Neighbors (KNN)	26
2.5.9	Other Ensemble Methods	26
2.6	Performance Measure	27
2.6.1	Coefficient of Determination (R^2).....	28
2.6.2	Adjusted R^2	28
2.6.3	Root Mean Squared Error (RMSE)	29
2.6.4	Mean Absolute Error (MAE).....	29
2.6.5	Mean Absolute Percentage Error (MAPE).....	30
2.6.6	Most Commonly Used Evaluation Metrics	30
2.7	Hyperparameter Optimization	30
2.7.1	Trial and Error	31
2.7.2	Grid Search Optimization	31
2.7.3	Random Search Optimization.....	32

2.7.4	Bayesian Optimization.....	32
2.8	Feature Selection and Engineering.....	33
2.8.1	Sequential Feature Selection	34
2.8.2	Automated Feature Engineering and Selection (Autofeat)	34
2.8.3	Effects of Feature Selection on the Input Variables.....	35
2.9	Explainable AI (XAI) Techniques	35
2.9.1	SHapley Additive exPlanations (SHAP).....	35
2.9.2	Neural Interpretation Diagram (NID).....	36
2.9.3	Partial Dependence Plot (PDP)	36
2.10	Research Gap	37
 CHAPTER 3 INTEGRATION OF FEATURE OPTIMIZATION METHODS AND MACHINE LEARNING MODELS FOR ULTRA-HIGH-PERFORMANCE CONCRETE (UHPC) COMPRESSIVE STRENGTH PREDICTION.....		38
3.1	Introduction.....	38
3.2	Collected Data Explanation	42
3.2.1	Description of Dataset.....	42
3.2.2	Statistical Analysis of Dataset	44
3.2.3	Exploratory Data Analysis (EDA)	45
3.3	Methodological Overview.....	49
3.3.1	K-Nearest Neighbors.....	49
3.3.2	Decision Tree Regressor	49
3.3.3	Random Forest Regressor	50
3.3.4	Extreme Gradient Boost Regressor	50
3.3.5	Support Vector Regressor	51
3.3.6	Linear Regression	51
3.3.7	Research Modeling Procedure	52

3.3.8	Model Evaluation	53
3.3.9	Optimization Strategy.....	54
3.4	Results and Discussion.....	55
3.4.1	Initial Iteration Results.....	55
3.4.2	Comparing Results with Similar Studies	58
3.4.3	Effects of Feature Selection on the Dataset.....	59
3.5	Conclusion and Recommendation	63
3.5.1	Conclusion	63
3.5.2	Recommendations	63
CHAPTER 4 CONCLUSIONS AND RECOMMENDATIONS		65
4.1	Conclusion	65
4.2	Recommendations for Practical Application.....	66
4.2.1	Adoption in Industry	66
4.2.2	Best Practices for Model Use.....	66
4.3	Limitations and Future Work.....	67
4.3.1	Data Limitations.....	67
4.3.2	Model Interpretability	68
4.3.3	Opportunities for Future Research	68
REFERENCES.....		70
VITAE		79

LIST OF TABLES

Table 2-1: Summary of published work on UHPC compressive strength.....	9
Table 3-1: Predictive Models in Previous Studies of UHPC Compressive Strength	40
Table 3-2: Description of features in the dataset with their units	43
Table 3-3: Summary statistics of each feature in the dataset	44
Table 3-4: Correlation coefficient between each input feature and the target (fcu)	44
Table 3-5: Performance Measure of Each Model for both training and test sets.....	56
Table 3-6: Hyperparameter Optimization of Each Model Before Feature Selection	56
Table 3-7: Comparison of study's performance with previous studies	59
Table 3-8: Performance Measure of Each Model for both training and test sets.....	60
Table 3-9: Hyperparameter Optimization of Each Model After Feature Selection .	61

LIST OF FIGURES

Figure 1-1: Precast Bridge Girder Made with UHPC (Lyksor.com).....	1
Figure 2-1: Diagrammatic Representation of ANN (Medium.com).....	20
Figure 2-2: Diagrammatic Representation of DNN (GeeksforGeeks.org)	21
Figure 2-3: Simplified Representation of SVM (Medium.com).....	22
Figure 2-4: Typical Structure of DT (Smartdraw.com)	22
Figure 2-5: Simplified Representation of RF (Medium.com)	23
Figure 2-6: Graphical Representation of GB (Medium.com)	25
Figure 2-7: KNN Visual Representation (Javapoint.com)	26
Figure 2-8: Stacking Ensemble Method.....	27
Figure 2-9: Voting Ensemble Method	27
Figure 2-10: Visual representation of Grid and Random Search (Sabrepc.com)	32
Figure 3-1: Feature Correlation Heatmap	46
Figure 3-2: Feature Distribution	47
Figure 3-3: Scatter Plot of Each Feature with the Target.....	48
Figure 3-4: Experimental Setup	53
Figure 3-5: Feature Importance for Tree-based Models before Feature Selection ..	57
Figure 3-6: Test performance of each ML model before feature selection	57
Figure 3-7: Feature Importance for Tree-based Models after Feature Selection.....	61
Figure 3-8: Testing Performance of each ML Model with Feature Selection	62

LIST OF ABBREVIATIONS

ML	:	Machine Learning
UHPC	:	Ultra-High Performance Concrete
RFR	:	Random Forest Regressor
XGBR	:	Extreme Gradient Boost Regressor
ETR	:	Extra Tree Regressor
LR	:	Linear Regression
KNN	:	K-Nearest Neighbors
SVR	:	Support Vector Regressor
DTR	:	Decision Tree Regressor
ANN	:	Artificial Neural Network
ABR	:	Adaboost Regressor
GBR	:	Gradient Boost Regressor
AutoML	:	Automated Machine Learning
BR	:	Bagging Regressor
VR	:	Voting Regressor
DNN	:	Deep Neural Network
LGBMR	:	Light Gradient Boost Machine Regressor
CBR	:	Catboost Regressor
SR	:	Stacking Regressor
GEP	:	Gene Expression Programming
EDA	:	Exploratory Data Analysis
ANFIS	:	Adaptive Neuro-Fuzzy Inference System
GNDO		Generalized Normal Distribution Optimization

COA		Coots Optimization Algorithms
HBA	:	Honey Badger Algorithm
ANGN		ANFIS + GNDO
ANCO		ANFIS + COA
ANHB		ANFIS + HBA

ABSTRACT

Full Name : [Muhammed Olawale Balogun]
Thesis Title : [Machine-learning-aided optimization of ultra-high-performance concrete (UHPC) compressive strength]
Major Field : [Civil Engineering]
Date of Degree : [December 2024]

Ultra-high-performance concrete (UHPC) is known for its superior mechanical properties and durability for infrastructure applications. The prediction of UHPC compressive strength is crucial in the preliminary stages of structural design. Traditional empirical mixture formulations for non-proprietary UHPC are often labor-intensive and time-consuming, highlighting the need for more efficient approaches. This study explores the feasibility of machine learning (ML) models to optimize and predict the UHPC compressive strength. Several ML models were employed for the prediction process, including K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), Support Vector Machine (SVM), and Linear Regression. For the modeling purpose, a dataset based on UHPC mix design parameters was collected from the open-source literature, comprising 810 observations with 15 input parameters (“features”). At first, the correlation rank feature selection technique was adopted to reduce the input parameters from 15 to 8, thereby optimizing the prediction of UHPC compressive strength. Model evaluation metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and determination coefficient (R^2), were used to assess the model’s predictive accuracy. Results indicated that RF and XGB models outperformed other models in predictive accuracy. The findings underscore the potential of ML models to enhance the efficiency and cost-effectiveness of UHPC mix design significantly. This study provides a robust computer-aided framework for rapid and precise optimization of UHPC, paving the way for its broader application in the construction industry.

ملخص الرسالة

الاسم الكامل: محمد أولوال والي بالوگون

عنوان الرسالة: التحسين بمساعدة التعلم الآلي لقوة الانضغاط للخرسانة فائقة الأداء (UHPC)

التخصص: الهندسة المدنية

تاريخ الدرجة العلمية: ديسمبر 2024

الخرسانة فائقة الأداء (UHPC) معروفة بخصائصها الميكانيكية الفائقة والمتانة في تطبيقات البنية التحتية. يُعتبر التنبؤ بمقاومة الانضغاط للـ UHPC أمراً بالغ الأهمية كمرحلة سابقة لإجراءات تصميم الهياكل. تُعد الصيغ التجريبية التقليدية للـ UHPC غالباً مجتهدة وتستغرق وقتاً طويلاً، مما يستدعي الحاجة إلى نهج أكثر كفاءة. تستكشف هذه الدراسة جدوى استخدام نماذج التعلم الآلي (ML) لتحسين وتوقع مقاومة الانضغاط للـ UHPC. تم استخدام عدة نماذج للتعلم الآلي في عملية التنبؤ، بما في ذلك نموذج الجيران الأقرب (KNN)، شجرة القرار (DT)، الغابة العشوائية (RF)، تعزيز التدرج المتطرف (XGB)، آلة المتجهات الداعمة (SVM)، بالإضافة إلى الانحدار الخطي. لأغراض النمذجة، تم جمع بيانات الـ UHPC بناءً على معايير تصميم الخليط من الأدبيات مفتوحة المصدر والتي شملت 810 ملاحظة و15 معلمة إدخال "ميزات". في البداية، تم اعتماد تقنية اختيار الميزات بناءً على تصنيف الارتباط لتقليل معلومات الإدخال من 15 إلى 8، مما أدى إلى تحسين التنبؤ بمقاومة الانضغاط للـ UHPC. تم استخدام مقاييس تقييم النماذج، بما في ذلك متوسط الخطأ المطلق (MAE)، الجذر التربيعي لمتوسط الخطأ المربع (RMSE)، ومعامل التحديد (R^2)، لتقييم دقة توقع النموذج. أشارت النتائج إلى أن نماذج الـ RF والـ XGB تفوقت على النماذج الأخرى في دقة التنبؤ. وتؤكد هذه النتائج على إمكانيات نماذج التعلم الآلي في تحسين كفاءة وفعالية تكلفة تصميم الخليط للـ UHPC بشكل كبير. تقدم هذه الدراسة إطاراً قوياً يعتمد على الحاسوب يمكن استخدامه لتحسين تصميم الـ UHPC بسرعة ودقة، مما يمهد الطريق لتطبيقه بشكل أوسع في صناعة البناء.

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Ultra-high-performance concrete (UHPC) represents a significant advancement in construction materials, characterized by its superior mechanical properties, durability, and resistance to environmental stressors [1]. UHPC is typically formulated with a high binder content, low water-to-binder ratio, supplementary cementitious materials (e.g. silica fume), mineral admixtures, and steel fibers [2]. This composition not only enhances UHPC's strength and durability but also ensures its performance under extreme conditions, making it ideal material for critical infrastructure applications such as high-rise buildings, structural repairs, and bridges [3], [4].



Figure 1-1: Precast Bridge Girder Made with UHPC (Lyksor.com)

The optimization of UHPC, particularly its compressive strength, is critical for maximizing its potential. Compressive strength is a key performance indicator that determines the material's capacity to withstand axial loads, which is vital for structural integrity and safety [5]. Traditionally, optimizing UHPC mix designs to achieve desired compressive strengths has involved empirical methods and extensive laboratory testing [6]. While these methods have been somewhat effective, they are often labor-intensive, and time-consuming, and do not fully capture the complex interactions between various constituents and their effects on the final concrete properties [6].

With advancements in computational technologies, machine learning (ML) has emerged as a powerful tool for modeling and predicting the properties of complex materials [7]. Machine learning algorithms can analyze large datasets to identify patterns and correlations that are not apparent through traditional methods. By applying ML techniques to UHPC mix design, researchers can develop predictive models that accurately estimate compressive strength based on various input parameters. This approach not only accelerates the optimization process but also enhances the precision of predictions, leading to more efficient and cost-effective mix designs [8], [9].

This study aims to leverage ML to optimize the compressive strength of UHPC. The study involved developing and training ML models using a dataset of UHPC mix compositions and their corresponding compressive strengths. Key aspects of the study include feature selection to identify the most influential parameters, hyperparameter tuning to improve model performance, and validation of the models using experimental data. The ultimate goal is to create a robust, data-driven framework for UHPC mix design optimization that

can be used to achieve higher compressive strengths with greater efficiency than traditional methods.

1.2 Research Significance

The data-driven approach to the prediction of UHPC compressive strength is very promising as it eliminates or reduces the cost of the experimental approach to UHPC material property optimization. The machine learning (ML) techniques in previous studies faced a lot of limitations: the lack of sufficient data to train the ML models which causes low accuracy and generalization issues in some of the existing models, the high dimensionality of experimental data due to the large number of features used in predicting UHPC compressive strength, and the non-availability of most of the existing models to end users due to the complexity of the ML algorithms used.

This research plans to address these limitations by developing ML models that are robust enough to predict UHPC compressive strength with high accuracy and easy enough to apply by end users. Feature selection approaches will be explored to reduce the high dimensionality of the datasets and optimize the process significantly.

1.3 Research Objectives

The main objective of this study is to optimize the compressive strength of ultra-high-performance concrete (UHPC) through machine learning principles. To achieve this objective, the following specific objectives are outlined:

1. To analyze the relationship between UHPC mixture parameters and compressive strength through machine learning (ML) principles.
2. To develop ML models capable of predicting UHPC compressive strength.
3. To optimize UHPC mixture parameters using the developed ML models targeting enhanced compressive strength.
4. To investigate the effect of feature selection on the performance of the ML predictive models for UHPC compressive strength

1.4 Scope of the study

The scope of this study focuses on the application of machine learning (ML) models to predict and optimize the compressive strength of Ultra-High-Performance Concrete (UHPC). By utilizing a dataset of 810 observations, the study employs various ML algorithms, such as RF, XGB, and SVM, and compares their performance based on three performance metrics (R^2 , MAE, and RMSE). The study also incorporates feature selection techniques to identify key UHPC mix parameters, improving model accuracy and reducing complexity. Additionally, the study addresses the limitations of traditional empirical methods for UHPC optimization, proposing ML as a more efficient and accurate approach for understanding the relationships between mixed design components and compressive strength, while exploring the prospects of ML in civil engineering materials research.

1.5 Thesis Organization

The thesis is organized into four main chapters. The outline of each chapter is given below:

- Chapter 1 introduces the background of research, significance, objectives, scope, and thesis organization.
- Chapter 2 presents the literature review.
- Chapter 3 presents an article manuscript, titled “Integration of feature optimization methods and machine learning models for ultra-high performance concrete (UHPC) compressive strength prediction”.
- Chapter 4 presents the over conclusion drawn from the thesis and suggested recommendations for future studies.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Ultra-high-performance concrete (UHPC) is a highly durable and mechanically robust material that has gained considerable attention in civil engineering [10]. Characterized by its dense microstructure, low permeability, and exceptional compressive strength, UHPC is frequently used in applications demanding superior performance, such as high-rise buildings, bridges, and marine structures [11]. The optimization of UHPC properties, particularly compressive strength, is essential for maximizing its benefits and enhancing its use in various structural applications [12].

Machine learning (ML) has emerged as a powerful tool for predicting and optimizing material properties, offering a data-driven alternative to traditional empirical methods [13]. By analyzing complex relationships among mix design parameters, ML models enable accurate predictions of UHPC performance based on diverse input variables, such as cement content, silica fume, fiber additions, and curing conditions [14]. These models can capture intricate, non-linear interactions among variables that empirical models often overlook, thereby improving efficiency in UHPC mix design and reducing experimental costs [15].

The application of ML in UHPC research has expanded significantly in recent years, with studies exploring various predictive models, feature selection techniques, and hyperparameter optimization methods. Despite these advancements, challenges remain in data availability, model interpretability, and the complexity of UHPC mix design. This literature review aims to assess the current state of ML applications in UHPC, identifying key techniques, dataset characteristics, performance measures, and explainability methods that contribute to accurate and interpretable UHPC predictions.

This review discusses data collection approaches and dataset characteristics, including dataset size and preprocessing methods. It further delves into the key variables influencing UHPC compressive strength. The ML algorithms commonly applied to UHPC are explored and the performance metrics used to evaluate model accuracy are reviewed. Hyperparameter optimization techniques and feature selection methods are also discussed. The review closes with Explainable AI (XAI) techniques, highlighting approaches to interpret model outputs and gain insights into the contribution of various UHPC mix components.

2.2 Data Collection

Data collection plays a pivotal role in developing accurate predictive models for Ultra-High-Performance Concrete (UHPC). The reviewed studies adopted various strategies to gather comprehensive datasets, ensuring the models were trained with diverse and representative data. Many studies targeted key mix components such as cement, silica fume, fly ash, slag, fibers, and chemical admixtures, focusing on their impact on UHPC

compressive strength. This targeted data collection enabled researchers to isolate and understand the effects of individual components on UHPC performance, leading to more precise model predictions. The data collection efforts primarily focused on two main approaches: extracting data from existing literature and combining this with experimental results.

2.2.1 Experimental Data from Published Journals

A significant portion of the studies sourced data from comprehensive literature reviews [16], [17], [18], [19]. This involved compiling information from previously published research on UHPC mix designs, testing methodologies, and material properties. By leveraging the results of past experiments, researchers assembled large datasets, often comprising hundreds or even thousands of data points. These datasets provided a broad foundation for developing machine learning models, with a wide array of UHPC formulations represented.

2.2.2 Combining Literature Data with Experimental Results

In some studies, researchers opted for a hybrid approach, augmenting data gathered from the literature with their experimental findings [20], [21], [22]. This strategy enriched the datasets by incorporating real-world experimental data, which helped validate or extend the literature. By blending theoretical data with practical results, researchers were able to capture a more comprehensive range of UHPC performance characteristics, thus improving the robustness and applicability of their predictive models.

2.2.3 Size of Datasets

The dataset size plays a crucial role in training machine learning models for accurately predicting the compressive strength and other properties of Ultra-High-Performance Concrete (UHPC). The studies reviewed varied significantly in dataset sizes, ranging from hundreds to nearly a thousand observations, with data sourced from experimental work and published literature. Larger datasets generally improve model accuracy and generalization. Table 2-1 shows the summary of journal publications on UHPC compressive strength prediction using ML techniques from 2012 to 2024. The data size used in each journal is presented in the table which ranges from slightly above 50 observations to slightly below 1000 observations.

Table 2-1: Summary of published work on UHPC compressive strength

Year	ML Algorithm(s)	#IP	Input Parameters (IP)	Data Size	R2	Ref
2012	ANN	7	C, S, SF, QP, W, SP, Fi	53	0.96 - 0.99	[23]
2017	ANN	11	C, S, CA, W, FA, SF, SP, W/B, TA/B, S/TA, Dmin–Dmax	78	0.998	[24]
2018	ANN	5	W/B, Df, Lf, AsF, Fi	162	0.990	[25]
2020	ANN	8	C, Fi, FA, S, SF, QP, W, Ad	110	0.801	[26]
2020	RFR, GBR, ETR	15	C, SF, SL, FA, QP, LP, NS, W, Fi, SP, S, CA, T, RH, Age	810	0.926 - 0.962	[6]
2021	GEP-ANN, ANN	5	W/B, C, SF, LP, S	80	0.831 - 0.970	[27]
2022	RFR	17	C, SF, FA, SL, GP, RHA, FC3R, MK, LP, W, SP, QP, TA, Dmax, W/B, W/P, VPD	931	0.904	[28]
2022	ANN	12	C, SF, GP, FC3R, LP, QP, SP, W, S, W/B, W/P, VPD, Age	265	0.960 - 0.983	[29]
2022	XGBR	17	C, SF, FA, SL, GP, RHA, FC3R, MK, LP, W, SP, QP, TA, Dmax, W/B, W/P, VPD	931	0.892	[30]

Year	ML Algorithm(s)	#IP	Input Parameters (IP)	Data Size	R2	Ref
2022	XGBR, ABR, BR	10	C, FA, SL, SF, S, W, SP, Fi, AsF, Age	372	0.78 - 0.90	[12]
2022	DTR, BR, GBR	21	C, FA, SL, NS, QP, LP, S, CA, SP, Fi, Fp, W, CT, CS, CSC, Dmax, Lf, Df, Lp, Dp	626	0.94 - 0.95	[31]
2022	AutoML	24	C, FA, SF, SL, NS, QP, LP, S, CA, SP, Fi, Fp, W, CT, CS, CSC, Dmax, Lf, Df, Lp, Dp, Age, Si, Li	626	0.87 - 0.99	[32]
2023	DNN, XGBR, ETR, VR, GBR	15	C, SF, SL, FA, QP, LP, NS, W, Fi, SP, S, CA, T, RH, Age	810	0.98 - 0.99	[33]
2023	XGBR	12	C, SF, QP, GP, TA, W, SP, Dmax, W/B, W/P, VPD	309	0.95	[34]
2023	AutoML	16	C, CT, FA, SL, SF, MK, NS, LP, QP, S, Dmax, W, SP, Fi, Si, AsF	379	0.95	[35]
2023	XGBR, ABR, CBR, RFR, LGBMR	10	CS, SF/C, S/C, Age, SP/C, W/B, W/C, Fi, Lf/Df, L/B	400	0.883 - 0.945	[36]
2023	XGBR, ABR, BR, SR	15	C, SF, SL, FA, QP, LP, NS, W, Fi, SP, S, CA, T, RH, Age	810	0.951 - 0.958	[37]
2023	GEP	15	C, SF, SL, FA, QP, LP, NS, W, Fi, SP, S, CA, T, RH, Age	810	0.977 - 0.992	[38]
2023	ANHB, ANCO, ANGN	8	C, SF/C, FA/C, S/C, QP/C, W/C, Ad/C, Fi/C	132	0.987 - 0.996	[5]
2024	LassoR, AutofeatR	15	C, SF, LP, QP, FA, NS, SL, W, S, CA, Fi, SP, RH, T, Age, B	810	0.950 - 0.962	[39]

2.3 Data Preparation and Preprocessing

After collecting the relevant data from the literature and experimental results, thorough preprocessing was conducted to ensure the quality, consistency, and reliability of the datasets used for developing ML models. Data preprocessing is essential in ensuring that the ML models trained on UHPC datasets are both accurate and generalizable. This step is

critical in addressing potential issues like missing values, outliers, and inconsistencies that could negatively impact model performance.

2.3.1 Handling Missing Data

One of the common challenges in dataset preparation was handling missing values. Missing data points can introduce bias or reduce the accuracy of machine learning models if left unaddressed. Studies applied various techniques to manage this issue, such as imputation methods, where missing values were replaced by the mean/median or zeros. In some cases, records with a significant amount of missing data were excluded entirely to maintain the integrity of the dataset.

2.3.2 Outlier Removal

Outliers have extreme values that deviate significantly from other observations. They were identified and addressed to prevent skewed results. Outliers were detected and removed. In some cases, robust models, like XGBR and SVR, were used that were less sensitive to the influence of outliers, reducing the need for their removal while preserving the dataset's diversity.

2.3.3 Standardization and Normalization

To ensure consistency across different features, data standardization or normalization was often applied. Since UHPC mix components such as cement content, silica fume, and fibers exist in different units and scales, these values were transformed into a common scale to

prevent features with larger magnitudes from disproportionately influencing the model. Standardization (z-score normalization) and min-max scaling were the most common techniques applied to achieve this.

2.3.4 Data Partitioning

Once preprocessing was completed, the datasets were partitioned into training and testing subsets. A common practice involved splitting the dataset, typically with 70-80% allocated to training and 20-30% to testing. This partitioning ensured that the models could be trained on a large portion of the data while preserving an independent testing set to evaluate the model's performance. In some cases, cross-validation techniques such as k-fold cross-validation were applied to further improve the reliability of the model evaluation.

2.4 Prediction Variables for UHPC Compressive Strength

The mechanical performance of UHPC is influenced by multiple factors, which serve as critical input parameters for predicting its compressive strength. Experimental data sourced from the literature often encompasses a broad spectrum of UHPC mix design parameters. These variables represent the diverse constituents and production conditions of UHPC, each contributing significantly to its compressive strength. Numerous studies have identified the most influential parameters for predicting UHPC compressive strength, including age, fiber content, silica fume content, cement content, and water content. A detailed discussion of these key factors is provided below:

- **Cement content (C):** The amount of cement in the mix directly influences the binding properties of UHPC, playing a crucial role in compressive strength. Some studies [40], [41] included cement strength class (CSC) and cement type (CT) as part of the input variables for UHPC compressive strength prediction. UHPC typically requires a high volume of cementitious materials, often ranging from 700 to 1000 kg/m³ [42], [43]. While a high cement content is typically required, it is important to optimize the mix design to ensure efficient cement utilization and long-term performance [44], [45], [46]. Rajasekar A et al. found that higher cement content UHPC mix produced lower strength, which could be due to the inadequacy in the content of quartz powder [47].
- **Silica fume content (SF):** Silica fume fills the voids in the cement matrix, enhancing the packing density and reducing the porosity of UHPC mix, thereby contributing to both strength and durability [48], [49]. SF also reacts with calcium hydroxide (CH), a byproduct of cement hydration, to form additional calcium silicate hydrate (C-S-H) gel [45], [50]. This pozzolanic reaction is a key contributor to the high strength of UHPC [49]. While SF contributes to long-term strength development through the pozzolanic reaction, its effect on early strength is not as significant as nano-silica [51], [52]. Also, Sadrmomtazi et al. [46] has reported reduced UHPC compressive strength with increasing cement replacement by SF.
- **Slag content (SL):** Slag may include copper slag, steel slag, and phosphorous slag. The addition of slag as a cementitious material can improve UHPC's workability and strength, though its impact depends on the amount added to UHPC mix [47].

When used as a cement replacement in UHPC, SL was reported to cause reduction in the early-stage compressive strength of UHPC, although it has a marginal effect on its long-term strength [43].

- **Fly ash content (FA):** Fly ash improves the long-term strength development of UHPC [42], [45] and reduces permeability [53], making it a key feature in the mix design. The pozzolanic reactions of fly ash with calcium hydroxide generate additional calcium silicate hydrate (C-S-H), which strengthens the concrete matrix [45], [50]. Alsalman A et al. [45] found that at 28 and 90 days, UHPC mixtures with fly ash content of 20% and 30% showed greater compressive strength than a control mixture, while a mixture with 40% fly ash showed slightly less strength at 28 days. Wu Z et al. recommend a 20% replacement of cement with fly ash when producing UHPC [53].
- **Quartz powder content (QP):** Quartz powder enhances the material's density and strength by filling voids in the mix [54], [55], contributing to a more compact microstructure. In some cases, quartz powder can be used to replace a portion of the cement in UHPC. Yu R et al. found that replacing approximately 20% of the cement with quartz powder did not result in a significant decrease in compressive strength [44]. Janković K et al.'s study indicated that with the increase of quartz powder, the compressive and flexural strength of UHPC first increased and then decreased slightly when the content was more than 3% due to agglomeration of particles [55].

- Limestone powder content (LP):** Limestone powder is often used to enhance UHPC's workability and reduce cement content, indirectly influencing compressive strength [56]. It is also used as an alternative filler to quartz powder in UHPC, although its contribution to strength is slightly lower than that of quartz powder [54]. LP can accelerate the early hydration of cement due to its filler effect, which provides additional surfaces for the nucleation and growth of hydration products [54]. While LP can improve early strength, its impact on long-term strength is more complex. Wang C et al. found that UHPC mixtures with 20% LP had higher compressive strengths at 28, 56, 90, 180 and 365 days when compared to a mixture without LP. However, the mixture with 40% LP exhibited lower strengths when compared to the control mixture at these same ages [56].
- Nano silica content (NS):** Nano silica is added as a nano-filler to improve packing density of the UHPC matrix, enhancing both compressive and flexural strength in UHPC [55]. Nano-silica can significantly promote the early hydration of cement in UHPC, leading to a higher rate of strength development [51]. This is due to the high pozzolanic reactivity of nano-silica, which consumes calcium hydroxide (CH) and forms additional calcium silicate hydrate (C-S-H) [55], [57], [58]. Studies have shown that the ideal amount of nano-silica can range from 0.5% to 4% by weight of cementitious material [51], [55], [59]. Nano-silica has a higher specific surface area and pozzolanic reactivity compared to silica fume, making it more effective at improving the strength of UHPC [52].

- **Water content (W):** Water plays a critical role in the hydration process. However, excess water can weaken UHPC, so maintaining an optimal water-to-binder ratio is crucial for strength [45]. According to Sadrmomtazi et al. [46], the compressive strength of UHPC decreases by 11.5% as the water-to-binder ratio increases from 0.18 to 0.22.
- **Fine aggregate content (S):** Fine aggregates are the primary constituent material of UHPC whose particles are glued together by cementitious materials, and are essential for improving the stiffness of UHPC. Their particle size and distribution affect both strength and durability [60].
- **Coarse aggregate content (CA):** The inclusion of coarse aggregates, though minimal in UHPC, can still influence the material's compressive strength, especially when larger aggregates are used [50].
- **Fiber content (Fi):** Ductility is an essential feature of UHPC, which is achieved with the inclusion of micro- and macro-fibers in its mixture. Steel is the most used fiber to improve UHPC's toughness and resistance to cracking [61], [62]. Other fibers used include polystyrene and glass fibers. While they primarily affect tensile and flexural properties, fiber content can also impact compressive strength [44], [51], [61]. The increase in compressive strength is attributed to the fibers' ability to bridge cracks and restrict their propagation [62], [63], [64]. Polyethylene fibers generally reduce the compressive strength of UHPC when used as a replacement for steel fibers [65]. The addition of glass fibers initially increases the compressive

strength, but the strength increase stops at a certain fiber content, and the strength may even decrease with higher glass fiber content [63]. The impact of fibers on UHPC compressive strength depends on some factors such as fiber type, volume fraction, length, aspect ratio, orientation, dispersion, and the bond between the fibers and the matrix [63], [65].

- **Superplasticizer content (SP):** Superplasticizers are used to improve the workability of UHPC without increasing water content, making them a crucial variable in achieving both strength and fluidity in the mix [45], [46], [56]. Increasing SP from 1.2 to 1.4% by volume of concrete led to a 10% gain in UHPC compressive strength [46]. SPs function by dispersing the cement particles in the mix. This prevents agglomeration, resulting in a more homogenous mix and a more efficient hydration process [51].
- **Temperature (T):** Curing temperature affects the hydration process and strength development of UHPC [66]. Higher temperatures typically accelerate strength gain [50]. Compared to normal curing, UHPC specimens cured in steam and heat exhibit an average increase of about 8% and 14% in compressive strength, respectively, regardless of the inclusion of copper slag [47].
- **Relative humidity (RH):** Ambient humidity during curing impacts the hydration of UHPC, influencing its final strength [49]. A sufficient moisture supply ensures that cement particles can react with water to form the calcium-silicate-hydrate (C-S-H) gel responsible for the concrete's strength [45]. High RH promotes adequate

hydration, reduces drying shrinkage and the risk of cracking, promotes the pozzolanic reaction of SCMs, and helps in developing a dense, compact microstructure. Low RH, on the other hand, can lead to incomplete hydration, increased shrinkage, and reduced strength [46].

- **Curing time (Age):** The age of hardened UHPC, defined as the time elapsed from the addition of water to cement in the mixture, is one of the most important factors in predicting compressive strength. Compressive strength increases over time as hydration progresses [48].
- **Maximum size of aggregate (D_{max}):** Larger aggregates can lead to variations in strength, although UHPC typically uses smaller aggregate particles to achieve its dense matrix [67].
- **Packing density:** The packing density, which measures how efficiently particles pack together, is a key predictor of the mechanical performance of UHPC. Yu et al. have shown that optimizing particle packing and using hybrid steel fibers enables the production of UHPC with relatively low binder content [61]. Accordingly, a mixture containing 1.5 and 0.5% vol. of long and short straight steel fibers, respectively, was reported to exhibit a 28-day compressive strength of 142 MPa.
- **Fiber diameter (D_f):** Although fiber diameter may not significantly impact compressive strength, it plays a role in other properties such as toughness and flexural strength [61].

- **Fiber length (Lf):** Like fiber diameter, fiber length has more influence on properties like flexural strength rather than compressive strength [62].
- **Aggregate strength:** The strength of aggregates, though not always a primary factor in compressive strength, can influence the overall durability and integrity of UHPC. Aggregates with higher strength can contribute to a higher overall compressive strength of UHPC [67].

2.4.1 Number of Features for UHPC Compressive Strength Prediction

From Table 2-1, the number of features utilized in UHPC compressive strength prediction differs from one journal to another. Some publications used only seven (7) input variables for prediction while others used up to twenty-four (24) features for building their predictive models. The dataset size should be large enough to be commensurate with the number of features for building predictive models. This is hardly the case in most real-world datasets. UHPC datasets compiled and used for compressive predictions in earlier studies are not large enough but still achieved good performance even with many features. A large number of input variables with a small dataset causes high dimensionality of the datasets with some features not contributing to performance of the ML models. To solve this issue, feature selection techniques can be used to eliminate redundant features and can also help to improve the models' outcomes.

2.5 Machine Learning Algorithms

ML algorithms have gained increasing prominence in civil engineering due to their ability to model complex, non-linear relationships between input variables and material properties. In the context of UHPC, ML techniques have been widely applied to predict compressive strength, a critical property for structural applications. Various algorithms have been utilized to capture the intricate interplay between UHPC mix components and their resulting mechanical performance.

2.5.1 Artificial Neural Network

Artificial Neural Networks (ANNs) have become one of the most popular ML models in predicting concrete properties, including UHPC compressive strength [23], [68]. Due to their multi-layered architecture, ANNs use a single hidden layer to capture highly complex and non-linear relationships in the data. They have been particularly useful in cases where UHPC behavior is influenced by numerous variables, including mix design components, curing conditions, and environmental factors.

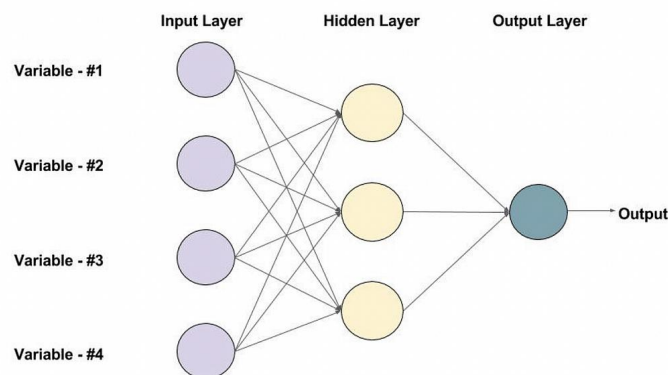


Figure 2-1: Diagrammatic Representation of ANN (Medium.com)

2.5.2 Deep Neural Network

Deep Neural Networks (DNNs) are an extension of traditional neural networks that employ additional hidden layers to enhance their capacity for learning from data. DNNs have been mentioned in the context of modeling UHPC compressive strength [33]. Their ability to process vast amounts of data and identify deep patterns makes them ideal for high-dimensional datasets, such as those used in UHPC research.

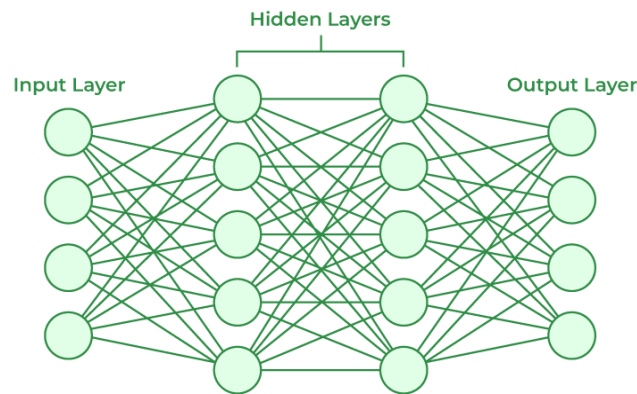


Figure 2-2: Diagrammatic Representation of DNN (GeeksforGeeks.org)

2.5.3 Support Vector Regressor

Support Vector Regressor (SVR) is a robust algorithm known for its ability to model complex relationships, particularly in cases with high-dimensional data. In the studies reviewed, SVR has been applied to predict the compressive strength of UHPC and other composite materials, demonstrating strong predictive performance when compared to traditional regression techniques. By using kernel functions, SVM is capable of handling non-linear patterns, which is especially important in UHPC applications where material behavior is governed by multiple interacting factors.

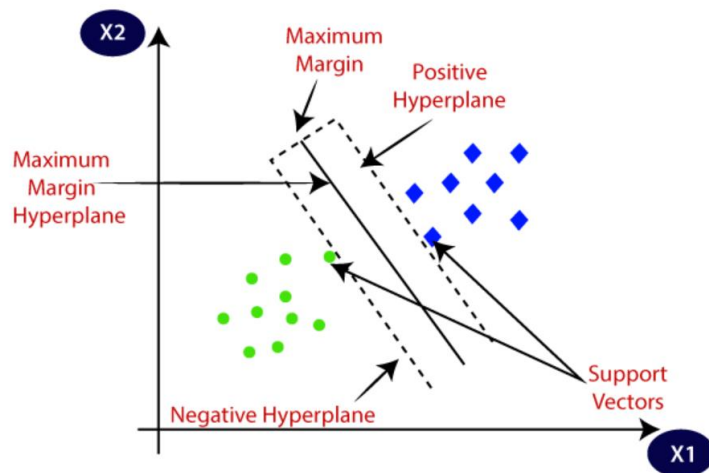


Figure 2-3: Simplified Representation of SVM (Medium.com)

2.5.4 Decision Tree Regressor

Decision Tree Regressor (DTR) is a foundational algorithm in machine learning, often used for regression tasks like UHPC compressive strength prediction. While DTs can capture non-linear relationships, they are prone to overfitting unless properly tuned or combined with ensemble methods.

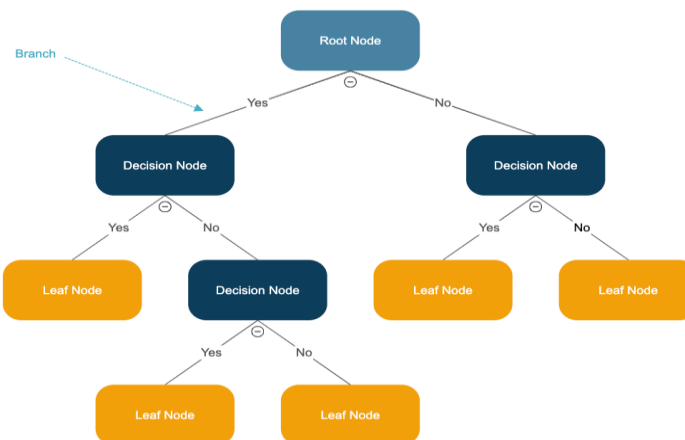


Figure 2-4: Typical Structure of DT (Smartdraw.com)

2.5.5 Random Forest Regressor

Random Forest Regressor (RFR) is an ensemble learning method that constructs multiple decision trees and averages their predictions to improve accuracy and prevent overfitting. RFR has been widely applied to UHPC compressive strength prediction [28], [6], demonstrating strong performance due to its ability to model non-linear relationships and handle interactions between input components.

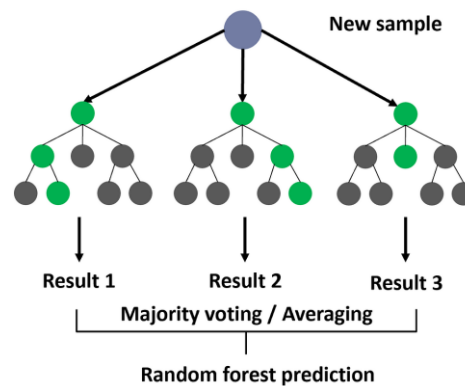


Figure 2-5: Simplified Representation of RF (Medium.com)

2.5.6 Extra Trees Regressor

Extra trees regressor (ETR) is another ensemble learning method similar to RF. However, it introduces more randomness in the tree-building process, which can reduce variance and improve model generalization. ETR has been applied to various concrete prediction tasks, including UHPC compressive strength modeling [6], [33].

2.5.7 Gradient Boosting Regressor

Gradient Boosting Regressor (GBR) is a powerful ensemble method that builds models sequentially, where each new model corrects the errors of its predecessor. This iterative process has proven to be highly effective for predicting the compressive strength of UHPC [6], [31]. GBR's ability to handle linear and non-linear interactions makes it a preferred choice for many researchers working with complex datasets. There are several variants of GBR that have shown better performance from previous studies.

- **Light Gradient Boosting Machine**

Light gradient boosting machine (LightGBM) is a variant of GBR designed to be more efficient, particularly for large datasets. It uses histogram-based algorithms to speed up the training process while maintaining accuracy. LightGBM has been successfully used in predicting UHPC properties, offering similar predictive power as GBR but with improved computational efficiency [35], [36].

- **Extreme Gradient Boost Regressor**

Extreme Gradient Boost Regressor (XGBR), a more advanced version of GBM, is designed for speed and performance. It has been applied to UHPC compressive strength prediction, demonstrating superior accuracy by leveraging regularization techniques to prevent overfitting [12], [12], [30], [33], [34], [37]. Its scalability and efficiency make it ideal for large and complex UHPC datasets.

- **Categorical Boosting Regressor**

Categorical Boosting Regressor (CBR) is another gradient-boosting algorithm that handles categorical data efficiently, reducing the need for extensive preprocessing. It has been successfully applied to the prediction of concrete compressive strength, including UHPC, yielding competitive performance compared to other boosting algorithms.

- **Adaptive Boosting Regressor**

Adaptive Boosting Regressor (ABR) is another ensemble method that boosts the performance of weaker models by combining them iteratively. It has been applied to concrete compressive strength prediction, including UHPC, showing solid performance in cases where other models struggled.

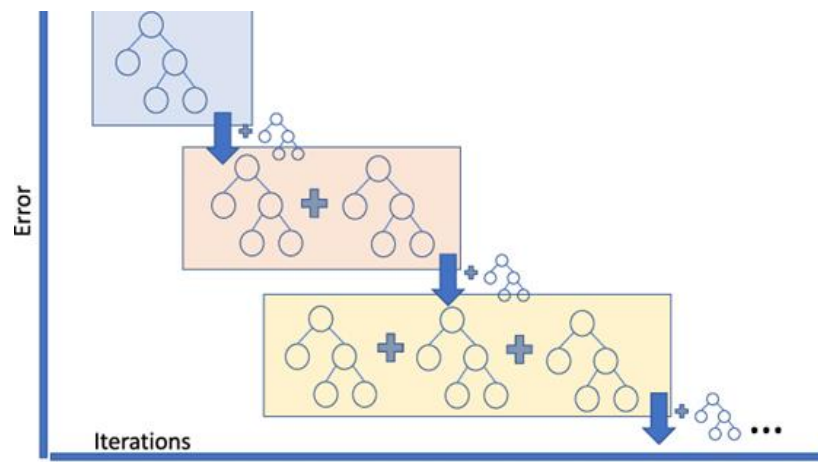


Figure 2-6: Graphical Representation of GB (Medium.com)

2.5.8 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a simple, yet effective, non-parametric algorithm that predicts the compressive strength of UHPC by averaging the values of its nearest neighbors in the feature space. Despite its simplicity, KNN has shown good performance in cases where the relationship between input parameters and compressive strength is not strictly linear.

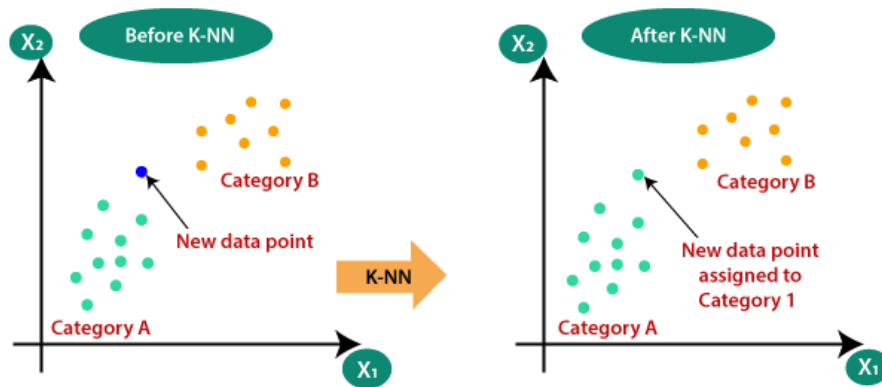


Figure 2-7: KNN Visual Representation (Javapoint.com)

2.5.9 Other Ensemble Methods

Ensemble Methods combine multiple algorithms to improve model accuracy and robustness. Techniques like stacking and voting, where multiple models' predictions are combined, have shown excellent performance in UHPC compressive strength prediction. Most published journals in this area (UHPC compressive strength prediction) used default ensemble algorithms such as RF, XGB, etc., but few publications utilized custom ensemble techniques by combining multiple algorithms to achieve better performance.

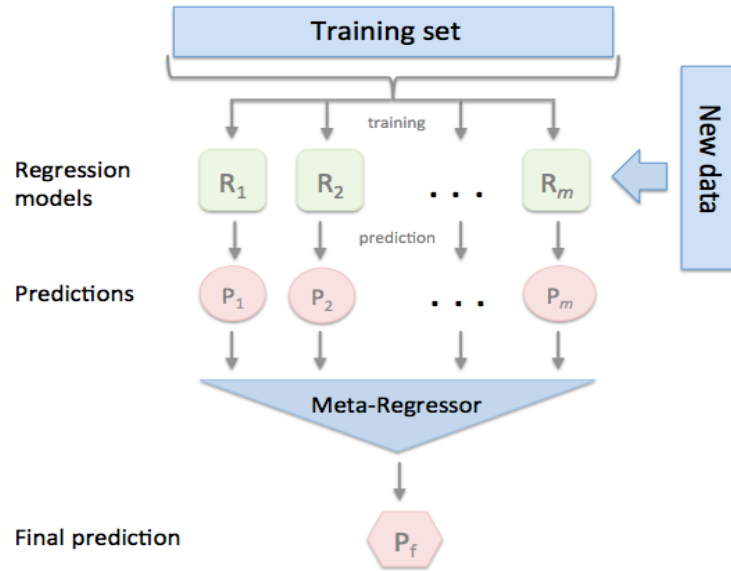


Figure 2-8: Stacking Ensemble Method

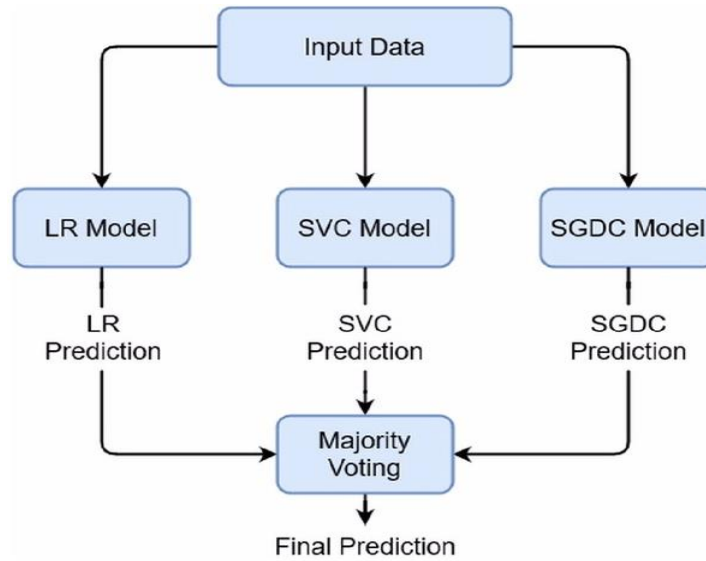


Figure 2-9: Voting Ensemble Method

2.6 Performance Measure

The effectiveness of machine learning models in predicting Ultra-High-Performance Concrete (UHPC) compressive strength is evaluated using a variety of performance

metrics. These metrics assess how well the model's predicted values align with the actual experimental results, enabling researchers to quantify model accuracy, bias, and overall fit. The performance measures used in the reviewed studies include both error-based and variance-based metrics, providing a comprehensive evaluation of model performance.

2.6.1 Coefficient of Determination

R-squared (R^2), also known as the coefficient of determination, measures the proportion of variance in the dependent variable (compressive strength) that can be predicted from the independent variables (UHPC mix components). It ranges from 0 to 1, with values closer to 1 indicating a better model fit.

$$R^2 = 1 - \frac{\sum_{k=0}^n (y_k - \hat{y}_k)^2}{\sum_{k=0}^n (y_k - \bar{y}_k)^2} \quad (2.1)$$

Where n is the number of observations, y_k is the actual value, $\hat{y}_k$ is the predicted value.

2.6.2 Adjusted R^2

Adjusted R-squared modifies the R^2 metric to account for the number of independent variables in the model, penalizing models that include irrelevant predictors. It is useful when comparing models with different numbers of features, as it discourages overfitting.

$$Adj. R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \quad (2.2)$$

Where n is the number of observations, p is the number of predictors.

2.6.3 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is a widely used measure of the differences between predicted and actual values. It provides an absolute measure of fit in the units of the dependent variable (e.g., MPa for compressive strength), with lower values indicating better model performance. RMSE is particularly sensitive to large errors, making it useful for detecting outliers.

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=0}^n (y_k - \hat{y}_k)^2} \quad (2.3)$$

Where n is the number of observations, y_k is the actual value, $\hat{y}_k$ is the predicted value.

2.6.4 Mean Absolute Error

Mean Absolute Error (MAE) calculates the average absolute difference between predicted and actual values. Like RMSE, a lower MAE indicates better prediction accuracy, but MAE is less sensitive to outliers.

$$MAE = \frac{1}{n} \sum_{k=0}^n |y_k - \hat{y}_k| \quad (2.4)$$

Where n is the number of observations, y_k is the actual value, $\hat{y}_k$ is the predicted value.

2.6.5 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) expresses the average absolute percentage difference between predicted and actual values, making it scale-independent. This metric is useful when the dataset contains a wide range of values, as it provides an intuitive measure of prediction accuracy in percentage terms.

$$MAE = \frac{1}{n} \sum_{k=0}^n \left| \frac{y_k - \hat{y}_k}{y_k} \right| * 100 \quad (2.5)$$

Where n is the number of observations, y_k is the actual value, $\hat{y}_k$ is the predicted value.

2.6.6 Most Commonly Used Evaluation Metrics

Aside from the five model evaluation metrics listed above, several other metrics have been reported in previous studies on UHPC compressive strength. Examples include weighted mean absolute percentage error (WMAPE), variance account factor (VAF), performance Index (PI), maximum error (ME), and others. The most widely used performance metrics in all the reviewed literature on UHPC compressive strength prediction are MAE, RMSE and R^2 .

2.7 Hyperparameter Optimization

Hyperparameter optimization (HPO) plays a critical role in improving the performance of machine learning models by fine-tuning parameters that govern the learning process. In predicting the compressive strength of UHPC, several HPO techniques have been applied

to find the optimal settings for models such as RFR, XGR, and others. The choice of optimization method often depends on factors such as problem complexity, data characteristics, and computational resources.

2.7.1 Trial and Error

The trial-and-error method is a straightforward approach for hyperparameter tuning that involves manually adjusting parameters based on model performance observations. This technique, though labor-intensive and time-consuming, allows researchers to test parameter combinations iteratively until satisfactory results are achieved. In the reviewed studies, the trial-and-error method was frequently used for simpler models and smaller datasets to determine optimal hyperparameters for algorithms like RFR, ETR, GBR, and VR. While this method is less efficient than automated approaches, it can still be effective for models with few hyperparameters.

2.7.2 Grid Search Optimization

Grid search is a classic hyperparameter optimization technique that systematically explores a predefined grid of hyperparameter values. Grid search guarantees that the optimal set of hyperparameters is identified by exhaustively looping through each possible parameter combination. However, it is computationally expensive, especially in high-dimensional spaces with multiple parameters. Many studies utilized grid search in combination with k-fold cross-validation to enhance the model's robustness while tuning hyperparameters for models in UHPC compressive strength prediction.

2.7.3 Random Search Optimization

Random search is an alternative to grid search that randomly samples parameter combinations from a specified distribution. This approach is often more efficient than grid search for high-dimensional parameter spaces. It is faster than grid search as it reduces the number of evaluations by focusing on random subsets. In UHPC applications, random search has been employed as a more practical solution for complex models where computational resources are limited.

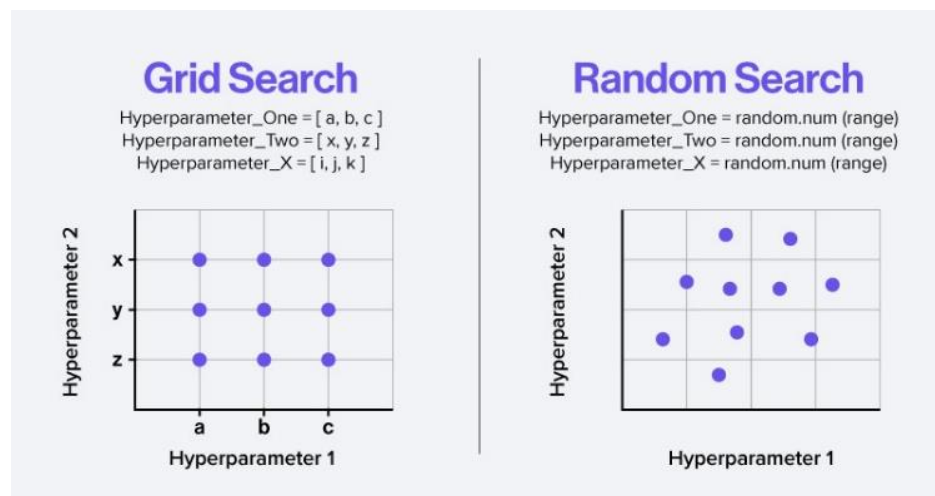


Figure 2-10: Visual representation of Grid and Random Search (Sabrepc.com)

2.7.4 Bayesian Optimization

Bayesian optimization (BO) is a probabilistic model-based approach that seeks to minimize the objective function by estimating the best parameter combination iteratively. BO is more computationally efficient than grid or random search because it strategically explores the parameter space, focusing on regions likely to contain optimal values. The reviewed studies highlighted Bayesian optimization's effectiveness in fine-tuning hyperparameters for

complex models like RFR. By continuously updating its search strategy based on previous evaluations, BO can achieve high accuracy with fewer iterations, making it suitable for UHPC research with limited computational resources.

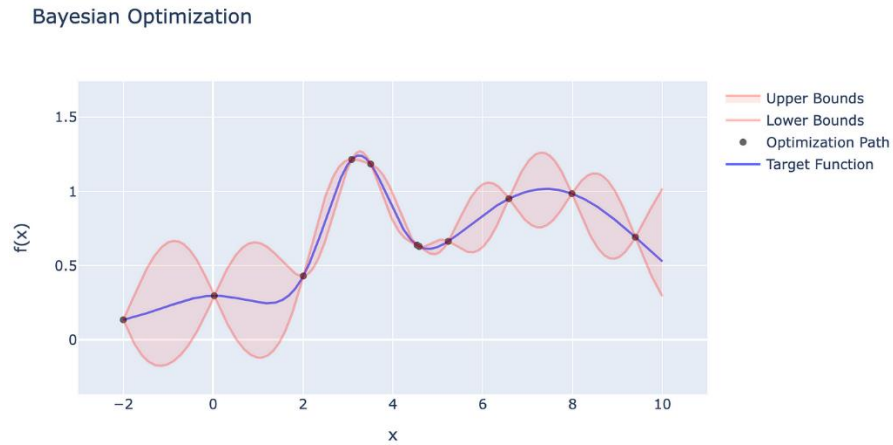


Figure 2-11: Visual representation of Bayesian Optimization (Medium.com)

2.8 Feature Selection and Engineering

Feature selection is a critical process in machine learning that helps enhance model accuracy, reduce dimensionality, and improve interpretability by identifying the most relevant input variables. For predicting the compressive strength of UHPC, various feature selection techniques have been employed to extract meaningful features that contribute to model precision. Each method is chosen based on factors such as data structure, model type, and computational resources.

2.8.1 Sequential Feature Selection

Sequential Feature Selection (SFS) is a wrapper method that sequentially adds features to the model, evaluating the contribution of each feature based on an objective function. This process continues until adding further features does not significantly improve the model's performance. SFS has been frequently applied in combination with models such as Artificial Neural Networks (ANN) and Back-Propagation Neural Networks (BPNN) to enhance the accuracy of UHPC compressive strength predictions while reducing computation time [69]. By iteratively selecting features, SFS minimizes redundancy, helping the model focus on the most impactful variables.

2.8.2 Automated Feature Engineering and Selection

Automated Feature Engineering and Selection (Autofeat) is an advanced feature engineering technique that generates a comprehensive pool of potential features, including linear, interaction, and non-linear transformations of the original variables. By automatically generating new features and selecting the most informative ones, Autofeat enhances the model's performance, particularly in complex datasets where non-linear relationships are prevalent. In UHPC research, Autofeat has been used to create features that may not be immediately apparent but have a significant impact on compressive strength predictions [32], [35], [39].

2.8.3 Effects of Feature Selection on the Input Variables

The impact of Feature Selection techniques on UHPC compressive strength dataset has not been well studied, especially as it relates to dimensionality reductions within the dataset. Feature selection techniques like correlation rank, recursive feature elimination (RFS), and LASSO regression can help to remove redundant features and contribute towards the robustness of the ML models.

2.9 Explainable AI Techniques

Explainable AI (XAI) techniques play a crucial role in enhancing the interpretability of machine learning models, allowing researchers and practitioners to understand how different input variables influence predictions. In the context of UHPC compressive strength prediction, XAI helps to address the “black-box” nature of complex algorithms by providing insights into the relationships between UHPC composition and mechanical properties. The following XAI techniques have been applied to UHPC strength prediction models to improve transparency and facilitate a deeper understanding of model behavior.

2.9.1 Shapley Additive Explanations

Shapley Additive Explanations (SHAP) is one of the most widely used XAI techniques, especially for highly complex models. SHAP provides interpretability by calculating the contribution of each feature to the model's predictions, offering both global and local insights into the model's behavior. In UHPC research, SHAP has been used to analyze the relationship between concrete composition and predicted compressive strength. By

quantifying how each feature affects the prediction outcome, SHAP allows for a more transparent understanding of how material properties like cement content, water-to-binder ratio, and curing conditions contribute to the UHPC strength.

2.9.2 Neural Interpretation Diagram

Neural Interpretation Diagram (NID) is a visualization tool used to assess the influence of each input variable on the model's output, particularly in neural network models. NID provides a graphical representation of the relationships between input features and predictions, allowing researchers to visually inspect the significance of each variable. This technique has been used to validate the selected parameters in models predicting UHPC compressive strength, adding a layer of interpretability to the machine learning process.

2.9.3 Partial Dependence Plot (PDP)

Partial Dependence Plot (PDP) is an XAI tool that visualizes the marginal effect of one or two input features on the target variable while keeping all other features constant. This method helps interpret how changes in a particular feature impact the predicted compressive strength of UHPC. PDPs are valuable for understanding non-linear relationships between input features and output variables, particularly in complex datasets where multiple factors interact.

2.10 Research Gap

The lack of comprehensive datasets for predicting UHPC compressive strength, combined with the high dimensionality of existing datasets, presents challenges in developing robust machine learning (ML) models for UHPC applications. This study aims to address these issues by critically exploring feature selection techniques to reduce the number of input variables used in UHPC compressive strength prediction. By tackling the high dimensionality of current datasets, this research seeks to enhance model robustness and facilitate the practical implementation of ML models in the construction industry for UHPC applications.

CHAPTER 3

INTEGRATION OF FEATURE OPTIMIZATION METHODS AND MACHINE LEARNING MODELS FOR ULTRA-HIGH-PERFORMANCE CONCRETE (UHPC) COMPRESSIVE STRENGTH PREDICTION

3.1 Introduction

Ultra-high-performance concrete (UHPC) is an advanced concrete used to construct critical infrastructures like bridges, high-rise buildings, and structural repairs [3], [68]. Although UHPC is very expensive and requires high expertise to manufacture [1], [34], UHPC is preferred to conventional concrete for some construction works due to its superior mechanical properties, exceptional durability, and resistance to extreme environmental conditions [10]. UHPC differs slightly from normal concrete in its composition. It requires high binder content, low water-to-binder ratio, high supplementary cementitious materials (SCMs), mineral admixtures, and steel fibers [70]. The predominant SCM in UHPC is silica fume which contributes to the optimum packing density of UHPC mix [71]. Other SCMs mostly incorporated into the UHPC mix include fly ash (FA), quartz powder (QA), limestone powder (LP), granulated blast furnace slag (SL), and nano-silica (NS) [71]. The impact of each constituent of UHPC on its strength and other properties has been well investigated to optimize UHPC mix design for specific purposes [1], [72].

The optimization of UHPC, particularly its compressive strength, is critical for maximizing its potential [26]. Compressive strength is a key performance indicator that determines the material's capacity to withstand axial loads, which is vital for structural integrity and safety [73]. The minimum compressive strength of UHPC is set at 120 MPa per ASTM C1856 [34]. Several efforts have been made to optimize the UHPC mix design to achieve the desired compressive strength through experimental works [27]. While UHPC mix optimization has majorly been through extensive laboratory testing and empirical methods, the process is very expensive, time-consuming, and labor-intensive [6]. In addition, these processes do not fully capture the complexity of the interaction between UHPC's various components and their final effects on UHPC properties.

With advancements in computational technologies, machine learning (ML) has emerged as a powerful tool for modeling and predicting the properties of complex materials [39]. ML algorithms can analyze large datasets to identify patterns and correlations that are not apparent through traditional methods. By applying ML techniques to UHPC mix design, researchers can develop predictive models that accurately estimate compressive strength based on various input parameters. This approach not only accelerates the optimization process but also enhances the precision of the predictions, leading to more efficient and cost-effective mix designs [35].

Recent studies have explored the use of ML to predict the compressive strength of UHPC based on its mixture design parameters. Table 3-1 presents the summary of some of the earlier ML attempts for UHPC compressive strength predictions. Table 3-1 shows the input parameters, ML techniques, number of observations, and performance metrics. Although

three major evaluation metrics (R^2 , MAE, and RMSE) have been popularly adopted in most studies, only R^2 has been included in Table 3-1.

Table 3-1: Predictive Models in Previous Studies of UHPC Compressive Strength

Year	ML Algorithm(s)	No	Input Parameters*	Data Size	R^2	Ref
2017	ANN	11	C, S, CA, W, FA, SF, SP, W/B, TA/B, S/TA, Dmin–Dmax	78	0.998	[24]
2018	ANN	5	W/B, Df, Lf, AsF, Fi	162	0.990	[25]
2020	ANN	8	C, Fi, FA, S, SF, QP, W, Ad	110	0.801	[26]
2020	RFR, GBR, ETR	15	C, SF, SL, FA, QP, LP, NS, W, Fi, SP, S, CA, T, RH, Age	810	0.926 - 0.962	[6]
2021	GEP-ANN, ANN	5	W/B, C, SF, LP, S	80	0.831 - 0.970	[27]
2022	RFR	17	C, SF, FA, SL, GP, RHA, FC3R, MK, LP, W, SP, QP, TA, Dmax, W/B, W/P, VPD	931	0.904	[28]
2022	ANN	12	C, SF, GP, FC3R, LP, QP, SP, W, S, W/B, W/P, VPD, Age	265	0.960 - 0.983	[29]
2022	XGBR	17	C, SF, FA, SL, GP, RHA, FC3R, MK, LP, W, SP, QP, TA, Dmax, W/B, W/P, VPD	931	0.892	[30]
2022	XGBR, ABR, BR	10	C, FA, SL, SF, S, W, SP, Fi, AsF, Age	372	0.78 - 0.90	[12]
2022	DTR, BR, GBR	21	C, FA, SL, NS, QP, LP, S, CA, SP, Fi, Fp, W, CT, CS, CSC, Dmax, Lf, Df, Lp, Dp	626	0.94 - 0.95	[31]
2022	AutoML	24	C, FA, SF, SL, NS, QP, LP, S, CA, SP, Fi, Fp, W, CT, CS, CSC, Dmax, Lf, Df, Lp, Dp, Age, Si, Li	626	0.87 - 0.99	[32]
2023	DNN, XGBR, ETR, VR, GBR	15	C, SF, SL, FA, QP, LP, NS, W, Fi, SP, S, CA, T, RH, Age	810	0.98 - 0.99	[33]

Year	ML Algorithm(s)	No	Input Parameters*	Data Size	R ²	Ref
2023	XGBR	12	C, SF, QP, GP, TA, W, SP, Dmax, W/B, W/P, VPD	309	0.95	[34]
2023	AutoML	16	C, CT, FA, SL, SF, MK, NS, LP, QP, S, Dmax, W, SP, Fi, Si, AsF	379	0.95	[35]
2023	XGBR, ABR, CBR, RFR, LGBMR	10	CS, SF/C, S/C, Age, SP/C, W/B, W/C, Fi, Lf/Df, L/B	400	0.883-0.945	[36]
2023	XGBR, ABR, BR, SR	15	C, SF, SL, FA, QP, LP, NS, W, Fi, SP, S, CA, T, RH, Age	810	0.951 - 0.958	[37]
2023	GEP	15	C, SF, SL, FA, QP, LP, NS, W, Fi, SP, S, CA, T, RH, Age	810	0.977 - 0.992	[38]
2023	ANHB, ANCO, ANGN	8	C, SF/C, FA/C, S/C, QP/C, W/C, Ad/C, Fi/C	132	0.987 - 0.996	[5]
2024	LassoR, AutofeatR	15	C, SF, LP, QP, FA, NS, SL, W, S, CA, Fi, SP, RH, T, Age, B	810	0.950 - 0.962	[39]

*Note: Cement (C), Fine aggregate (S), Coarse aggregate (CA), Water (W), Fly ash (FA), Silica Fume (SF), Superplasticizer (SP), Total aggregate (TA), Binder (B), Minimum aggregate size (Dmin), Maximum aggregate size (Dmax), Steel fiber (Fi), Quartz powder (QP), Nano silica (NS), Slag (SL), Admixture (Ad), Aspect ratio of fiber (AsF), Curing time (Age), Cement type (CT), Cement strength (CS), Cement strength class (CSC), Temperature (T), Relative humidity (RH), Limestone powder (LP), Metakolin (MK), Glass powder (GP), Length of steel fiber (Lf), Diameter of steel fiber (Df), Polystyrene fibre (Fp), Length of polystyrene fiber (Lp), Diameter of polystyrene fiber (Df), Length of sample (Li), Virtual parking density (VPD), Fluid catalytic residue (FC3R).

While ML application in UHPC compressive strength prediction holds great promise, several challenges are hindering its adoption. One significant challenge is the availability of large-quality datasets required to train ML algorithms effectively and build accurate models for UHPC strength prediction. Furthermore, there are lots of input parameters used in the predictive model which makes the available datasets highly dimensional and difficult to implement. Some studies also utilize complex ML techniques like Deep Learning, which are difficult to interpret and understand the underlying factors behind the predictions. This research aims to address these limitations by using feature selection techniques to develop explainable ML models that are robust enough to predict UHPC compressive strength with

high accuracy and easy enough to apply by end users. Feature selection approaches will be applied to reduce the high dimensionality of the datasets and optimize the process significantly.

3.2 Collected Data Explanation

3.2.1 Description of Dataset

The dataset used for this study was obtained from the work of Merani et al, 2020 [74]. The dataset contains 810 observations with 16 features gathered from 28 research journals [6]. Table 3-2 shows the description of the dataset and the units of each parameter. The dataset was checked for quality and tidiness issues.

Some important assumptions were made during the gathering of the dataset. The assumptions are listed below.

1. Only the cement content is considered. Cement types and cement strength classes were not included in the dataset.
2. Only the contents of supplementary cementitious materials (SCMs) [SL, SF, FA, NS, LP, QP] were included. The SCMs' types reported in the manuscripts were not considered.
3. Only superplasticizer is included as mineral admixture. Other forms of admixtures such as retarder and water-reducer were not added.

4. Only steel fiber is added. Other forms of fibers such as polystyrene and polypropylene are not considered. Also, physical properties of steel fiber such as length and diameter were excluded from the dataset.
5. Three curing conditions (T and RH and Age) were considered.
6. The mixture constituents' absolute mass in kg per unit volume in m^3 is recorded in the dataset.
7. The particle size distribution, density, and other physical properties of UHPC constituents were excluded from the dataset.

Table 3-2: Description of features in the dataset with their units

Feature	Description	Unit
C	Cement content	kg/m^3
SL	Granulated blast furnace slag content	kg/m^3
SF	Silica fume content	kg/m^3
LP	Limestone powder content	kg/m^3
QP	Quartz powder content	kg/m^3
FA	Fly ash content	kg/m^3
NS	Nano silica content	kg/m^3
W	Water content	kg/m^3
S	Sand content	kg/m^3
CA	Coarse aggregate content	kg/m^3
Fi	Steel fiber content	kg/m^3
SP	Superplasticizer content	kg/m^3
T	Curing temperature of UHPC	$^{\circ}\text{C}$
RH	Curing relative humidity of UHPC	%
Age	Curing time	days
fcu	Compressive strength of UHPC	MPa

3.2.2 Statistical Analysis of Dataset

Table 3-3 presents the summary statistics of the features in the dataset. The summary statistics include the mean, first quarter (Q25), second quarter (median), third quarter (Q75), standard deviation, minimum, and maximum values of each attribute. Table 3-4 provides the Pearson pairwise correlation coefficient of each input variable to the output variable.

Table 3-3: Summary statistics of each feature in the dataset

Feature	Mean	Q25	Median	Q75	Standard Deviation	Minimum	Maximum
C	737.91	620.20	770.5	850.00	173.46	270.00	1251.20
SL	25.19	0.00	0.0	0.00	74.36	0.00	375.00
SF	136.98	43.70	144.0	219.00	104.14	0.00	433.70
LP	41.93	0.00	0.0	0.00	133.13	0.00	1058.20
QP	33.27	0.00	0.0	0.00	79.67	0.00	397.00
FA	26.26	0.00	0.0	0.00	67.46	0.00	356.00
NS	3.64	0.00	0.0	4.00	7.77	0.00	47.50
W	179.89	163.00	177.0	192.50	25.57	90.00	272.60
S	995.33	786.40	1021.0	1231.00	283.27	0.00	1502.80
CA	154.78	0.00	0.0	0.00	357.57	0.00	1195.00
Fi	56.04	0.00	0.0	156.00	75.23	0.00	234.00
SP	30.03	18.00	30.2	44.20	13.99	1.10	57.00
T	97.98	100.00	100.0	100.00	7.81	50.00	100.00
RH	23.92	21.00	21.0	23.00	16.21	20.00	210.00
Age	37.09	7.00	28.0	28.00	53.11	1.00	365.00
fcu	123.13	96.00	122.3	154.27	40.24	28.51	220.5

Table 3-4: Correlation coefficient between each input feature and the target (fcu)

Input feature	Correlation coefficient	Absolute Rank
SF	+0.52113	1
Fi	+0.44034	2
Age	+0.41536	3
SP	+0.39574	4

Input feature	Correlation coefficient	Absolute Rank
C	+0.33735	5
S	-0.27486	6
QP	+0.23796	7
NS	-0.20579	8
T	+0.15084	9
RH	-0.13505	10
CA	-0.13078	11
LP	-0.12063	12
W	+0.08681	13
FA	-0.05817	14
SL	+0.05549	15

3.2.3 Exploratory Data Analysis (EDA)

Preliminary data visualizations were created to assess the dataset's quality further and to visualize the distribution of each variable. The correlation heatmap (Figure 3-1) was plotted to visualize each variable's Pearson pairwise correlation coefficient with other variables. The correlation heatmap provides insights into possible collinearity issues in the dataset. The distribution of each feature was assessed with Histograms (Figure 3-2). Scatter plots (Figure 3-3) were used to determine the linear relationship between each input feature and the target variable.

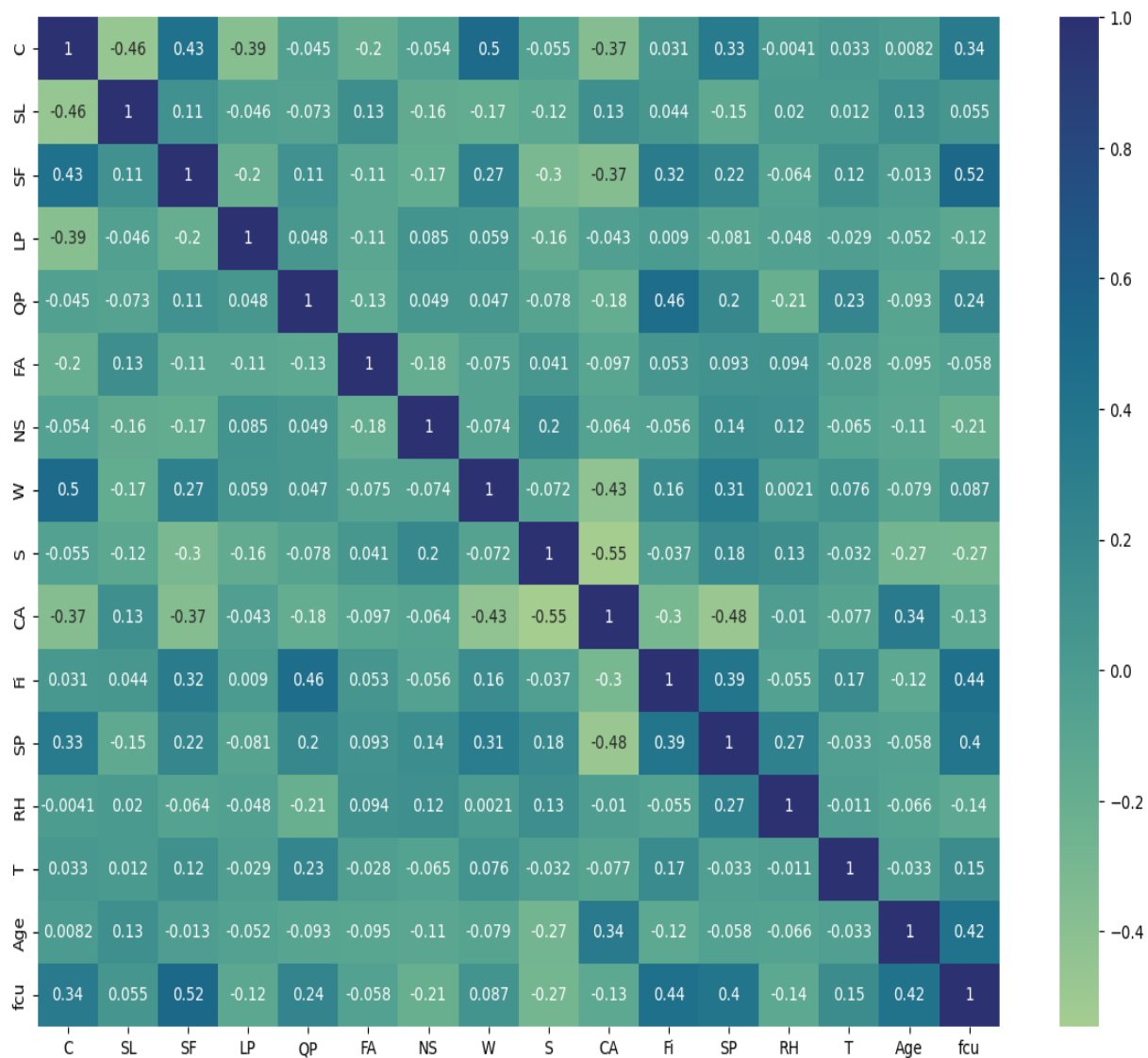


Figure 3-1: Feature Correlation Heatmap

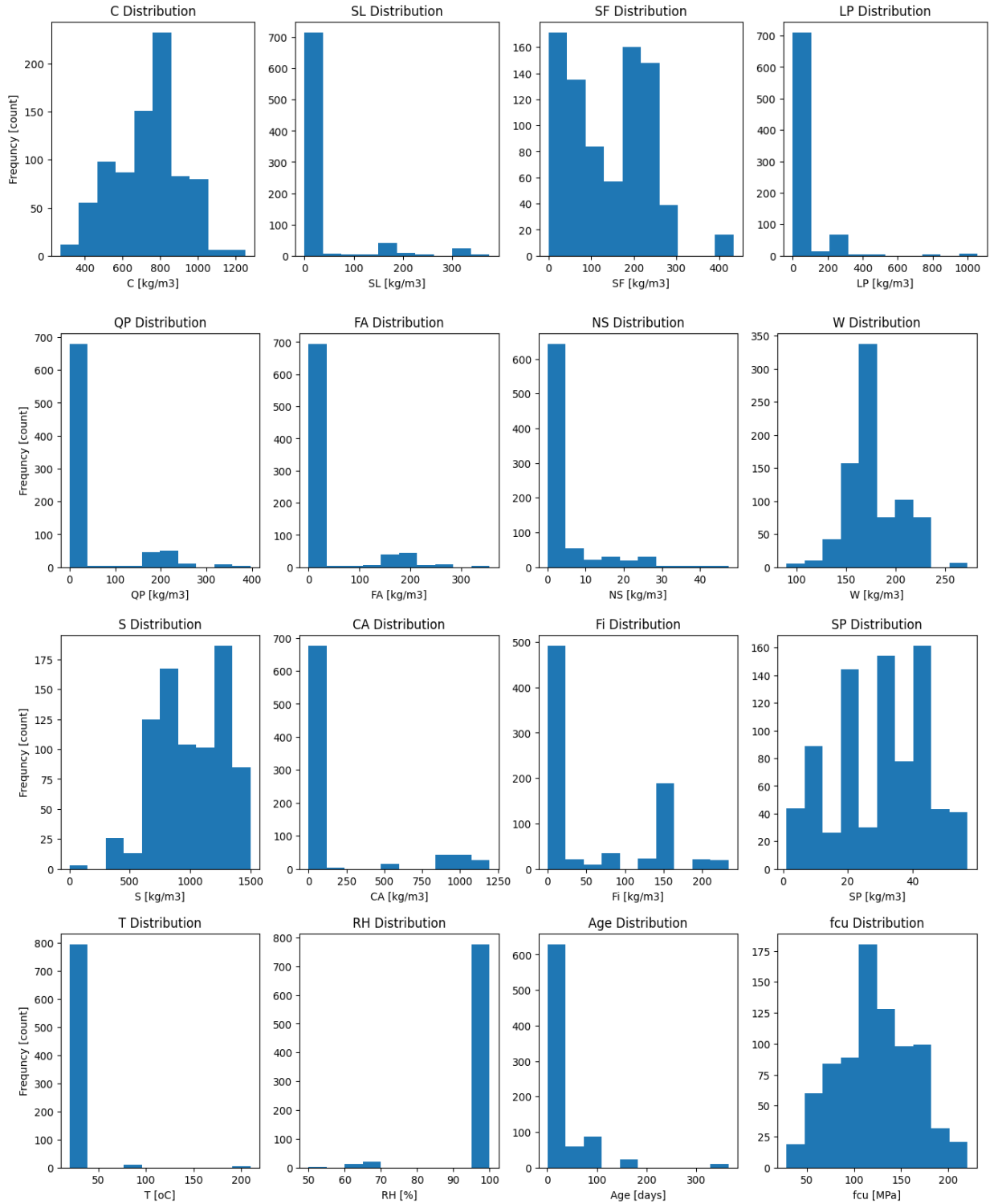


Figure 3-2: Feature Distribution

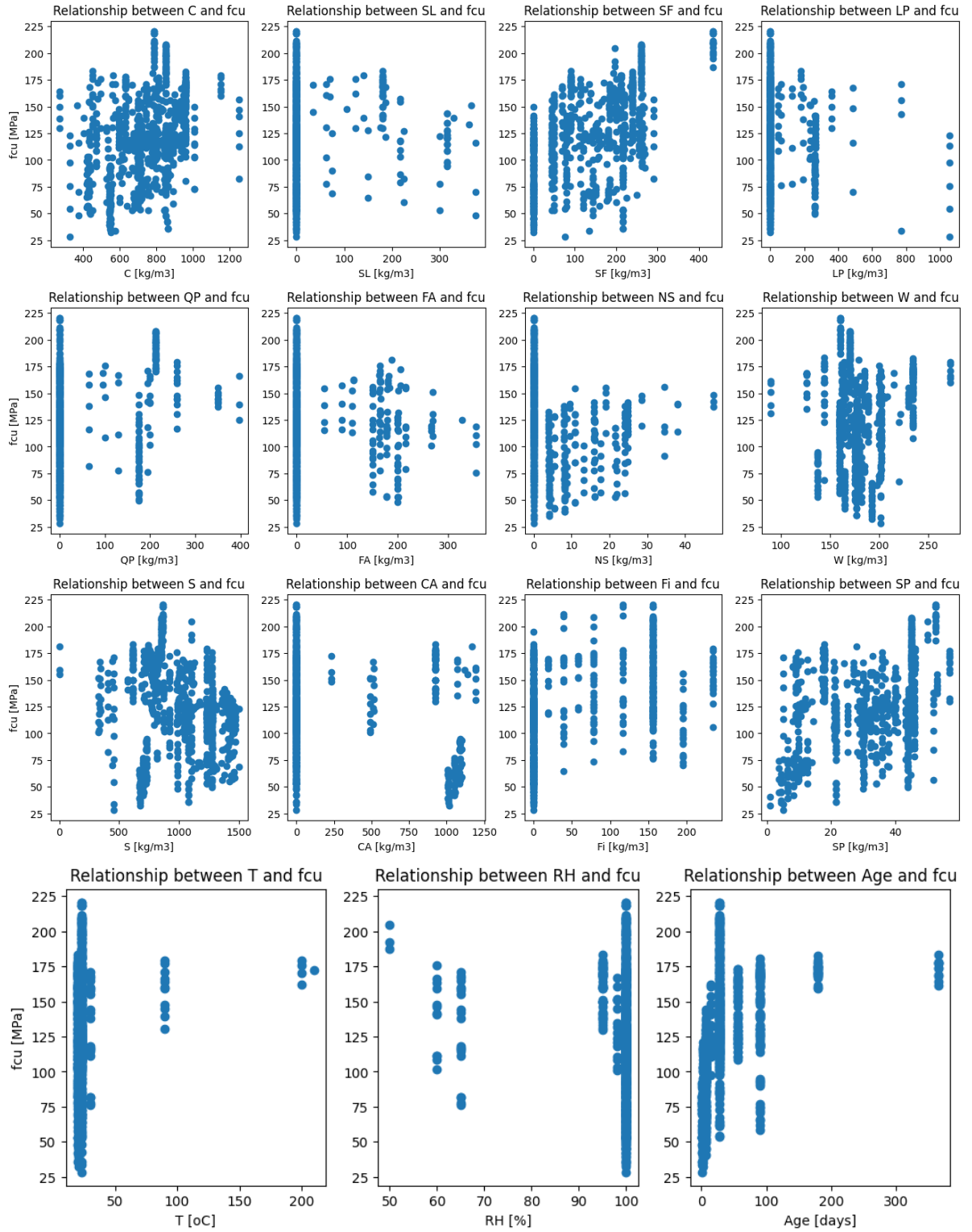


Figure 3-3: Scatter Plot of Each Feature with the Target

3.3 Methodological Overview

3.3.1 K-Nearest Neighbors

K-Nearest Neighbors (KNN) Regressor is a non-parametric method used for regression tasks [75]. It predicts the target value for a given input by averaging the values of its k -nearest neighbors in the feature space. The Euclidean distance is commonly used to determine the nearest neighbors. The predicted value $\hat{y}$ for a given input value x is calculated as:

$$\hat{y} = \frac{1}{k} \sum_{i=1}^k y_i \quad (3.1)$$

Where y_i are the target values of k -nearest neighbors.

3.3.2 Decision Tree Regressor

Decision Tree Regressor is a non-linear model that uses a tree-like structure to make predictions. It recursively splits the data into subsets based on the feature that provides the highest information gain, or the lowest mean squared error. The process continues until a stopping criterion is met, such as a maximum tree depth or a minimum number of samples per leaf. The prediction for a given input is the mean value of the target variable in the leaf node where the input falls.

3.3.3 Random Forest Regressor

Random Forest Regressor is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and control over-fitting. Each tree in the forest is trained on a bootstrap sample of the dataset, and features are randomly selected at each split. The final prediction is the average of the predictions from all individual trees. The Random Forest algorithm reduces variance and enhances generalization compared to a single decision tree.

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B \hat{y}_b(x) \quad (3.2)$$

where $\hat{y}_b(x)$ is the prediction of the b-tree, and B is the number of trees.

3.3.4 Extreme Gradient Boost Regressor

Extreme Gradient Boosting (XGBoost) is a scalable and efficient machine learning algorithm that belongs to the family of gradient boosting techniques. It is widely used for supervised learning tasks, including classification, regression, and ranking problems. XGBoost builds an ensemble of weak learners (typically decision trees) sequentially, with each new learner correcting the errors made by the previous ones. XGBoost enhances this process by employing a regularization term to prevent over-fitting and using parallel processing to speed up computation. The prediction of the ensemble model is given by:

$$\hat{y} = \sum_{m=1}^M \alpha_m f_m(x) \quad (3.3)$$

where α_m is the learning rate, $f_m(x)$ the prediction from the m-tree, and M is the number of boosting rounds.

3.3.5 Support Vector Regressor

The Support Vector Machine (SVM) is a popular statistical-based supervised machine learning algorithm widely applied for both regression and classification problems. SVM aims to enhance class separation and minimize prediction errors. It's renowned for its effectiveness with linear and non-linear data and its exceptional handling of the curse of dimensionality. It employs a linear or non-linear kernel to map input features into higher-dimensional space where a linear regression is performed.

3.3.6 Linear Regression

Linear regression is one of the simplest and most widely used statistical techniques for modeling the relationship between a dependent variable (target) and one or more independent variables (predictors) [76]. It assumes a linear relationship between the predictor variables and the target variable, making it particularly useful for tasks where this assumption holds true or can be reasonably approximated.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \quad (3.4)$$

where y is the predicted compressive strength, β_0 is the intercept, β_n are the coefficients.

3.3.7 Research Modeling Procedure

The Google Collaboratory Environment with a CPU accelerator and Python 3 Runtime was used to build the regression models for predicting the compressive strength of UHPC. Before training the model, the dataset was thoroughly accessed for quality and tidiness issues. The dataset was split into an 80% training set and a 20% test set using Scikit-Learn “train_test_split” with the random state set to 42. A linear regression model was built as the baseline model while five other models were created (K-Nearest Neighbors Regressor, Decision Tree Regressor, Random Forest Regressor, Extreme Gradient Boost Regressor, and Support Vector Regressor). The Scikit-Learn library was utilized in the ML training of all models except the XGBoost Regressor which was trained with XGB library. Each ML model was trained using the raw dataset and scaled dataset (scaled with a Standard scaler from Scikit-Learn). The performance of each model for both scaled and unscaled datasets was similar. After each training, the performance of the models was evaluated with three regression metrics (R^2 score, RMSE, and MAE) imputed from Scikit-Learn library. Hyperparameter tuning was performed with GridSearchCV from Scikit-Learn library with 10-fold cross-validation techniques. The effect of feature selection was experimented on the datasets to refine the input parameters further. The top eight (8) features were selected using each input feature's absolute correlation rank coefficients with the target variable.

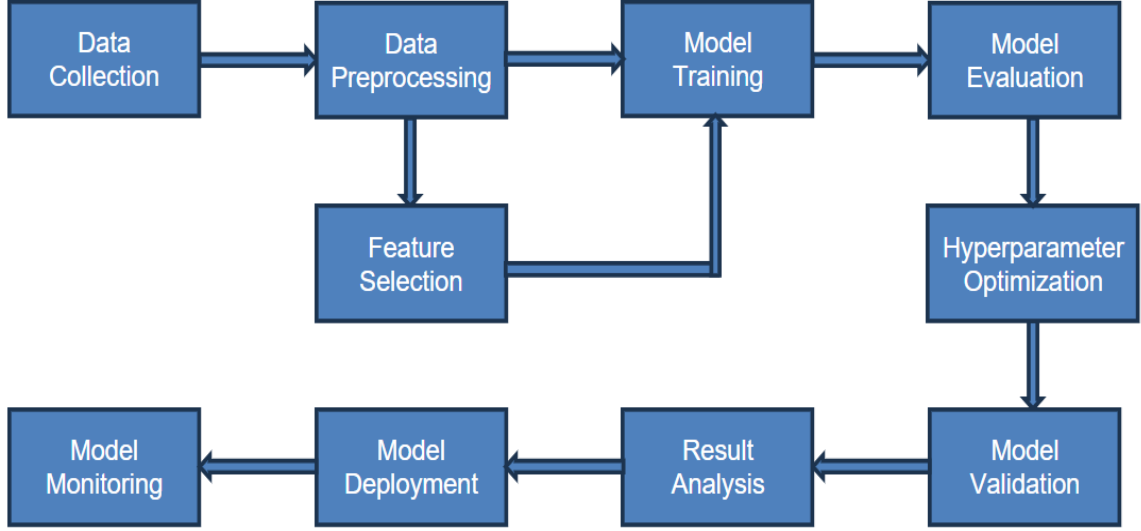


Figure 3-4: Experimental Setup

3.3.8 Model Evaluation

Three evaluation metrics (R^2 score, RMSE, and MAE) were used to evaluate the performance of each model built for UHPC compressive strength prediction.

- Determination Coefficient (R^2)

$$R^2 = 1 - \frac{\sum_{k=0}^n (y_k - \hat{y}_k)^2}{\sum_{k=0}^n (y_k - \bar{y}_k)^2} \quad (3.5)$$

where n is the number of observations, y_k is the actual value, $\hat{y}_k$ is the predicted value, and $\bar{y}_k$ is the mean of the actual value.

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=0}^n (y_k - \hat{y}_k)^2} \quad (3.6)$$

where n is the number of observations, y_k is the actual value, and $\hat{y}_k$ is the predicted value.

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{k=0}^n |y_k - \hat{y}_k| \quad (3.7)$$

where n is the number of observations, y_k is the actual value, and $\hat{y}_k$ is the predicted value.

3.3.9 Optimization Strategy

GridSearchCV with 10-fold cross-validation was used to identify the optimum parameter for each model with the best performance. The hyperparameter grid for each algorithm is provided below.

- **KNN:** `n_neighbors` {1, 2, 3, 4, 5, 6, 7}, `algorithm` {auto, ball_tree, kd_tree, brute}, `weights` {uniform, distance}, `metric` {minkowski, euclidean, manhattan}.
- **DTR:** `max_depth` {None, 1, 2, 3, 4, 5}, `min_samples_split` {2, 3, 4, 5, 6, 7}, `max_features` {None, 1, 2, 3, 4, 5, 6, 7}
- **RFR:** `n_estimators` {50, 100, 200, 300, 400, 500}, `max_depth` {None, 1, 2, 3, 4, 5}, `min_samples_split` {2, 3, 4, 5, 6, 7}, `max_features` {None, 2, 5, 10, 15}
- **XGBR:** `n_estimators` {50, 100, 200, 300, 400, 500}, `max_depth` {None, 3, 5, 10}, `learning_rate` {2, 3, 4, 5, 6, 7}, `subsample` {0.5, 0.7, 1.0}

- **SVR:** kernel {linear, poly, rbf, sigmoid}, C {1.0, 10.0, 100.0, 1000.0}, gamma {scale, auto}

3.4 Results and Discussion

3.4.1 Initial Iteration Results

Table 3-5 presents the comparison of the machine learning (ML) models for this research after hyperparameter optimization of each model and before experimenting with feature selection. The hyperparameter optimization values for each model are presented in Table 6. From Table 3-5, XGBR achieved the best performance with an R^2 score of 96.83%, followed by RFR which achieved a score of 95.62%. SVM Regressor has the lowest performance with a score of 70.01% worse than the baseline model (LR) which has a score of 70.49%. The performance of each model can be easily visualized in Figure 3-6 which shows the scatter plots of the actual compressive strength of UHPC to the predicted values for each algorithm.

Also, the three tree-based models (DTR, RFR, and XGBR) were the best-performing models of the six models created. Further exploration of these models shown in Figure 3-5 revealed the input variables that contributed the most to the outcome of their predictions. For the XGBR model, the five most important features that contributed towards its predictions are SF, RH, Age, SP and Fi respectively. On the other hand, the five most important features for DTR and RFR are Age, SF, SP, Fi, and C respectively. There exist

some similarities and differences in the way that each of the three models ranked the importance of each feature to its prediction.

Table 3-5: Performance Measure of Each Model for both training and test sets

Dataset	Performance Measure	LR	KNN	DTR	RFR	XGBR	SVR
Train	R ²	0.7183	0.8583	0.9933	0.9889	0.9921	0.7031
	RMSE	21.5321	15.2739	3.3274	4.2712	3.6100	22.1035
	MAE	17.2476	11.462	1.1821	2.7830	2.1084	16.7571
Test	R ²	0.7049	0.7746	0.9065	0.9562	0.9683	0.7001
	RMSE	20.9881	18.3396	11.8145	8.0882	6.8801	21.1583
	MAE	16.5898	13.8558	7.7100	5.8697	4.9144	16.0633

Table 3-6: Hyperparameter Optimization of Each Model Before Feature Selection

ML Models	Hyperparameters	Values
KNN	Algorithm	Brute
	Metrics	Manhattan
	N_neighbors	5
	Weights	Uniform
DTR	Max_depth	None
	Max_features	none
	Min_sample_split	2
RFR	Max_depth	None
	Max_features	10
	Min_sample_split	2
	N_estimators	200
XGBR	Max_depth	5
	Learning_rate	0.1
	Subsample	0.7
	N_estimators	400

ML Models	Hyperparameters	Values
	Gamma	0.0
SVR	Kernel	Linear
	C	1.0

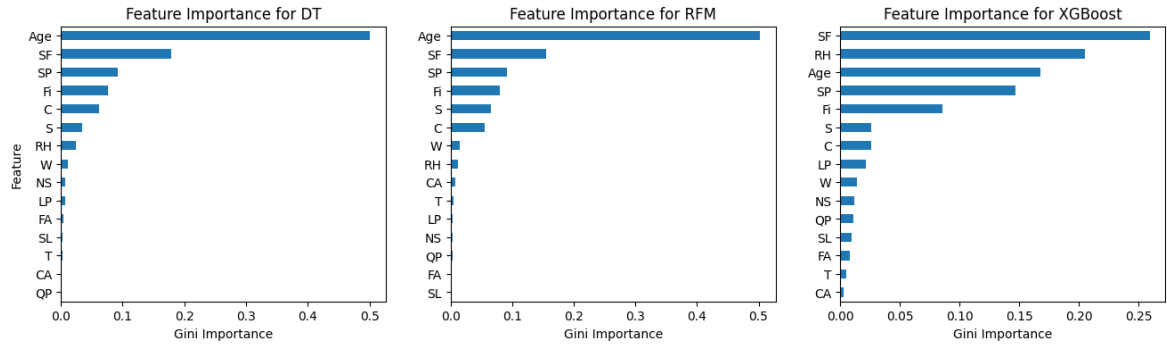


Figure 3-5: Feature Importance for Tree-based Models before Feature Selection

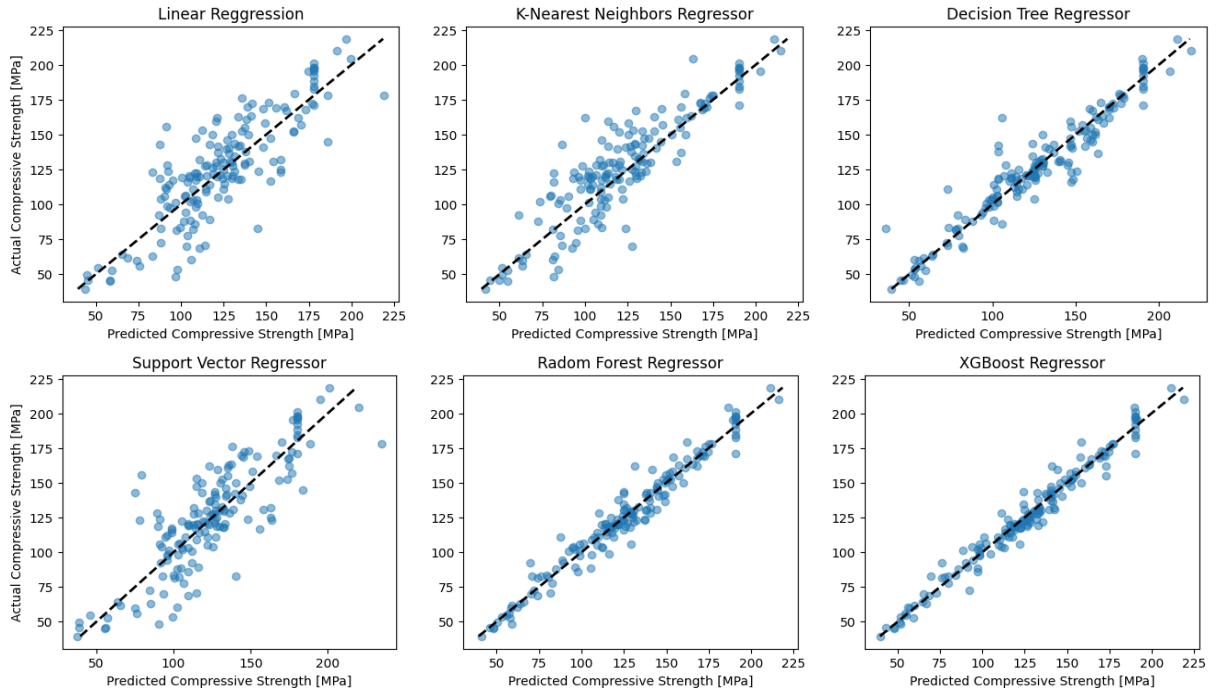


Figure 3-6: Test performance of each ML model before feature selection

3.4.2 Comparing Results with Similar Studies

On the dataset used for this study, five journals have already been published from 2020 to 2024. Table 3-7 presents the comparison of the performance of the models in those publications to the outcome of this research. From Table 3-7, it is evidence that several ML algorithms have been utilized in previous studies to predict UHPC compressive strength. This research took a different approach using a data-centric framework known as “Let the data speak”. The dataset is first fitted by simple models and their performance was evaluated. The complexity of the ML algorithms is gradually increased based on the performance of the previous models. This approach helps to determine whether simple models, which are usually easy to build and interpret, can unravel the relationship between the input features and the target. It is only when simple models’ performances were unsatisfactory that complex models, where computational efficiency and explainability are usually a challenge, will be built.

The best-performing model (XGBR) in this research matches the performance of most models reported in the previous articles. There are few models however that have higher performance than the one reported in this article. ETR, XGBR, and GBR reported by Kumar et al., 2023 [33] and Gene Expression Programming method (GEP) reported by Alabduljabbar et al., 2023 [38] have scores of 99.3% and 99.2% respectively. These models outperform the XGBR model presented in this study by 2.58%. Other metrics of measurement such as RMSE and MAE seem to be quite similar in all the studies except the values recorded by Kumar et al., 2023 which are extremely low. These values appear to be problematic because they almost mirror the ideal conditions without much

justification for those results. The main contribution of this research however is the implementation of feature selection which is discussed in the next section.

Table 3-7: Comparison of study's performance with previous studies

Ref.	ML Algorithm	R2	RMSE	MAE
[12]	Lasso model	0.950	8.959	6.621
	Autofeat model	0.962	7.873	6.215
[25]	XGBR	0.969	7.238	5.387
	Ada-XGBR	0.961	6.347	8.111
	Bagging-XGBR	0.966	7.552	5.728
	Stacking-XGBR	0.970	7.163	5.412
	Voting-XGBR	0.968	7.298	5.601
[26]	GEP	0.992	3.635	4.762
[23]	DNN	0.953	0.046	0.031
	GBR	0.993	0.018	0.007
	XGBR	0.993	0.018	0.007
	RFM	0.954	0.045	0.032
	ETR	0.993	0.018	0.007
	VR	0.964	0.040	0.029
[28]	RFR	0.926	10.729	7.240
	GBR	0.954	8.467	5.465
	ETR	0.942	9.473	6.029
This study	LR	0.704	20.988	16.590
	KNN	0.775	18.340	13.856
	DT	0.907	11.815	7.710
	RFM	0.956	8.088	5.870
	XGBR	0.968	6.880	4.914
	SVM	0.700	21.158	16.063

3.4.3 Effects of Feature Selection on the Dataset

Eight (8) features out of the fifteen (15) features in the dataset that have 0.2 or more absolute correlation values with the target feature (UHPC compressive strength) were selected to build predictive models. The eight features are C, SL, SF, QP, NS, SP, Fi, and

Age. Table 3-8 presents the performance of each of the models created and Table 3-9 shows the hyperparameter optimization values used to achieve those performances.

The performance of each algorithm is quite similar to the performance with the entire dataset. The feature selection process was able to eliminate noise in the dataset. XGBR still performs better than all the other models with a score of 95.72%, RFR is the next in performance with a score of 94.68% and the least performing algorithm is SVM with a score of 59.30%. However, KNN Regressor performs better (80.79%) with 8 variables than it did (77.49%) with the complete datasets. Figure 3-8 provides a visual representation of the actual values against the predicted values of UHPC compressive strength.

The three top-performing ML techniques remain the tree-based models (XGBR, RFR, and DTR). Figure 3-7 presents the feature importance chart of each of the three models. Although the degree of contribution in each model differs, the three models show that the most important features are Age, SF, and SP. The performance of XGBR is sufficient to approximate the compressive strength of UHPC and can be used for predictive models for UHPC compressive strength.

Table 3-8: Performance Measure of Each Model for both training and test sets

Dataset	Performance Measure	LR	KNN	DTR	RFR	XGBR	SVM
Training	R ²	0.6466	0.9917	0.9917	0.9872	0.9909	0.6219
	RMSE	24.1199	3.7032	3.6983	4.5935	3.8640	24.9477
	MAE	20.0036	1.4707	1.4688	3.0577	2.1601	19.6145
	R ²	0.6153	0.8079	0.9108	0.9468	0.9572	0.5930

Test	RMSE	23.9629	16.9348	11.5353	8.9114	7.9924	24.6454
	MAE	19.2726	12.8384	7.8454	6.5440	5.5096	18.9378

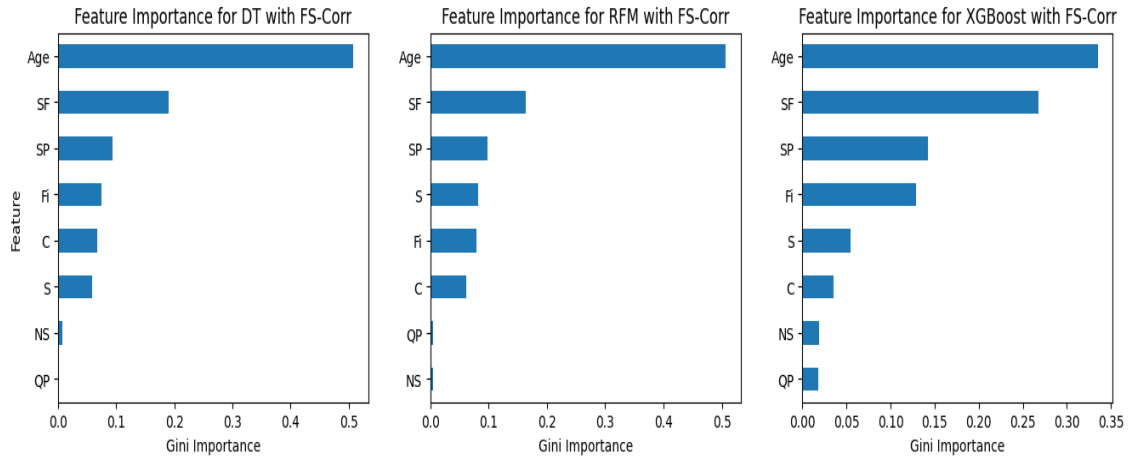


Figure 3-7: Feature Importance for Tree-based Models after Feature Selection

Table 3-9: Hyperparameter Optimization of Each Model After Feature Selection

ML Models	Hyperparameters	Values
KNN	Algorithm	Ball_tree
	Metrics	Manhattan
	N_neighbors	16
	Weights	Distance
DT	Max_depth	None
	Max_features	None
	Min_sample_split	2
RFR	Max_depth	None
	Max_features	7
	Min_sample_split	2
	N_estimators	100
XGBR	Max_depth	5

ML Models	Hyperparameters	Values
	Learning_rate	0.3
	Subsample	0.7
	N_estimators	400
	Gamma	1.5
SVR	Kernel	Linear
	C	1.0

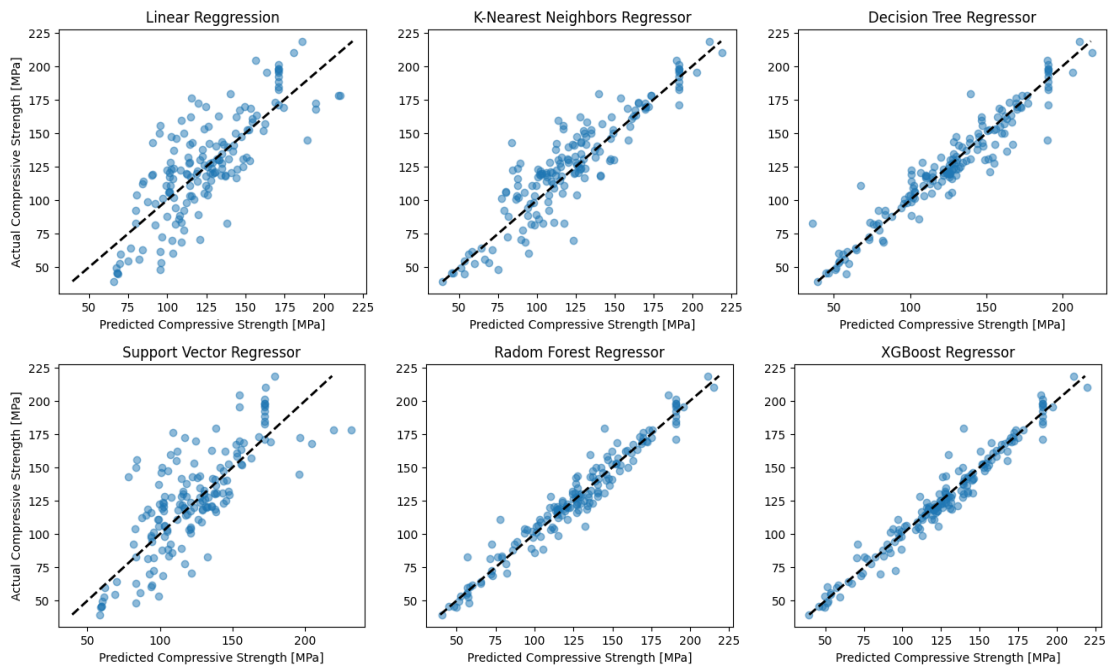


Figure 3-8: Testing Performance of each ML Model with Feature Selection

3.5 Conclusion and Recommendation

3.5.1 Conclusion

This study has demonstrated the efficacy of ML techniques to optimize the compressive strength of UHPC. By employing a variety of ML algorithms, including LR, KNN, DTR, RFR, XGBR, and SVR, the study successfully developed predictive models that accurately estimate UHPC compressive strength based on input parameters.

Key findings from the study indicate that ensemble methods like RFR and XBGR, provide superior predictive performance to traditional empirical methods.

The feature selection technique using correlation rank effectively identified the eight (8) most influential parameters in UHPC compressive strength prediction, thereby improving model accuracy and reducing computational complexity.

The research also highlighted the importance of a robust dataset and the need for meticulous data preprocessing to ensure the reliability of the predictions. The models developed in this study can significantly streamline the UHPC mix design process, reducing the time and resources required for experimental trials and leading to more efficient and cost-effective construction practices.

3.5.2 Recommendations

- Future research should focus on expanding the dataset to include more diverse and comprehensive data points. This will enhance the generalizability of the machine

learning models and improve their predictive accuracy across different UHPC formulations.

- Real-world UHPC projects should be undertaken to validate the practical applicability of the developed models. This will help to assess the models' performance under actual construction conditions and may lead to further refinement.

CHAPTER 4

CONCLUSIONS AND RECOMMENDATIONS

4.1 Conclusion

The study elucidates the relationship between dependent variables (input features) and the target variables (UHPC compressive strength). Scatter plot diagram reveal that the relationship is known linear and complex of the dataset was effectively captured by ML models created.

This study demonstrates the efficiency of ML in predicting the compressive strength of UHPC, marking a significant advancement over traditional empirical approaches. By leveraging diverse ML models such as RFR, XGBR, and SVR, this research has shown that ML algorithms can capture complex, non-linear relationships between UHPC mix design parameters and their resulting compressive strength. The models developed through this study achieved high prediction accuracy, as reflected in metrics like R^2 , RMSE, and MAE.

The application feature selection techniques through correlation rank help to reduce the dimensionality of the dataset by reducing the input features from 15 to 8.

The research outcomes underscore the potential of ML as a reliable, data-driven tool for UHPC optimization, offering benefits in terms of accuracy, efficiency, and scalability. By

integrating ML into the UHPC design process, the construction industry could accelerate mix design optimization, ultimately improving material performance while reducing costs associated with experimental testing.

4.2 Recommendations for Practical Application

4.2.1 Adoption in Industry

The integration of ML into the design and optimization of UHPC offers significant potential for enhancing efficiency and reducing the costs associated with traditional empirical methods. To support practical implementation, construction and materials engineering firms are encouraged to adopt ML models in their UHPC mix design processes. These models can streamline the identification of optimal mix proportions for specific performance targets, such as high compressive strength, without relying solely on costly and time-intensive experimental testing. Industry stakeholders should consider creating accessible, standardized datasets to further facilitate the application of ML across various UHPC projects.

4.2.2 Best Practices for Model Use

For effective application of ML models in UHPC mix optimization, adherence to certain best practices is recommended:

- **Dataset Quality:** Ensure that datasets are comprehensive, diverse, and include a wide range of mix proportions to increase model accuracy and generalizability. High-quality data from a variety of sources, including literature and experimental findings, can significantly enhance prediction reliability.
- **Model Selection:** Select ML models based on specific project needs and available computational resources. Ensemble models like RF and XGBR have shown strong predictive performance, making them suitable choices for applications where high accuracy is critical.
- **Interpretability Tools:** Utilize Explainable AI (XAI) tools, such as SHAP and Partial Dependence Plot (PDP), to gain insight into the influence of individual mix components. This enables engineers to understand the impact of key parameters, making it easier to adjust designs based on specific performance requirements.

4.3 Limitations and Future Work

4.3.1 Data Limitations

While this study successfully applied ML to predict UHPC compressive strength, certain limitations remain. First, the availability and quality of data pose significant constraints. The datasets used in this research were primarily collected from published studies, which may not fully represent the wide variability in UHPC mix designs. Additionally, many datasets lack comprehensive details on curing conditions, environmental factors, and the

specific properties of aggregates, all of which can impact UHPC performance. Expanding data sources to include these factors could improve model accuracy and applicability.

4.3.2 Model Interpretability

Although Explainable AI (XAI) techniques such as SHAP and Partial Dependence Plot (PDP) were employed to enhance interpretability, some ML models remain challenging to interpret, especially for practitioners without a background in data science. Ensemble models like RFR and XGBR, while accurate, are complex and can act as "black boxes," limiting insights into the precise influence of each parameter. Future work could focus on developing more interpretable models or integrating additional XAI techniques to improve transparency without sacrificing accuracy.

4.3.3 Opportunities for Future Research

- **Expand Dataset Diversity:** Collecting more diverse datasets, including experimental results from various UHPC formulations and conditions, would enable more robust model training.
- **Hybrid Model Development:** Combining traditional empirical models with machine learning approaches could yield hybrid models that leverage the strengths of both methodologies. Such models would provide interpretable, data-driven insights while ensuring reliability.

- **Advanced Explainability Methods:** Incorporating advanced explainability methods, such as counterfactual analysis or causal inference, could provide clearer insights into model predictions, making them more accessible to engineers and stakeholders without ML expertise.

References

- [1] M. Amran, S. S. Huang, A. M. Onaizi, N. Makul, H. S. Abdelgader, and T. Ozbakkaloglu, “Recent trends in ultra-high performance concrete (UHPC): Current status, challenges, and future prospects,” Oct. 17, 2022, *Elsevier Ltd.* doi: 10.1016/j.conbuildmat.2022.129029.
- [2] D. Ravichandran, P. R. Prem, S. K. Kaliyavaradhan, and P. S. Ambily, “Influence of fibers on fresh and hardened properties of Ultra High Performance Concrete (UHPC)—A review,” Oct. 01, 2022, *Elsevier Ltd.* doi: 10.1016/j.jobbe.2022.104922.
- [3] J. Abellán-García, J. S. Carvajal-Muñoz, and C. Ramírez-Munévar, “Application of ultra-high-performance concrete as bridge pavement overlays: Literature review and case studies,” Jan. 05, 2024, *Elsevier Ltd.* doi: 10.1016/j.conbuildmat.2023.134221.
- [4] M. K. Almustafa and M. L. Nehdi, “Machine learning prediction of structural response of steel fiber-reinforced concrete beams subjected to far-field blast loading,” *Cem Concr Compos*, vol. 126, 2022, doi: 10.1016/j.cemconcomp.2021.104378.
- [5] N. Gong and N. Zhang, “Predict the compressive strength of ultra high-performance concrete by a hybrid method of machine learning,” *Journal of Engineering and Applied Science*, vol. 70, no. 1, 2023, doi: 10.1186/s44147-023-00274-w.
- [6] A. Marani, A. Jamali, and M. L. Nehdi, “Predicting ultra-high-performance concrete compressive strength using tabular generative adversarial networks,” *Materials*, vol. 13, no. 21, pp. 1–24, 2020, doi: 10.3390/ma13214757.
- [7] A. Kumar *et al.*, “Compressive Strength Prediction of Lightweight Concrete: Machine Learning Models,” *Sustainability (Switzerland)*, vol. 14, no. 4, 2022, doi: 10.3390/su14042404.
- [8] C. Tavares, X. Wang, S. Saha, and Z. Grasley, “Machine learning-based mix design tools to minimize carbon footprint and cost of UHPC. Part 1: Efficient data collection and modeling,” *Cleaner Materials*, vol. 4, 2022, doi: 10.1016/j.clema.2022.100082.

- [9] C. Tavares and Z. Grasley, "Machine learning-based mix design tools to minimize carbon footprint and cost of UHPC. Part 2: Cost and eco-efficiency density diagrams," *Cleaner Materials*, vol. 4, 2022, doi: 10.1016/j.clema.2022.100094.
- [10] S. Jonnalagadda and S. Chava, "Ultra-High-Performance Concrete (UHPC): A state-of-the-art review of material behavior, structural applications and future," 2023, *Department of Civil and Environmental Engineering*. doi: 10.56748/ejse.23426.
- [11] S. Saleem, F. Hejazi, and N. Ostovar, "Health monitoring of ultra high fiber performance reinforced concrete communication tower using machine learning algorithms," *J Civ Struct Health Monit*, 2023, doi: 10.1007/s13349-023-00688-3.
- [12] Z. Shen, A. F. Deifalla, P. Kamiński, and A. Dyczko, "Compressive Strength Evaluation of Ultra-High-Strength Concrete by Machine Learning," *Materials*, vol. 15, no. 10, 2022, doi: 10.3390/ma15103523.
- [13] E. Ford, S. Kailas, K. Maneparambil, and N. Neithalath, "Machine learning approaches to predict the micromechanical properties of cementitious hydration phases from microstructural chemical maps," *Constr Build Mater*, vol. 265, 2020, doi: 10.1016/j.conbuildmat.2020.120647.
- [14] G. Sun, M. Du, B. Shan, J. Shi, and Y. Qu, "Ultra-high performance concrete design method based on machine learning model and steel slag powder," *Case Studies in Construction Materials*, vol. 17, 2022, doi: 10.1016/j.cscm.2022.e01682.
- [15] A. Ibrahim Bibi Farouk, J. Zhu, J. Ding, and S. I. Haruna, "Prediction and uncertainty quantification of ultimate bond strength between UHPC and reinforcing steel bar using a hybrid machine learning approach," *Constr Build Mater*, vol. 345, 2022, doi: 10.1016/j.conbuildmat.2022.128360.
- [16] M. Huu Nguyen, T.-A. Nguyen, and H.-B. Ly, "Ensemble XGBoost schemes for improved compressive strength prediction of UHPC," *Structures*, vol. 57, 2023, doi: 10.1016/j.istruc.2023.105062.
- [17] S. Mahjoubi, R. Barhemat, W. Meng, and Y. Bao, "AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC)," *Resour Conserv Recycl*, vol. 189, 2023, doi: 10.1016/j.resconrec.2022.106741.
- [18] D. L. Zou, L. L. Wu, Y. F. Hao, L. Xu, and J. J. Chen, "Composition-strength relationship study of ultrahigh performance fiber reinforced concrete (UHPFRC)

- using an interpretable data-driven approach,” *Constr Build Mater*, vol. 392, 2023, doi: 10.1016/j.conbuildmat.2023.131973.
- [19] A. Marani, A. Jamali, and M. L. Nehdi, “Predicting ultra-high-performance concrete compressive strength using tabular generative adversarial networks,” *Materials*, vol. 13, no. 21, pp. 1–24, 2020, doi: 10.3390/ma13214757.
 - [20] M. I. Khan, Y. M. Abbas, G. Fares, and F. K. Alqahtani, “Strength prediction and optimization for ultrahigh-performance concrete with low-carbon cementitious materials – XG boost model and experimental validation,” *Constr Build Mater*, vol. 387, 2023, doi: 10.1016/j.conbuildmat.2023.131606.
 - [21] J. Abellán-García, “Study of nonlinear relationships between dosage mixture design and the compressive strength of UHPC,” *Case Studies in Construction Materials*, vol. 17, 2022, doi: 10.1016/j.cscm.2022.e01228.
 - [22] N. H. Nguyen, J. Abellán-García, S. Lee, E. Garcia-Castano, and T. P. Vo, “Efficient estimating compressive strength of ultra-high performance concrete using XGBoost model,” *Journal of Building Engineering*, vol. 52, Jul. 2022, doi: 10.1016/j.jobbe.2022.104302.
 - [23] E. Ghafari, M. Bandarabadi, H. Costa, and E. Júlio, “Design of UHPC using artificial neural networks,” in *Brittle Matrix Composites 10, BMC 2010*, 2012, pp. 61–69. doi: 10.1533/9780857099891.61.
 - [24] J. Zhang, Y. Zhao, and H. Li, “Experimental Investigation and Prediction of Compressive Strength of Ultra-High Performance Concrete Containing Supplementary Cementitious Materials,” *Advances in Materials Science and Engineering*, vol. 2017, 2017, doi: 10.1155/2017/4563164.
 - [25] D. Qu, X. Cai, and W. Chang, “Evaluating the effects of steel fibers on mechanical properties of ultra-high performance concrete using artificial neural networks,” *Applied Sciences (Switzerland)*, vol. 8, no. 7, Jul. 2018, doi: 10.3390/app8071120.
 - [26] O. R. Abuodeh, J. A. Abdalla, and R. A. Hawileh, “Assessment of compressive strength of Ultra-high Performance Concrete using deep machine learning techniques,” *Applied Soft Computing Journal*, vol. 95, 2020, doi: 10.1016/j.asoc.2020.106552.
 - [27] D. Fan *et al.*, “Precise design and characteristics prediction of Ultra-High Performance Concrete (UHPC) based on artificial intelligence techniques,” *Cem Concr Compos*, vol. 122, Sep. 2021, doi: 10.1016/j.cemconcomp.2021.104171.

- [28] J. Abellán-García, “Study of nonlinear relationships between dosage mixture design and the compressive strength of UHPC,” *Case Studies in Construction Materials*, vol. 17, 2022, doi: 10.1016/j.cscm.2022.e01228.
- [29] J. Abellán García, J. Fernández Gómez, and N. Torres Castellanos, “Properties prediction of environmentally friendly ultra-high-performance concrete using artificial neural networks,” *European Journal of Environmental and Civil Engineering*, vol. 26, no. 6, pp. 2319–2343, 2022, doi: 10.1080/19648189.2020.1762749.
- [30] N. H. Nguyen, J. Abellán-García, S. Lee, E. Garcia-Castano, and T. P. Vo, “Efficient estimating compressive strength of ultra-high performance concrete using XGBoost model,” *Journal of Building Engineering*, vol. 52, Jul. 2022, doi: 10.1016/j.jobbe.2022.104302.
- [31] Y. Qian, M. Sufian, O. Accouche, and M. Azab, “Advanced machine learning algorithms to evaluate the effects of the raw ingredients on flowability and compressive strength of ultrahigh-performance concrete,” 2022. doi: 10.1371/journal.pone.0278161.
- [32] S. Mahjoubi, W. Meng, and Y. Bao, “Auto-tune learning framework for prediction of flowability, mechanical properties, and porosity of ultra-high-performance concrete (UHPC),” *Appl Soft Comput*, vol. 115, 2022, doi: 10.1016/j.asoc.2021.108182.
- [33] R. Kumar, B. Rai, and P. Samui, “A comparative study of prediction of compressive strength of ultra-high performance concrete using soft computing technique,” *Structural Concrete*, vol. 24, no. 4, pp. 5538–5555, 2023, doi: 10.1002/suco.202200850.
- [34] M. I. Khan, Y. M. Abbas, G. Fares, and F. K. Alqahtani, “Strength prediction and optimization for ultrahigh-performance concrete with low-carbon cementitious materials – XG boost model and experimental validation,” *Constr Build Mater*, vol. 387, 2023, doi: 10.1016/j.conbuildmat.2023.131606.
- [35] S. Mahjoubi, R. Barhemat, W. Meng, and Y. Bao, “AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC),” *Resour Conserv Recycl*, vol. 189, 2023, doi: 10.1016/j.resconrec.2022.106741.
- [36] D. L. Zou, L. L. Wu, Y. F. Hao, L. Xu, and J. J. Chen, “Composition-strength relationship study of ultrahigh performance fiber reinforced concrete (UHPFRC)

- using an interpretable data-driven approach,” *Constr Build Mater*, vol. 392, 2023, doi: 10.1016/j.conbuildmat.2023.131973.
- [37] M. Huu Nguyen, T.-A. Nguyen, and H.-B. Ly, “Ensemble XGBoost schemes for improved compressive strength prediction of UHPC,” *Structures*, vol. 57, 2023, doi: 10.1016/j.istruc.2023.105062.
 - [38] H. Alabduljabbar, M. Khan, H. H. Awan, S. M. Eldin, R. Alyousef, and A. M. Mohamed, “Predicting ultra-high-performance concrete compressive strength using gene expression programming method,” *Case Studies in Construction Materials*, vol. 18, 2023, doi: 10.1016/j.cscm.2023.e02074.
 - [39] N.-H. Nguyen, J. Abellán-García, S. Lee, and T. P. Vo, “From machine learning to semi-empirical formulas for estimating compressive strength of Ultra-High Performance Concrete,” *Expert Syst Appl*, vol. 237, 2024, doi: 10.1016/j.eswa.2023.121456.
 - [40] Y. Qian, M. Sufian, O. Accouche, and M. Azab, “Advanced machine learning algorithms to evaluate the effects of the raw ingredients on flowability and compressive strength of ultrahigh-performance concrete,” 2022. doi: 10.1371/journal.pone.0278161.
 - [41] S. Mahjoubi, W. Meng, and Y. Bao, “Auto-tune learning framework for prediction of flowability, mechanical properties, and porosity of ultra-high-performance concrete (UHPC),” *Appl Soft Comput*, vol. 115, 2022, doi: 10.1016/j.asoc.2021.108182.
 - [42] R. Yang *et al.*, “Low carbon design of an Ultra-High Performance Concrete (UHPC) incorporating phosphorous slag,” *J Clean Prod*, vol. 240, 2019, doi: 10.1016/j.jclepro.2019.118157.
 - [43] X. Zhang, S. Zhao, Z. Liu, and F. Wang, “Utilization of steel slag in ultra-high performance concrete with enhanced eco-friendliness,” *Constr Build Mater*, vol. 214, pp. 28–36, 2019, doi: 10.1016/j.conbuildmat.2019.04.106.
 - [44] R. Yu, P. Spiesz, and H. J. H. Brouwers, “Mix design and properties assessment of Ultra-High Performance Fibre Reinforced Concrete (UHPFRC),” *Cem Concr Res*, vol. 56, pp. 29–39, 2014, doi: 10.1016/j.cemconres.2013.11.002.
 - [45] A. Alsalman, C. N. Dang, and W. Micah Hale, “Development of ultra-high performance concrete with locally available materials,” *Constr Build Mater*, vol. 133, pp. 135–145, 2017, doi: 10.1016/j.conbuildmat.2016.12.040.

- [46] A. Sadrmomtazi, S. Tajasosi, and B. Tahmouresi, “Effect of materials proportion on rheology and mechanical strength and microstructure of ultra-high performance concrete (UHPC),” *Constr Build Mater*, vol. 187, pp. 1103–1112, Oct. 2018, doi: 10.1016/j.conbuildmat.2018.08.070.
- [47] A. Rajasekar, K. Arunachalam, and M. Kottaisamy, “Assessment of strength and durability characteristics of copper slag incorporated ultra high strength concrete,” *J Clean Prod*, vol. 208, pp. 402–414, 2019, doi: 10.1016/j.jclepro.2018.10.118.
- [48] H. Zhang, T. Ji, B. He, and L. He, “Performance of ultra-high performance concrete (UHPC) with cement partially replaced by ground granite powder (GGP) under different curing conditions,” *Constr Build Mater*, vol. 213, pp. 469–482, 2019, doi: 10.1016/j.conbuildmat.2019.04.058.
- [49] S.-H. Kang, S.-G. Hong, and J. Moon, “The use of rice husk ash as reactive filler in ultra-high performance concrete,” *Cem Concr Res*, vol. 115, pp. 389–400, 2019, doi: 10.1016/j.cemconres.2018.09.004.
- [50] X. Liang, C. Wu, Y. Su, Z. Chen, and Z. Li, “Development of ultra-high performance concrete with high fire resistance,” *Constr Build Mater*, vol. 179, pp. 400–412, 2018, doi: 10.1016/j.conbuildmat.2018.05.241.
- [51] R. Yu, P. Spiesz, and H. J. H. Brouwers, “Effect of nano-silica on the hydration and microstructure development of Ultra-High Performance Concrete (UHPC) with a low binder amount,” *Constr Build Mater*, vol. 65, pp. 140–150, 2014, doi: 10.1016/j.conbuildmat.2014.04.063.
- [52] M. Gesoglu, E. Güneyisi, D. S. Asaad, and G. F. Muhyaddin, “Properties of low binder ultra-high performance cementitious composites: Comparison of nanosilica and microsilica,” *Constr Build Mater*, vol. 102, pp. 706–713, 2016, doi: 10.1016/j.conbuildmat.2015.11.020.
- [53] Z. Wu, C. Shi, K. H. Khayat, and L. Xie, “Effect of SCM and nano-particles on static and dynamic mechanical properties of UHPC,” *Constr Build Mater*, vol. 182, pp. 118–125, 2018, doi: 10.1016/j.conbuildmat.2018.06.126.
- [54] S.-H. Kang, Y. Jeong, K. H. Tan, and J. Moon, “The use of limestone to replace physical filler of quartz powder in UHPFRC,” *Cem Concr Compos*, vol. 94, pp. 238–247, 2018, doi: 10.1016/j.cemconcomp.2018.09.013.
- [55] K. Janković, S. Stanković, D. Bojović, M. Stojanović, and L. Antić, “The influence of nano-silica and barite aggregate on properties of ultra high

- performance concrete,” *Constr Build Mater*, vol. 126, pp. 147–156, 2016, doi: 10.1016/j.conbuildmat.2016.09.026.
- [56] C. Wang, C. Yang, F. Liu, C. Wan, and X. Pu, “Preparation of Ultra-High Performance Concrete with common technology and materials,” *Cem Concr Compos*, vol. 34, no. 4, pp. 538–544, 2012, doi: 10.1016/j.cemconcomp.2011.11.005.
- [57] E. Ghafari, H. Costa, E. Júlio, A. Portugal, and L. Durães, “The effect of nanosilica addition on flowability, strength and transport properties of ultra high performance concrete,” *Mater Des*, vol. 59, pp. 1–9, 2014, doi: 10.1016/j.matdes.2014.02.051.
- [58] A. Khaloo, M. H. Mobini, and P. Hosseini, “Influence of different types of nano-SiO₂ particles on properties of high-performance concrete,” *Constr Build Mater*, vol. 113, pp. 188–201, 2016, doi: 10.1016/j.conbuildmat.2016.03.041.
- [59] Z. Wu, C. Shi, K. H. Khayat, and S. Wan, “Effects of different nanomaterials on hardening and performance of ultra-high strength concrete (UHSC),” *Cem Concr Compos*, vol. 70, pp. 24–34, 2016, doi: 10.1016/j.cemconcomp.2016.03.003.
- [60] S. Ahmad, K. O. Mohaisen, S. K. Adekunle, S. U. Al-Dulaijan, and M. Maslehuddin, “Influence of admixing natural pozzolan as partial replacement of cement and microsilica in UHPC mixtures,” *Constr Build Mater*, vol. 198, pp. 437–444, Feb. 2019, doi: 10.1016/j.conbuildmat.2018.11.260.
- [61] R. Yu, P. Spiesz, and H. J. H. Brouwers, “Development of Ultra-High Performance Fibre Reinforced Concrete (UHPFRC): Towards an efficient utilization of binders and fibres,” *Constr Build Mater*, vol. 79, pp. 273–282, 2015, doi: 10.1016/j.conbuildmat.2015.01.050.
- [62] Q. Song, R. Yu, Z. Shui, X. Wang, S. Rao, and Z. Lin, “Optimization of fibre orientation and distribution for a sustainable Ultra-High Performance Fibre Reinforced Concrete (UHPFRC): Experiments and mechanism analysis,” *Constr Build Mater*, vol. 169, pp. 8–19, 2018, doi: 10.1016/j.conbuildmat.2018.02.130.
- [63] M. Gesoglu, E. Güneyisi, G. F. Muhyaddin, and D. S. Asaad, “Strain hardening ultra-high performance fiber reinforced cementitious composites: Effect of fiber type and concentration,” *Compos B Eng*, vol. 103, pp. 74–83, 2016, doi: 10.1016/j.compositesb.2016.08.004.

- [64] Z. Wu, C. Shi, W. He, and D. Wang, "Static and dynamic compressive properties of ultra-high performance concrete (UHPC) with hybrid steel fiber reinforcements," *Cem Concr Compos*, vol. 79, pp. 148–157, 2017, doi: 10.1016/j.cemconcomp.2017.02.010.
- [65] D.-Y. Yoo and M.-J. Kim, "High energy absorbent ultra-high-performance concrete with hybrid steel and polyethylene fibers," *Constr Build Mater*, vol. 209, pp. 354–363, 2019, doi: 10.1016/j.conbuildmat.2019.03.096.
- [66] Y. Li, K. H. Tan, and E.-H. Yang, "Synergistic effects of hybrid polypropylene and steel fibers on explosive spalling prevention of ultra-high performance concrete at elevated temperature," *Cem Concr Compos*, vol. 96, pp. 174–181, 2019, doi: 10.1016/j.cemconcomp.2018.11.009.
- [67] K. Wille and C. Boisvert-Cotulio, "Material efficiency in the design of ultra-high performance concrete," *Constr Build Mater*, vol. 86, pp. 33–43, 2015, doi: 10.1016/j.conbuildmat.2015.03.087.
- [68] D. Fan *et al.*, "Intelligent design and manufacturing of ultra-high performance concrete (UHPC) – A review," *Constr Build Mater*, vol. 385, 2023, doi: 10.1016/j.conbuildmat.2023.131495.
- [69] O. Abuodeh, J. A. Abdalla, and R. A. Hawileh, "Prediction of compressive strength of ultra-high performance concrete using SFS and ANN," in *2019 8th International Conference on Modeling Simulation and Applied Optimization, ICMSAO 2019*, 2019. doi: 10.1109/ICMSAO.2019.8880452.
- [70] C. Hung, S. El-Tawil, S. C.-J. of S. Engineering, and undefined 2021, "A review of developments and challenges for UHPC in structural engineering: Behavior, analysis, and designQ1Journal of Structural Engineering; H-Index: 157 SJR: Q1 CORE: NA ABDC: NA FT50: NA 2Journal of Structural Engineering; H-Index: 157 VHB: NA FNEGE: NA CoNRS: NA HCERE: NA CCF: NA BFI: 2 AJG: NA +," *ascelibrary.orgCC Hung, S El-Tawil, SH ChaoJournal of Structural Engineering, 2021•ascelibrary.org*, Accessed: May 19, 2024. [Online]. Available: [https://ascelibrary.org/doi/full/10.1061/\(ASCE\)ST.1943-541X.0003073](https://ascelibrary.org/doi/full/10.1061/(ASCE)ST.1943-541X.0003073)
- [71] X. Zhang, Z. Wu, J. Xie, X. Hu, and C. Shi, "Trends toward lower-carbon ultra-high performance concrete (UHPC) – A review," Mar. 22, 2024, *Elsevier Ltd*. doi: 10.1016/j.conbuildmat.2024.135602.

- [72] O. A. Mayhoub, E. S. A. R. Nasr, Y. A. Ali, and M. Kohail, “The influence of ingredients on the properties of reactive powder concrete: A review,” Mar. 01, 2021, *Ain Shams University*. doi: 10.1016/j.asej.2020.07.016.
- [73] S. Abdal, W. Mansour, I. Agwa, M. Nasr, A. A.- Buildings, and undefined 2023, “Application of ultra-high-performance concrete in bridge engineering: Current status, limitations, challenges, and future prospectsQ1Buildings; H-Index: 35 SJR: Q1 CORE: NA ABDC: NA FT50: NA,” *mdpi.com*, Accessed: May 19, 2024. [Online]. Available: <https://www.mdpi.com/2075-5309/13/1/185>
- [74] A. Marani, A. Jamali, and M. L. Nehdi, “Predicting Ultra-High-Performance Concrete Compressive Strength Using Tabular Generative Adversarial Networks 4.” [Online]. Available: www.mdpi.com/journal/materials
- [75] “k-nearest neighbors algorithm - Wikipedia.” Accessed: May 19, 2024. [Online]. Available: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm
- [76] M. Tsao, “Group least squares regression for linear models with strongly correlated predictor variables,” *Ann Inst Stat Math*, vol. 75, no. 2, pp. 233–250, Apr. 2022, doi: 10.1007/s10463-022-00841-7.

Vitae

Name	: Muhammed Olawale Balogun
Nationality	: Nigerian
Date of Birth	: 08/08/1991
Email	: muhammed.balogun123@gmail.com
Address	: 30 Ajoye Street, Losoro, Ibeju-Lekki, 105101, Lagos.
Academic Background	M.Sc. Civil Engineering, KFUPM, Saudi Arabia B.Eng. Civil Engineering FUTMinna, Nigeria