NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

EXPERIMENTAL EVALUATION AND MACHINE LEARNING PREDICTION OF GEOPOLYMER CONCRETE DURABILITY IN MARINE SULFATE ENVIRONMENTS

ZHANG YUNHAN
SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

2025

EXPERIMENTAL EVALUATION AND MACHINE LEARNING PREDICTION OF GEOPOLYMER CONCRETE DURABILITY IN MARINE SULFATE ENVIRONMENTS

ZHANG YUNHAN

School of Civil and Environmental Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Master of Engineering

2025

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

2025.9.25

Date

Zhang Yunhan

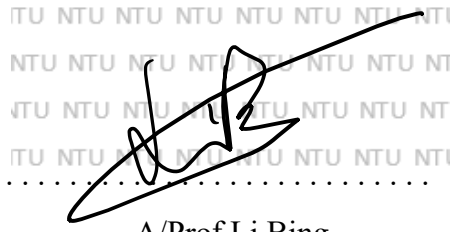
ZHANG YUNHAN

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

2025/9/24

Date



A/Prof Li Bing

Authorship Attribution Statement

This thesis does not contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

2025.9.25

Date

Zhang Yunhan

ZHANG YUNHAN

ACKNOWLEDGEMENTS

The research reported in this thesis was conducted in the School of Civil and Environmental Engineering, Nanyang Technological University, Singapore, as part of the requirement for the degree of Master of Engineering.

I would like to express my heartfelt gratitude to my supervisor, Prof. Li Bing, for his professional guidance, insightful advice, and continuous encouragement throughout my research journey. His support was invaluable in shaping the direction and quality of this work.

Special thanks are extended to the technical staff of the Protective Engineering Lab and the Construction Technology Lab at NTU.

I also wish to thank my colleagues and friends, Dr. Yang Zhihong, Dr. Yu Zecheng, Mr. Jin Yu, and Mr. Lin Haoran, for their valuable suggestions, discussions, and technical support. Their help greatly contributed to the successful completion of this research and made this journey an enriching and memorable one.

TABLE OF CONTENTS

ABSTRACT

The long-term performance of geopolymer concrete (GPC) in aggressive marine environments remains an important issue for sustainable infrastructure development. To address this, the present study investigates the degradation mechanisms and compressive strength evolution of metakaolin-slag geopolymer composites under sulfate exposure, using a combined methodology of experimental assessment and data-driven prediction.

Geopolymer specimens were synthesized using metakaolin (MK) and ground granulated blast furnace slag (GGBFS) as aluminosilicate binders, with a subset incorporating 1% basalt fiber reinforcement by volume. The specimens were exposed to a 5% sodium sulfate solution under two laboratory-simulated conditions: cyclic wetting-drying, representing tidal-zone exposure, and prolonged immersion, representing submerged exposure. Compressive strength, elastic modulus, density, and ultrasonic indicators were periodically evaluated, and low-field nuclear magnetic resonance (LF-NMR) was employed to quantify pore-structure evolution. The results show that basalt fiber reinforcement is associated with improved resistance to sulfate-induced deterioration under the investigated exposure regimes, particularly under cyclic wetting-drying, as reflected by better strength retention, reduced surface spalling, and moderated pore coarsening and microcrack-related void development.

To complement the experimental study, a machine learning framework was developed to predict compressive strength degradation under sulfate immersion exposure. A dataset comprising 369 mix-design records from 36 published studies was compiled under consistent sulfate-solution test conditions and standard 28-day curing. The model inputs are limited to mix-design variables and exposure duration as reported in the literature dataset; accordingly, the predictions are intended for the standardized sulfate-exposure conditions represented by these data. Multiple models, including Random Forest, XGBoost, Support Vector Regression, and Neural Networks, were evaluated. A hybrid approach integrating a Conditional Variational Autoencoder (CVAE) for feature learning and Optuna for hyperparameter

optimization was proposed, and the resulting CVAE-XGBoost model achieved the best overall predictive performance among the compared methods.

This thesis indicates that basalt fiber reinforcement can mitigate sulfate-induced degradation of geopolymer composites under the cyclic and immersion exposure conditions investigated in this study. Moreover, the successful application of advanced AI techniques demonstrates their potential for supporting strength-degradation prediction under sulfate exposure and informing mix design development within the studied exposure conditions. These findings provide useful insights for developing alkali-activated materials for marine construction and for supporting long-term structural resilience and sustainability.

LIST OF TABLES

LIST OF FIGURES

LIST OF NOTATIONS

a	Depth of corrosion attack (penetration)
A	Empirical constant (model parameter)
B	Empirical constant (model parameter)
A_t	Cross-sectional area of stirrup (transverse reinforcement)
b	Bias term (regression intercept)
b_1	Section width of uncorroded concrete member
b_c	Section width of corroded concrete member
C	Regularization parameter (SVR model)
d_1	Diameter of steel reinforcement bar (original)
d_c	Diameter of steel reinforcement bar (corroded)
E_c	Elastic modulus of concrete
E_s	Elastic modulus of uncorroded steel
$E_{s,c}$	Elastic modulus of corroded steel reinforcement
f_c	Initial concrete compressive strength
$f_{c,c}$	Concrete compressive strength after corrosion
f_t	Initial concrete tensile strength
$f_{t,c}$	Concrete tensile strength after corrosion
f_u	Ultimate tensile stress of uncorroded steel
$f_{u,c}$	Ultimate tensile stress of corroded steel
f_y	Yield strength of uncorroded steel reinforcement
$f_{y,c}$	Yield strength of corroded steel reinforcement
f_{yh}	Yield strength of stirrups (transverse reinforcement)
G_{fc}	Fracture energy of concrete in compression
G_{ft}	Fracture energy of concrete in tension
k	Bond strength reduction factor (for corroded reinforcement)
M_{cr}	Cracking moment
M_n	Nominal moment capacity
M_r	Radial moment (per unit width)
M_t	Tangential moment (per unit width)
M_u	Ultimate flexural capacity of RC slab
$Nk(x)$	Index set of the k-nearest neighbors of x (KNN model)

n_c	Number of corroded compressive reinforcement bars
R	Radius of circular column
r_i	Radius of isolated member (slab)
r	Correlation coefficient (roughness vs. diameter)
r_{cr}	Radius of crack-stabilized zone
r_s	Radius of cracked region
r_{sc}	Radius of reinforcement yield zone
r_v	Radius of load resultant location
r_y	Radius of critical shear crack
s	Spacing between stirrups
V_p	Punching shear strength of slab
V_r	Residual shear strength of slab
V_u	Ultimate shear strength of slab
w/c	Water-to-cement mass ratio
w	Weight vector (model coefficients)
W	Total crack width in concrete
α	Rust expansion factor (oxide-to-steel volume ratio)
δ	Slip (relative displacement in bond)
ζ	Curvature reduction factor (due to tension stiffening)
η	Efficiency factor for bending reinforcement (stiffness calculation)
θ	Slab rotation outside column region
μ	Corrosion level (fractional loss of steel cross-section)
ξ_i	Slack variable in SVR (lower tolerance)
ξ_i^*	Slack variable in SVR (upper tolerance)
ρ	Reinforcement ratio (steel-to-concrete area ratio)
τ_n	Nominal shear stress
τ_c	Ultimate bond strength of corroded rebar
φ_t	Tangential curvature (circumferential direction)

Chapter 1. INTRODUCTION

1.1 Research Background

Coastal and offshore infrastructure is constantly subjected to harsh marine conditions, which pose significant durability challenges for conventional ordinary Portland cement (OPC)-based concrete. The primary mechanisms of deterioration in these environments include chloride ingress and sulfate attack, both of which can lead to premature structural degradation. Chloride ions, derived from seawater, permeate the concrete matrix and reach the embedded steel reinforcement, initiating corrosion. This process not only weakens the structural integrity but also leads to the formation of expansive rust, resulting in cracking and spalling. Simultaneously, sulfate ions present in seawater or adjacent soils react with cement hydration products, such as calcium hydroxide and aluminates, forming expansive compounds like gypsum and ettringite. These reactions generate internal stresses that cause microcracking, swelling, and eventual surface deterioration.

Degradation is further exacerbated in tidal and splash zones, where cyclic wetting and drying intensify damage. During drying phases, salt crystallization within the concrete's pore structure exerts additional pressure, propagating microcracks. Moreover, repeated wet-dry cycles continuously replenish chloride and sulfate concentrations, preventing any natural recovery and leading to progressive material breakdown. Collectively, chloride-induced corrosion and sulfate attack significantly reduce the service life of OPC-based concrete in marine environments, which presents a critical challenge for the long-term sustainability of coastal infrastructure. Addressing these issues requires advanced material solutions to enhance durability and minimize maintenance costs.

In response to these challenges, geopolymer and alkali-activated binders have attracted growing attention for marine and coastal infrastructure, particularly in applications where sustainability and enhanced chemical stability are desired. Recent studies and pilot implementations have explored their use in precast elements, coastal protection units, and repair/rehabilitation materials exposed to seawater or marine

aerosols. This emerging adoption indicates that GPC is increasingly being considered as a viable alternative for marine-adjacent service environments.

However, it remains premature to assume that GPC is inherently durable under all marine exposures. Marine environments are multi-ionic and cyclic, involving not only chlorides but also sulfates, magnesium salts and pronounced wetting-drying actions in tidal and splash zones. The durability response of GPC can vary substantially with precursor chemistry and calcium content. For low-calcium systems, the reaction products are typically dominated by N-A-S-H type gels. For calcium-containing systems, C-(A)-S-H may coexist with N-A-S-H. Such differences in binder chemistry and gel structure have been reported to lead to divergent observations in sulfate resistance, volumetric stability and microcracking susceptibility in the literature. In particular, cyclic wet-dry sulfate exposure may accelerate salt crystallisation pressure and damage accumulation, yet systematic evidence for metakaolin-slag GPC and fiber-reinforced geopolymer composites under such marine-like regimes remains limited. This motivates the present study to experimentally elucidate deterioration pathways under controlled sulfate exposure scenarios and to develop predictive tools for performance evolution.

(a) (b)

Figure 1.1 Severe corrosion observed in (a) concrete pile cap and (b) concrete jacket after long-term marine exposure (Darmawan et al., 2024).

The significant environmental and durability challenges associated with OPC have spurred interest in alternative binders that offer superior chemical stability and sustainability. OPC production is highly energy-intensive and accounts for

approximately 5-7% of global anthropogenic CO_2 emissions, underscoring the urgent need for more eco-friendly alternatives. Among these, geopolymer binders synthesized via the alkali activation of aluminosilicate precursors such as fly ash, MK, and GGBFS, have emerged as a promising solution. Unlike OPC, which hardens through calcium silicate hydration, geopolymers form a three-dimensional aluminosilicate gel network, resulting in a denser and more chemically stable matrix.

This structural distinction is often associated with improved resistance in laboratory studies. Geopolymers have been widely reported to exhibit lower ion penetrability to aggressive ions, along with enhanced resistance to sulfate and acid attack. Their minimal calcium hydroxide ($Ca(OH)_2$) and tricalcium aluminate (C_3A) content mitigates expansive ettringite formation, significantly improving sulfate resistance. Furthermore, their refined pore structure and high alumina content enhance chloride-binding capacity while reducing ion diffusivity, thereby minimizing reinforcement corrosion risks. Beyond performance, GPC also addresses sustainability concerns by repurposing industrial byproducts like fly ash and slag, reducing landfill waste and carbon emissions. Accordingly, GPC is considered a promising alternative with potential durability and sustainability benefits in aggressive exposure conditions.

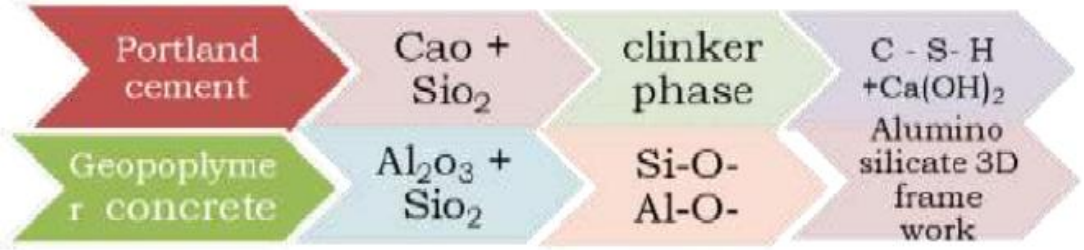


Figure 1.2 Schematic diagram of reaction mechanism in OPC and GPC (Poloju *et al*., 2020)

Geopolymers formulated with MK and GGBFS exhibit superior performance due to their complementary chemical contributions. MK provides high alumina and silica content, which are essential for geopolymerization, while GGBFS introduces calcium, enhancing early-age strength and promoting the formation of calcium-alumino-silicate hydrates (C-(A)-S-H) as secondary binding phases. Studies, such as that by Işıkdağ *et al*. (2023), have demonstrated that MK-GGBFS blended geopolymers

achieve higher early-age strength and greater durability compared to single-precursor systems. To capitalize on these synergistic effects, this thesis employs an MK-GGBFS blend activated by an optimized alkaline solution of sodium hydroxide and sodium silicate. This activator formulation enhances precursor dissolution and polymerization, resulting in a dense microstructure with low porosity and improved chemical resistance.

Despite their advantages, GPC can exhibit inherent brittleness and limited tensile capacity. To mitigate these limitations, fiber reinforcement—particularly basalt fibers—has been explored. Basalt fibers, derived from volcanic rock, offer high tensile strength, good alkali resistance, and environmental sustainability, making them suitable for harsh service conditions. Recent studies (e.g., Chakkor et al., 2025) have reported that basalt fiber-reinforced geopolymer composites can improve flexural performance, crack resistance, and water-uptake-related properties. At appropriate dosages, these fibers act as microcrack bridges, delaying crack propagation and enhancing post-cracking ductility. Such crack-control effects may also help limit crack-assisted ingress of aggressive ions (e.g., sulfates and chlorides), which is relevant for marine-adjacent exposures.

However, critical gaps remain in the understanding of GPC performance evolution under coupled deterioration mechanisms, particularly when wet-dry cycling acts concurrently with sulfate exposure. Prior studies have largely focused on OPC systems or on single-factor exposures, and systematic evidence for basalt fiber-reinforced geopolymer composites under marine-like regimes remains limited. To address this deficiency, the present work undertakes a coordinated experimental programme to elucidate the deterioration pathways and rates of metakaolin-slag geopolymer composites subjected to cyclic sulfate exposure with wet-dry cycling. Furthermore, it examines—relative to plain geopolymer matrices—the extent to which basalt fibers can suppress crack development, reduce damage accumulation, and improve compressive strength retention under the investigated exposure conditions.

In addition to experimental investigations, data-driven approaches—particularly machine learning (ML)—are increasingly used to support the prediction of material property evolution under aggressive exposures. Various ML models, including decision trees, neural networks, and ensemble regressors, have shown promising capability in learning complex mappings between mixture variables, exposure duration, and measured responses. For instance, Taffese et al. (2025) employed ML techniques to model sulfate-related performance changes in cementitious materials. However, existing studies predominantly focus on OPC systems and often simplify exposure conditions, while publicly available datasets capturing cyclic exposure effects remain limited.

To address these limitations within the scope of this thesis, a CVAE-based framework is proposed to predict compressive strength degradation of GPC under standardized marine-sulfate-simulated exposure conditions, where strength retention is treated as one deterioration-related indicator rather than a comprehensive measure of durability. The CVAE, as a generative neural network, is capable of learning high-dimensional nonlinear relationships and generating conditional synthetic samples to augment limited data. By training the framework on a combined dataset of in-house results and literature-derived data, this thesis aims to improve predictive robustness and provide a practical tool for strength-degradation prediction and performance evaluation under the investigated sulfate exposure scenarios.

1.2 Objectives and Scope

This thesis aims to advance the experimental understanding and data-driven prediction of the performance evolution of MK/GGBFS-based geopolymer concrete under marine sulfate-simulated exposure, with particular emphasis on tidal-like wet-dry cycling and continuous immersion. A metakaolin-slag geopolymer system, with and without basalt fiber reinforcement, is investigated to examine deterioration behaviour and associated performance indicators under controlled laboratory regimes. The specific objectives are as follows:

1.Experimental evaluation under marine sulfate-simulated exposure.

To experimentally evaluate the performance evolution of MK/GGBFS geopolymer concrete under sulfate-rich exposure conditions representative of tidal-zone wetting-drying and submerged immersion. Key performance indicators include compressive strength retention, elastic modulus evolution, density/mass change, and macroscopic surface damage.

2.Role of basalt fibers in deterioration mitigation.

To examine the influence of basalt fiber reinforcement (fixed at 1% by volume) by comparing mixes with and without fibers, with particular attention to crack development, damage evolution, and the retention of mechanical and acoustic indicators under cyclic exposure.

3.Machine learning prediction of compressive strength degradation under immersion exposure.

To develop a machine learning framework for predicting compressive strength degradation under continuous sulfate immersion, where strength retention is treated as one deterioration-related indicator. The model is trained using mix-design variables and exposure duration as inputs, and the CVAE-enhanced approach is benchmarked against conventional regressors and neural-network-based models using consistent evaluation metrics.

Scope and limitations:

This thesis is confined to laboratory-simulated sulfate exposure using a 5% Na_2SO_4 solution under two regimes: (i) cyclic wetting-drying and (ii) continuous immersion. The investigation focuses on the deterioration behaviour of the geopolymer matrix and associated performance indicators obtained from mechanical testing and microstructural characterization. Chloride-induced reinforcement corrosion is not addressed, as the specimens are unreinforced or reinforced only with non-corrodible basalt fibers.

The experimental programme adopts a fixed MK/GGBFS geopolymer formulation activated by sodium hydroxide and sodium silicate. Basalt fibers are incorporated at a constant dosage (1% by volume), and variations in fiber dosage and broader mix-design parameter ranges are beyond the scope of the experimental study. In addition, while real marine environments may involve additional actions (e.g., coupled chloride-sulfate exposure, temperature/humidity fluctuations, and biological effects), these factors are not explicitly simulated and are identified as directions for future work.

Scope of the Machine Learning Model

The machine learning component is developed using data from continuous sulfate immersion as reported in the compiled literature dataset, with inputs limited to mix-design variables and exposure duration. Accordingly, the model is intended to support prediction within the standardized exposure conditions represented by the dataset, rather than as a universal predictor across all in-situ marine environments or cyclic wet-dry exposure regimes. Within this defined scope, the ML study serves as a proof-of-concept for applying generative feature learning and regression models to forecast strength-degradation trends of geopolymer systems under sulfate exposure.

1.3 Organization of Thesis

This thesis is organized into six chapters to systematically present the research background, methodology, experimental findings, and the data-driven prediction framework.

Chapter 1 introduces the research background and problem context, highlighting the deterioration challenges faced by OPC-based concrete in marine environments and motivating the consideration of geopolymer concrete as a lower-carbon alternative. The chapter then states the research objectives and scope, and outlines the overall structure of the thesis.

Chapter 2 reviews the literature on concrete performance in marine exposure conditions. It summarizes the principal deterioration mechanisms relevant to marine

service, including chloride ingress and reinforcement corrosion, sulfate attack, and the aggravating role of wetting-drying cycles in tidal and splash zones. The chapter further discusses the reported performance trends of geopolymer and alkali-activated systems under chloride- and sulfate-related exposures, and reviews the role of fiber reinforcement—particularly basalt fibers—in crack control and performance retention. Finally, it surveys recent advances in data-driven prediction for cementitious materials, with emphasis on machine learning approaches for modelling strength development and degradation trends, and identifies research gaps addressed in this thesis.

Chapter 3 presents the experimental programme conducted to evaluate the performance evolution of MK/GGBFS geopolymer concrete with and without basalt fiber reinforcement under marine sulfate-simulated exposure. Two exposure regimes are investigated: cyclic wetting-drying and continuous immersion in a 5% Na_2SO_4 solution. The chapter reports the evolution of key performance indicators, including compressive strength, elastic modulus, ultrasonic pulse velocity, and bulk density/mass change. LF-NMR is used to support microstructural interpretation by tracking pore-structure evolution under a consistent conditioning protocol. Comparative analyses between fiber-reinforced and unreinforced specimens are provided to examine the influence of basalt fibers on damage development under the investigated exposure regimes.

Chapter 4 introduces the machine learning methodology for predicting compressive strength degradation of geopolymer concrete under sulfate immersion exposure. The chapter describes the dataset formulation and preprocessing strategy, and presents baseline regression models used for benchmarking. It then details the proposed CVAE-based feature learning and data augmentation framework, together with Optuna-based hyperparameter optimization, establishing the methodological basis for model training and evaluation.

Chapter 5 reports the implementation and evaluation of the machine learning models developed in Chapter 4. It compares baseline models against the CVAE-enhanced

approach using quantitative performance metrics and diagnostic plots (e.g., predicted versus measured strength, residual distributions, and degradation curves). Model interpretability is examined using SHAP analysis, and additional robustness checks and sensitivity analyses are conducted to assess the effect of latent feature dimensionality and key inputs on prediction performance.

Chapter 6 summarizes the main findings and conclusions of the thesis and provides recommendations for future research. It consolidates the experimental observations under the investigated sulfate exposure regimes and summarizes the predictive performance of the CVAE-enhanced model within the standardized dataset scope. The chapter concludes by outlining future directions to improve practical applicability, including broader environmental actions, additional performance indicators, and long-term field validation.

Chapter 2. BACKGROUND AND LITERATURE REVIEW

2.1 Introduction

Chapter 2 presents a comprehensive review of the literature relevant to the durability of GPC in marine environments, the role of basalt fiber reinforcement, and recent advances in ML for predicting concrete performance. This chapter synthesizes prior research to establish a foundation for understanding the key deterioration mechanisms and influencing factors governing the long-term durability of GPC under sulfate exposure and wetting-drying actions, and to identify the research gaps addressed by this thesis. Section 2.1 reviews the development of GPC and its sustainability advantages, together with a comparative discussion of its binder chemistry and material characteristics relative to OPC systems. Section 2.2 summarises the primary durability challenges in marine environments, including chloride ingress and sulfate attack, and reviews reported durability performance of OPC and GPC under relevant exposure actions. Section 2.3 discusses basalt fiber reinforcement with emphasis on crack control and its potential contribution to durability improvement in aggressive environments. Section 2.4 reviews ML-based methods for predicting concrete properties and performance evolution, with particular emphasis on modelling compressive-strength trends and degradation behaviour under standardized exposure conditions, including the use of CVAE-type generative learning for feature representation and data augmentation. Finally, the limitations of current research are critically evaluated, including the scarcity of long-term field evidence for geopolymer systems, the limited number of systematic durability studies on basalt fiber-reinforced GPC under marine-like exposure regimes, and the current gap in ML frameworks that can reliably model time-dependent compressive-strength degradation under consistently defined exposure conditions. These gaps motivate the present study and inform the research objectives defined in Chapter 1.

2.2 Geopolymer Concrete as a Sustainable Alternative to OPC

GPC is an inorganic polymer binder system formed by the alkaline activation of aluminosilicate precursors. In contrast to OPC, which relies on calcium-silicate-

hydrate (C-S-H) gel formation for strength development, GPC attains its structural integrity via the polycondensation of dissolved silica and alumina species, resulting in a three-dimensional aluminosilicate network. When sodium-based activators are employed, this binding phase is commonly termed sodium-alumino-silicate-hydrate (N-A-S-H) gel. The key raw materials for geopolymer synthesis include fly ash (particularly Class F, due to its high silica and alumina content), MK (a thermally activated clay with high aluminosilicate reactivity), and GGBFS, which introduces calcium silicates that influence reaction kinetics and microstructure. These precursors are combined with an alkaline activator, typically a solution of sodium hydroxide (NaOH) or potassium hydroxide (KOH) and sodium silicate (waterglass). The alkaline environment promotes the dissolution of silica and alumina from the solid precursors, followed by their reorganization into a polymeric Si-O-Al framework. This geopolymerization process yields a hardened binder with a dense, durable microstructure, offering superior mechanical properties, chemical resistance, and a lower carbon footprint compared to traditional OPC-based systems.

Meng *et al*. (2020) highlight the significant sustainability benefits of geopolymer technology. Unlike OPC which relies on energy-intensive clinker production, geopolymers utilize industrial by-products such as fly ash and slag or natural pozzolans like MK, eliminating the need for high-temperature calcination. This results in substantially lower CO_2 emissions and energy consumption. While the synthesis of alkali activators contributes to the carbon footprint of geopolymers, this impact remains far less severe than that of OPC clinker production. Additionally, geopolymer production aligns with circular economy principles by repurposing industrial waste materials that would otherwise end up in landfills.

Beyond environmental advantages, geopolymers demonstrate competitive mechanical performance. Rahman (2021) found that with a proper mix design optimizing alkali activator concentrations and SiO_2/Al_2O_3 ratios, GPC can achieve 28-day compressive strengths comparable to or exceeding those of conventional OPC-based concrete. For example, ambient-cured fly ash-based GPC can attain strengths of 40 MPa, equivalent to M40-grade OPC concrete. While heat curing (50-

80°C) accelerates early strength development, making geopolymers suitable for precast applications, recent advancements in slag-rich or calcium-enhanced formulations have enabled high-strength development under ambient conditions.

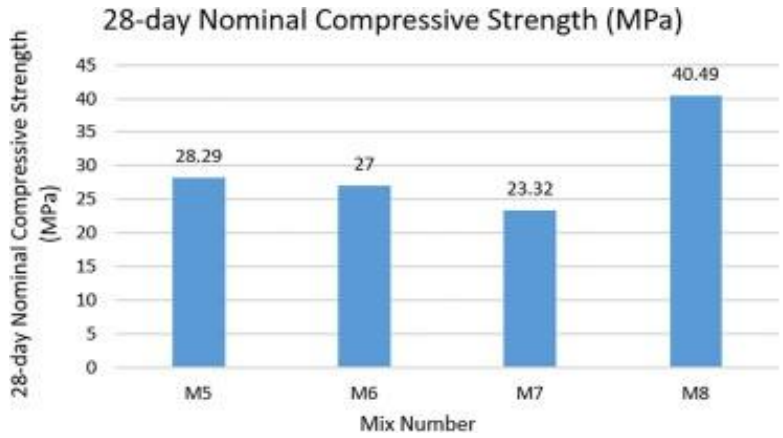


Figure 2.1 28d Compressive strength of SCGC specimens (Rahman, 2021)

A key distinction between geopolymers and OPC is in their chemical composition. As Zaidi et al. (2019) observed, geopolymer matrices lack portlandite [$Ca(OH)_2$], a weak and acid-soluble phase present in hydrated OPC pastes. This absence enhances the chemical durability of geopolymers, as discussed in Section 2.3. Furthermore, GPCs exhibit a refined pore structure dominated by gel pores, with fewer large capillary pores compared to OPC, particularly in low-calcium systems. This microstructure reduces permeability and sorptivity, improving resistance to aggressive agents in harsh environments such as marine exposure.

GPC exhibits considerable variability in properties due to differences in mix design and raw material composition. The calcium content, in particular, plays a critical role in determining the microstructure and durability performance of geopolymers. High-calcium systems, such as those incorporating GGBFS or Class C fly ash tend to form hydration products like calcium (alumino)silicate hydrate (C-(A)-S-H) gel and hydrocalumite alongside the geopolymer gel network. This results in a hybrid reaction mechanism that resembles Portland cement systems, potentially influencing durability behavior. Conversely, low-calcium aluminosilicate-based geopolymers, such as those derived from Class F fly ash or MK activated with sodium silicate

produce a predominantly geopolymeric gel structure with distinct chemical characteristics. Chen *et al*. (2021) highlighted that calcium content significantly influences durability mechanisms in GPC. While calcium-rich formulations (e.g., slag-modified geopolymers) enhance early-age strength and matrix densification, they may also introduce vulnerabilities in chemically aggressive environments. This is attributed to the formation of calcium-based phases such as C-S-H and ettringite, which are more prone to degradation under sulfate or acid attack.

Therefore, optimizing calcium content is crucial to balancing mechanical performance and long-term durability. A common strategy involves blending fly ash with slag to achieve synergistic benefits. For instance, Alomayri et al. (2022) demonstrated that incorporating 15-30% slag into a fly ash-based geopolymer improved compressive strength and reduced water absorption while maintaining sulfate resistance. Similarly, Deb (2013) observed that an 80/20 fly ash/slag geopolymer exhibited minimal strength loss and no significant cracking after prolonged exposure to sodium sulfate, confirming its superior sulfate resistance compared to OPC systems. Overall, well-formulated geopolymers—particularly those with dense microstructures and low permeability—demonstrate excellent durability. The following sections examine the specific challenges of marine environments and compare the performance of GPC with conventional OPC-based systems.

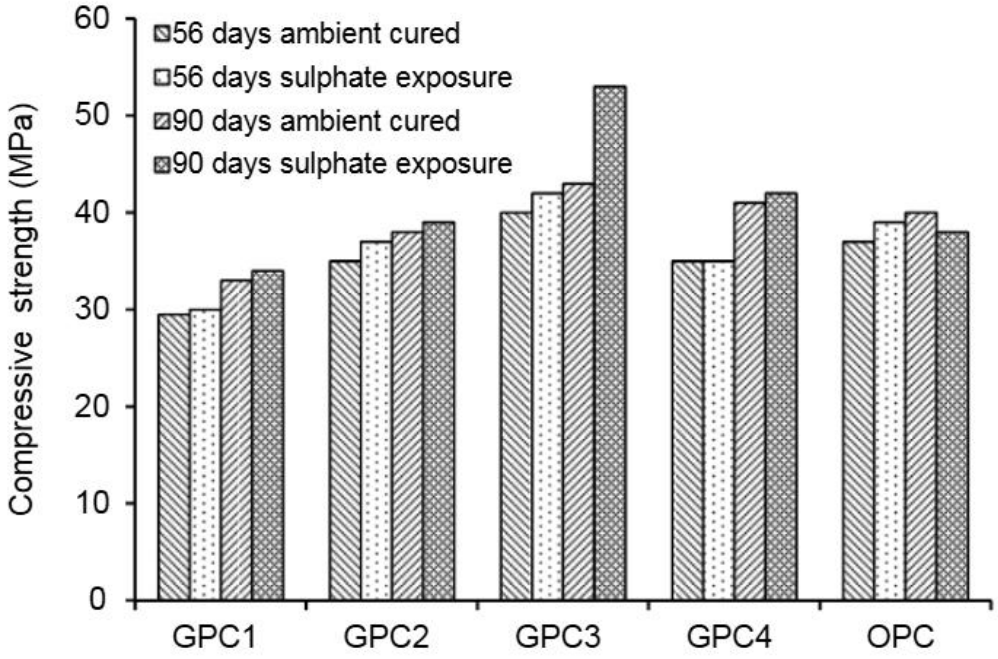


Figure 1.2 Change in Compressive strength of GPC and OPC concrete specimens immersed in 5% sodium sulphate solution (Deb, 2013)

2.3 Durability Challenges in Marine Environments

Marine environments present some of the harshest conditions for concrete structures due to the combined effects of multiple deterioration mechanisms. The primary durability challenges include chloride-induced reinforcement corrosion and sulfate attack, both of which degrade the structural integrity of concrete. Chloride penetration leads to steel reinforcement corrosion, while sulfate reactions disrupt the cementitious matrix. Furthermore, marine-exposed structures are subject to cyclic wetting and drying, salt crystallization, and, in some regions, significant temperature variations—all of which exacerbate deterioration processes. Rahman and Al-Ameri (2023) categorized marine exposure into four distinct zones: atmospheric, splash, tidal, and subaqueous. Among these, coastal saline soils often exhibit sulfate concentrations exceeding 2000 ppm, posing a severe durability risk. Structures in tidal and splash zones are particularly susceptible due to their simultaneous exposure to chloride ingress and cyclic wetting-drying conditions. Lin et al. (2010) demonstrated through numerical modeling that dry-wet cycles significantly accelerate the penetration of corrosive agents by combining diffusion and convection mechanisms. Their findings indicate that deterioration rates in tidal zones can be 3 to

10 times higher than in fully submerged regions. As a result, durability design and service life predictions for reinforced concrete must prioritize these critical exposure zones. In recent years, geopolymer and alkali-activated binders have been increasingly considered for marine and coastal infrastructure, particularly for precast elements, repair materials, and structures requiring reduced carbon footprint and improved chemical stability. Nonetheless, the extent to which geopolymer concretes can consistently deliver superior long-term durability in real marine environments remains an open question. Existing evidence is dominated by accelerated laboratory studies and short- to medium-term simulations, while long-duration field performance data are still limited. Moreover, marine exposure is inherently multi-ionic and cyclic, and the durability response of geopolymer binders varies with precursor type, calcium content, curing quality, and pore network characteristics. These uncertainties are particularly relevant in tidal and splash zones, where wetting-drying cycling can intensify salt crystallisation pressure and ion accumulation. Therefore, a mechanism-informed assessment of geopolymer durability under controlled marine-like exposure regimes remains necessary. This section reviews the dominant degradation mechanisms in marine environments and summarises reported performance trends of OPC and GPC under chloride and sulfate exposures, with emphasis on the critical tidal and splash zones.

2.3.1 Chloride Ingress and Steel Corrosion

Chloride-induced corrosion is the most pervasive durability issue for reinforced concrete in marine exposure. Chloride ions (predominantly from sodium chloride in seawater) penetrate the concrete cover and, once a threshold concentration reaches the steel rebar surface, they break down the passive oxide layer protecting the steel. This leads to active corrosion of steel, producing rust that expands and cracks the surrounding concrete. OPC concrete, being moderately porous and containing ionic pore solution, allows chlorides to gradually migrate inward under concentration gradients. The rate of chloride ingress depends on the concrete's permeability, pore structure, and the presence of binding phases. OPC paste does have some chloride-binding capability (through alumina in C_3A forming Friedel's salt), but a significant portion of chlorides remain free and contribute to corrosion risk.

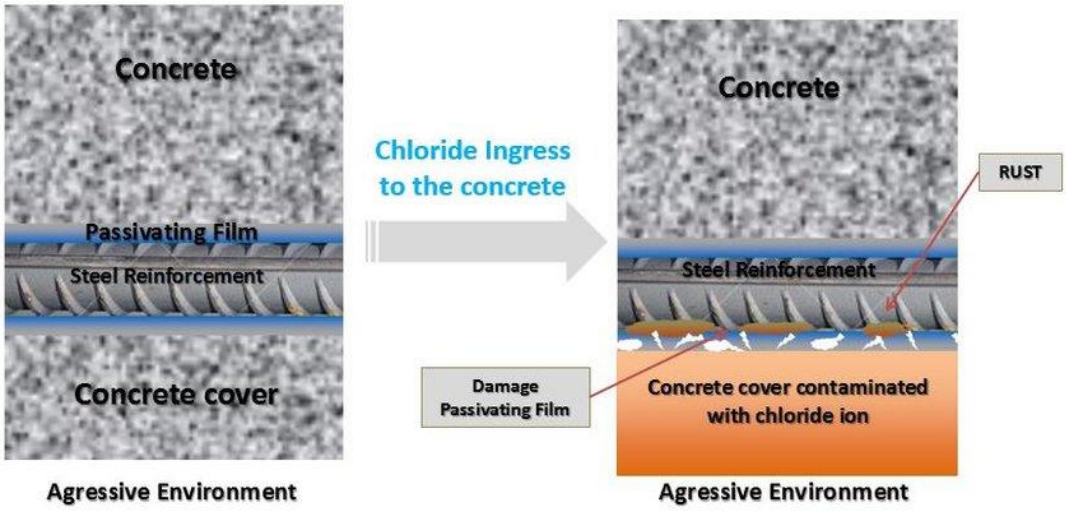


Figure 2.3 Corrosion Process Induced by Chloride Ingress (W Sutrisno, 2017)

GPC's behavior with respect to chlorides has been found to be equal or better than OPC in many studies. Geopolymers typically exhibit lower chloride ion penetrability in standard tests. Ganesan et al. (2015) compared OPC with GPC in rapid chloride permeability tests (ASTM C1202) and found the GPC had a total charge passed of ~1321 coulombs, whereas OPC concrete passed ~1764 coulombs - both in the "low" range, but GPC was clearly lower. The lower charge indicates that the geopolymer was less porous or its pores were less continuous for ion transport. In the same study, steel fiber-reinforced versions were tested and geopolymer still outperformed OPC in resisting chloride penetration. The microstructure of geopolymer (especially low-Ca ones) - a dense aluminosilicate gel - inherently limits chloride diffusion compared to OPC's C-S-H which has larger gel pores.

Table 2.1 Results of Bulk Diffusion Test and RCPT (Ganesan, 2015)

Mix	Depth of chloride penetration (cm)	Diffusion coefficient (m^2/s)	Charge passed (C)	Chloride ion penetrability as per ASTM
GPC	2.45	1.24×10^{-11}	1321	Low
SFRGPC1	2.42	1.21×10^{-11}	1445	Low
SFRGPC2	2.39	1.18×10^{-11}	1392	Low

Mix	Depth of chloride penetration (cm)	Diffusion coefficient (m^2/s)	Charge passed (C)	Chloride ion penetrability as per ASTM
SFRGPC3	2.36	1.15×10^{-11}	1566	Low
SFRGPC4	2.22	1.02×10^{-11}	1762	Low
CC	2.49	1.28×10^{-11}	1764	Low
SFRC2	2.47	1.26×10^{-11}	1423	Low

Additionally, research indicates that geopolymers possess a degree of chloride immobilization capability through chemical interactions. Amorim et al. (2021) investigated the corrosion rates of reinforcing steel in OPC and MK-based GPC, observing significantly lower corrosion currents in the geopolymer specimens. This reduction was attributed to the geopolymer matrix's enhanced capacity to immobilize chlorides, likely through adsorption or binding within the aluminosilicate gel structure. These findings suggest GPC offers superior protection to steel reinforcement compared to OPC, supporting the hypothesis that its reduced chloride permeability diminishes corrosion risk. However, the chloride-binding mechanisms in geopolymers differ from those in OPC. While OPC relies on calcium aluminate phases to form Friedel's salt, geopolymers, particularly those without supplementary slag lack these phases. Nevertheless, in slag-containing geopolymers, limited Friedel's salt formation may occur. Additionally, chlorides may be physically adsorbed within the gel structure or react with alkali constituents, further contributing to immobilization. Further research is necessary to fully elucidate the precise mechanisms governing chloride binding in geopolymer systems.

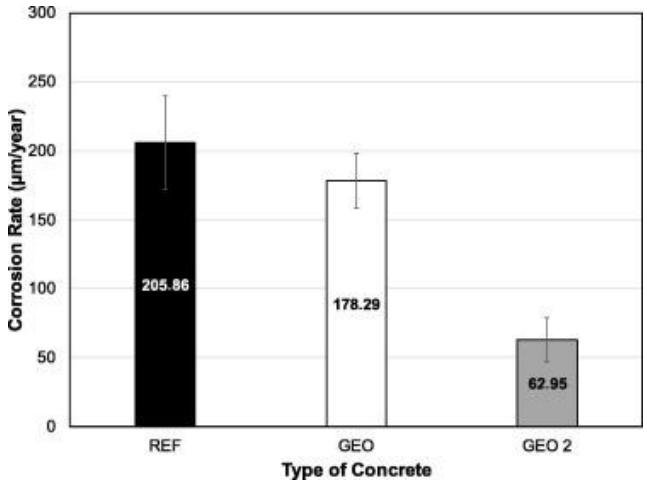


Figure 2.4 Corrosion rate of steel rebars inserted into the samples of geopolymer and conventional concrete, measured after the corrosion potential test. (Amorim, 2021)

Mohseni et al. (2019) observed that GPC structures exposed to marine atmospheres or splash zones may exhibit a delayed onset of corrosion initiation relative to OPC counterparts, assuming equivalent cover depths and strength grades. Their experimental investigation demonstrated that under simulated marine splash conditions, geopolymer mortar specimens exhibited negligible steel mass loss and significantly lower chloride ion penetration depths compared to OPC specimens. While GPCs are not entirely impervious to chloride ingress, particularly in cases of inadequate curing or elevated porosity—their resistance to chloride-induced corrosion is generally comparable to or exceeds that of conventional OPC-based concrete. The study further revealed that the incorporation of supplementary cementitious materials or discrete fibers can enhance durability performance. Specifically, the addition of up to 1% polypropylene fibers was found to reduce porosity and decrease chloride permeability in lightweight geopolymer composites. Additionally, the use of dense aggregate matrices and optimized water-to-binder ratios contributed to further improvements in chloride resistance. These findings suggest that material composition and mix design play a critical role in the long-term durability of geopolymer concretes in chloride-laden environments.

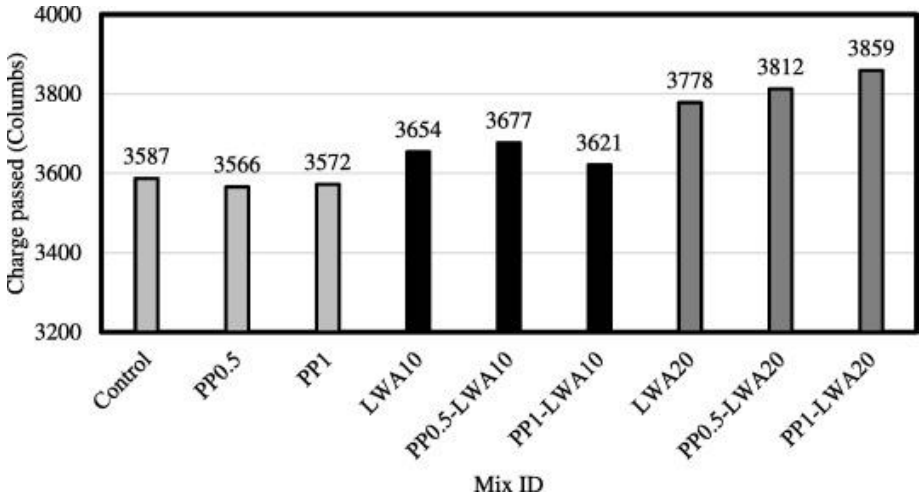


Figure 2.5 Rapid chloride permeability test results of geopolymer mixtures with the variation of PP fiber and LWA percentages
(Mohseni, 2019)

It is also important to consider the differences in pore solution chemistry between GPC and OPC systems. OPC concrete typically exhibits a highly alkaline pore solution (pH ~13-14), primarily due to the presence of NaOH, KOH, and $Ca(OH)_2$ (portlandite), which promotes chloride binding but also facilitates pitting corrosion once chlorides reach the reinforcement. In contrast, GPC possess a distinct pore solution composition, characterized by high Na^+ concentrations but often lower free OH^- levels upon reaction completion, as alkalis become partially incorporated into the geopolymer gel structure. Amorim *et al*. (2021) suggested that the inherent absence of portlandite and the unique ionic composition of geopolymer binders may contribute to slower chloride ion diffusion rates and a reduced availability of free chlorides at the steel-concrete interface, thereby delaying corrosion initiation. Furthermore, while carbonation in OPC systems leads to a substantial decrease in pore solution pH increasing the risk of steel depassivation even at moderate chloride concentrations—geopolymers undergo a different carbonation mechanism. Due to the lack of portlandite, carbonation in geopolymers primarily results in the formation of sodium carbonates without a significant pH reduction. This distinct reaction pathway suggests an enhanced durability performance of GPC in marine environments, where simultaneous exposure to carbon dioxide and chlorides is

common. Consequently, GPC may offer superior long-term corrosion resistance compared to OPC-based systems under such aggressive conditions.

2.3.2 Sulfate Attack and Wetting-Drying Cycles

Sulfate attack is a major durability concern for concrete structures in marine and coastal environments, particularly in regions with high sulfate concentrations, such as tidal flats or groundwater contaminated with seawater-derived sulfates. When sulfate ions (e.g., SO_4^{2-} from Na_2SO_4 or $MgSO_4$) penetrate the concrete matrix, they react with hydration products, primarily in OPC based systems. The most susceptible phases are calcium hydroxide (CH) and monosulfate (AFm), which undergo chemical reactions with sulfates to form gypsum ($CaSO_4 \cdot 2H_2O$) and ettringite ($Ca_6Al_2(SO_4)_3(OH)_{12} \cdot 26H_2O$, or AFt). The crystallization of these reaction products within the pore structure induces expansive stresses, leading to microcracking, spalling, surface scaling, and a gradual loss of cohesion. Notably, magnesium sulfate ($MgSO_4$) is more aggressive than sodium sulfate (Na_2SO_4) due to the additional degradation mechanism involving Mg^{2+} ions, which decalcify the C-S-H gel, forming non-cementitious phases such as brucite ($Mg(OH)_2$) and gypsum, thereby weakening the binder.

Extensive studies have documented substantial deterioration in the mechanical properties of OPC based materials in sulfate-rich environments. Bassuoni and Nehdi (2009) attributed this degradation to the cyclic hydration and dehydration of sulfate salts, which induce repeated volumetric changes and progressive damage. Wang *et al*. (2020) employed microstructural characterization techniques, including scanning electron microscopy (SEM) and X-ray diffraction (XRD), to confirm that sulfate attack predominantly results in the formation of ettringite and gypsum. Their findings indicate that these expansive phases can increase local volume by approximately 1.2-2.2 times, generating internal stresses that promote cracking and spalling, ultimately compromising the structural integrity of the concrete.

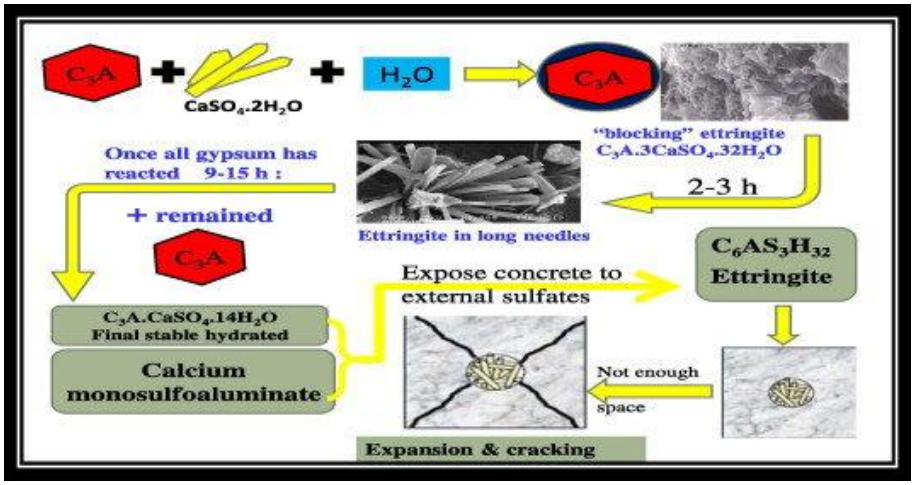


Figure 2.6 Mechanism of sulfate attack (Hodhod, 2013)

The pore structure and hydration products of geopolymers are quite different from those of cementitious materials due to differences in raw materials and reaction mechanisms. The reaction products of geopolymers are mainly high alkalinity gel products, zeolites, feldspars and nepheline-like structures. These products form dense three-dimensional network microstructures that give geopolymers unique chemical and physical properties, as well as excellent permeability and corrosion resistance. In addition, as a new type of cementitious material, the hydration products and pore structure of GPC are very different from those of traditional cement-based materials. GPC is cured by chemical reaction of silica-alumina materials, and the resulting hydration products mainly exist in the form of N-A-S-H and C-A-S-H. This difference suggests that the reaction mechanisms and performance of GPC in sulfate-aggressive environments may differ from those of conventional cement-based materials, warranting further study.

A comparative study on the strength performance of Portland cement and slag-based GPC has been conducted. Meng et al. (2024) reported that after 12 months of exposure to 5.0% Na_2SO_4 solution, the compressive strength of Portland cement concrete decreased by 25.0%, whereas slag GPC exhibited a lower strength loss of 17.0%. In a 5.0% $MgSO_4$ solution, the reductions were more severe, with Portland cement losing 37.0% of its strength compared to 23.0% for the slag geopolymer. Song

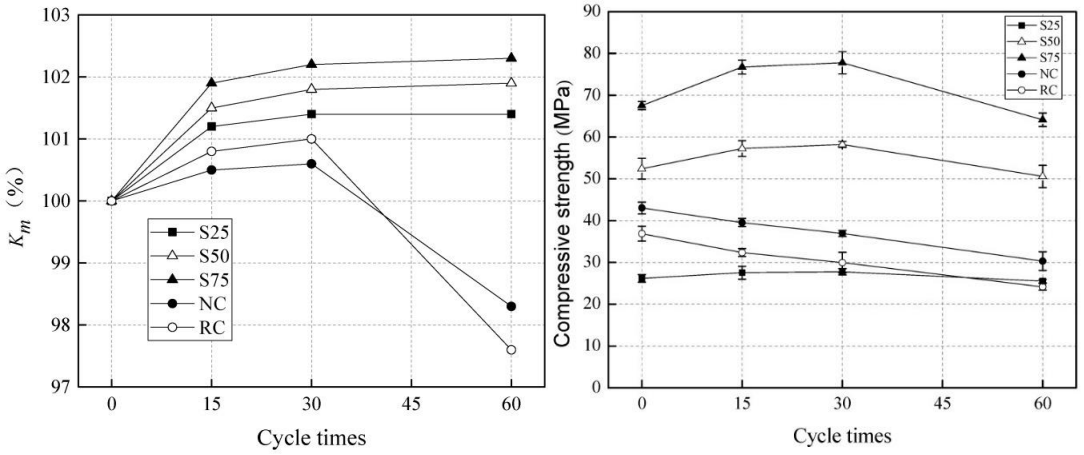
et al. (2005) further noted that Na_2SO_4 exposure can enhance polymerization reactions in geopolymer systems, thereby improving structural compactness. Clarke *et al*. (2007) observed that replacing tap water with seawater in slag geopolymer slurry formulations increased the bound water content and produced more crystalline reaction products, leading to a denser microstructure and improved resistance to seawater erosion. Overall, slag-based geopolymers demonstrate superior durability due to their dense matrix and corrosion-resistant reaction products.

GPCs, particularly those with low calcium content, exhibit superior sulfate resistance compared to OPC. This enhanced durability is primarily attributed to the absence or minimal presence of calcium aluminates, which significantly reduces the formation of expansive ettringite a major cause of degradation in OPC-based materials. Additionally, the dominant binding phase in geopolymers, the sodium aluminosilicate hydrate (N-A-S-H) gel, demonstrates remarkable stability in sulfate-rich environments. Beyond its chemical inertness, the N-A-S-H gel forms a dense microstructure that acts as a physical barrier, impeding sulfate ion penetration.

A compelling demonstration of this performance difference was provided by Yu *et al*. (2015), who compared geopolymer-stabilized soil with cement-stabilized soil under sulfate exposure. After 90 days of immersion in a 2.5% Na_2SO_4 solution, the cement-stabilized samples exhibited severe deterioration, including extensive cracking, spalling, and ettringite-filled voids. In contrast, slag-fly ash geopolymer-stabilized samples maintained structural integrity, showing only minor edge erosion and retaining a smooth surface. Notably, the optimal geopolymer mix retained approximately 87% of its strength after prolonged sulfate exposure, whereas the OPC-based specimens had failed completely. The authors attributed this resilience to the N-A-S-H gel's ability to restrict sulfate ingress and minimize reactions with vulnerable phases.

Further supporting evidence comes from Xie *et al* . (2019), who investigated RAC with a geopolymer binder (GGBFS + fly ash). Their findings revealed that the geopolymer RAC experienced negligible strength loss or mass change under sulfate

attack, whereas OPC-based RAC suffered significant expansion and cracking. The study concluded that geopolymer binders provide excellent sulfate resistance relative to conventional cement. This improved performance stems from a combination of chemical and microstructural advantages. Chemically, geopolymers lack reactive phases such as $Ca(OH)_2$ and C_3A, which in OPC produce deleterious expansion products. Physically, the dense, low-permeability matrix of geopolymers slows sulfate diffusion. Furthermore, the alkali-rich pore solution in geopolymers promotes the formation of less disruptive sodium sulfate salts (e.g., thenardite or mirabilite), unlike the destructive ettringite formation typical in OPC. These findings collectively highlight geopolymers as a durable alternative to OPC in sulfate-prone environments, offering both long-term stability and reduced maintenance demands.



(a) (b)

Figure 2.7 Influence of sulfate-attack cycles on (a) mass loss and (b) compressive strength of specimens (Jianhe Xie, 2019)

While high-calcium geopolymers—particularly those with a significant slag component—exhibit superior sulfate resistance compared to OPC, they are not entirely immune to sulfate-related degradation. When the geopolymer mix contains substantial calcium (derived from slag or cement additions), hydration byproducts such as C-A-S-H gel or calcium aluminosilicate hydrates may react with sulfates. Studies have documented minor ettringite or gypsum formation in slag-rich geopolymers exposed to sulfate environments. However, the resulting damage is typically far less severe than in OPC systems, primarily because the geopolymer matrix confines these crystalline phases within its gel pores, preventing disruptive

expansion. Alternatively, the limited quantity of these phases restricts their detrimental effects.

For example, Ariffin *et al*. (2022) observed ettringite deposition in fly ash/slag geopolymer mortar subjected to sulfuric acid, which creates a sulfate-bearing environment. Notably, while OPC mortar experienced significant strength loss, the geopolymer mortar retained most of its strength, as the binding gel structure remained intact despite crystal formation. These findings underscore that calcium content plays a critical role in sulfate resistance: low-calcium geopolymer systems demonstrate near-complete immunity to sulfate attack, whereas high-calcium (blended) formulations, while still highly resistant, may exhibit minor vulnerabilities under aggressive conditions.

Further supporting this, Chen *et al*. (2021) emphasize that calcium content influences the durability mechanisms of GPC. While higher calcium levels enhance mechanical properties, they may also introduce a degree of susceptibility to classical sulfate attack, particularly if substantial ettringite formation occurs. Thus, while high-calcium geopolymers generally outperform OPC in sulfate environments, their resistance is not absolute, and mix design optimization remains essential for long-term durability.

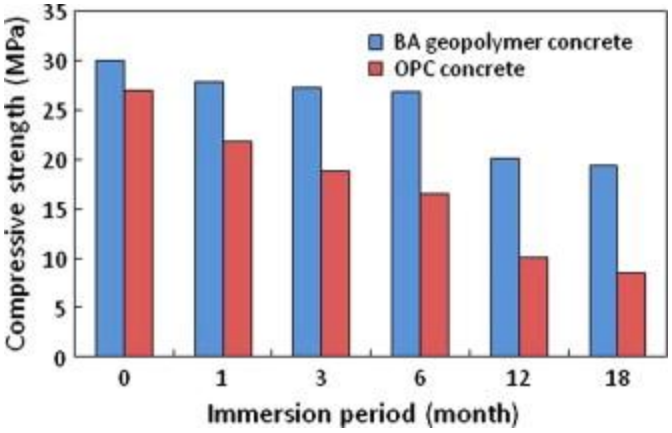


Figure 2.8 Compressive strength of concrete specimens exposed to sulfuric acid for 18 months (Ariffin, 2019)

Structures in the tidal or splash zone are subjected to periodic wetting by seawater followed by drying during low tide or solar radiation. These wetting-drying (W-D) cycles can amplify damage caused by both chlorides and sulfates. During the drying phase, salts present in the concrete pores crystallize, and the resulting crystallization pressure generates mechanical stress that widens pores and propagates cracks. Over time, repeated W-D exposure leads to progressive surface scaling, even in the absence of substantial chemical reactions. Additionally, fluctuations in internal moisture cause repeated expansion and contraction, which can fatigue the concrete matrix. In the presence of sulfate, W-D cycling further accelerates chemical deterioration. Drying concentrates sulfates within the pore structure, enhancing their interaction with hydration products, while rewetting replenishes the supply of aggressive ions. This cyclic concentration-dilution process produces more severe damage over fewer cycles than continuous immersion. OPC mortar may remain relatively intact after six months of continuous sulfate immersion, yet suffer substantial degradation after just 30 W-D cycles due to salt crystallization and alternating ettringite formation.

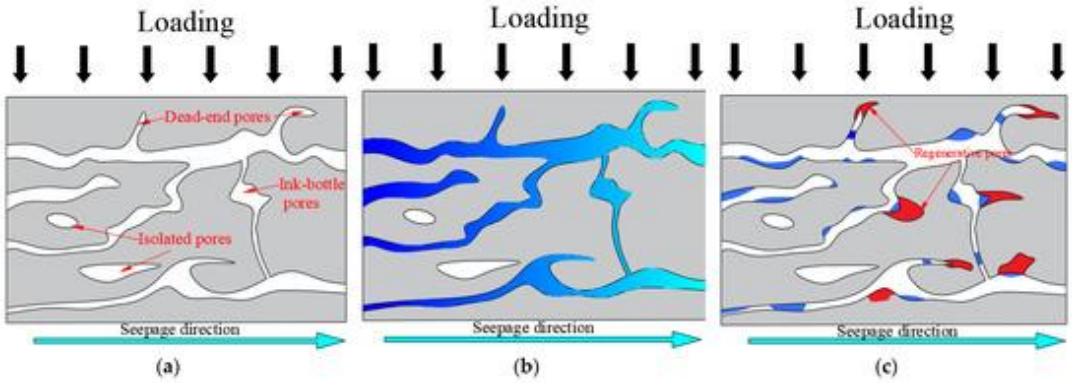


Figure 2.9 Concrete pore regeneration process: (a) initial diagram, (b) saturation diagram, and (c) naturally drying diagram (Wang, 2022)

Horpibulsuk *et al*. (2016) investigated the durability of geopolymer materials under wet-dry cycles in sulfate-rich environments and found that geopolymer systems outperformed their cement-based counterparts. In a comparative study of clay-fly ash geopolymer bricks and clay-cement bricks, the geopolymer specimens demonstrated negligible changes in microstructure and pH even after prolonged wet-dry exposure, while the cement-based bricks exhibited visible deterioration. Notably, the geopolymer samples showed no significant expansion or cracking, suggesting a

substantially longer service life under cyclic sulfate attack. Guo and Xiong (2021) investigated the performance of fiber-reinforced fly ash/steel-slag based geopolymer mortars subjected to drying-wetting cycles and observed that the majority of mass loss occurred during the initial cycles, likely due to early-stage leaching of residual alkalis. However, after approximately 15 cycles, the mass stabilized, and the specimens retained their structural integrity. In contrast, control samples made of OPC typically exhibited progressive mass gain due to gypsum formation, followed by surface cracking and subsequent mass loss as material delaminated. The superior resistance of geopolymers to internal damage from salt crystallization is attributed to their inherently low permeability, which restricts salt ingress, and possibly to a limited degree of self-healing or gel structure reorganization during cyclic exposure.

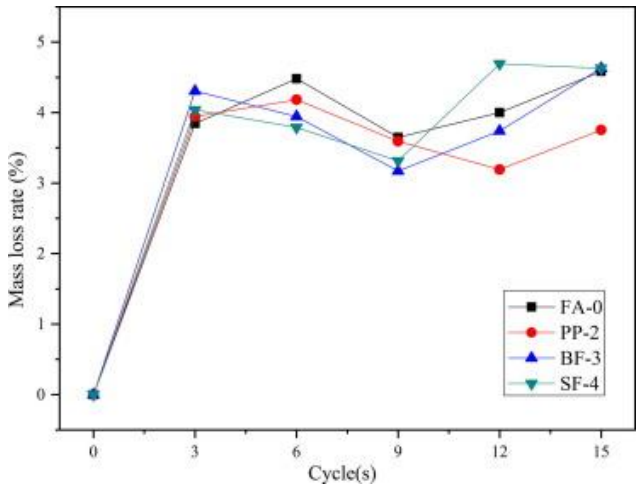


Figure 2.10 Mass loss rate of geopolymer mortar under double factor coupling (Guo, 2021)

In summary, marine durability requires concrete to resist chloride ingress and sulfate attack, often under cyclic wetting-drying exposure. OPC concrete is challenged on both fronts, as chlorides promote reinforcement corrosion while sulfates and salt crystallisation can induce cracking and surface scaling, especially in tidal and splash zones. Many laboratory studies report that geopolymer concrete can exhibit lower ion penetrability and improved stability compared with OPC under specific marine-simulated conditions, owing to its distinct binder chemistry and generally denser pore structure. Nevertheless, it is necessary to interpret these findings with caution. Real marine environments involve interacting factors such as combined chloride-sulfate

actions, temperature and humidity fluctuations, and mechanical loading. Long-term field evidence for geopolymer systems remains relatively limited, and performance is sensitive to mix design, curing quality, and pore connectivity. Poorly formulated mixtures with high porosity or incomplete geopolymerisation may not achieve the expected durability. These considerations motivate systematic research under controlled exposure regimes. Accordingly, the present thesis investigates a carefully designed MK-slag geopolymer system under cyclic wet-dry sulfate exposure and continuous immersion and provides experimental evidence to support the development of predictive models for performance evolution.

2.3.3 Role of Fiber Reinforcement in Geopolymer Durability

GPC exhibits similar brittle characteristics to conventional OPC, including low tensile strength, susceptibility to cracking, and a tendency toward failure during the damage process. To address these limitations, the incorporation of fibers into concrete has become a well-established method to improve mechanical performance, particularly in terms of tensile strength, ductility, and crack resistance. Fiber reinforcement is an effective strategy for improving the ductility of GPC by reducing its brittleness. Nagaraju *et al*. (2023) demonstrated that incorporating synthetic PP and polyolefin fibers into fly ash-based GPC significantly improves its creep and shrinkage performance. Liu *et al*. (2024) further found that the addition of fibers effectively inhibits the initiation and propagation of microcracks, thereby substantially reducing the adverse effects of drying and shrinkage at various stages.

In the context of durability, fibers further contribute by controlling crack growth, this reduction in crack size and quantity limits pathways for the ingress of deleterious substances such as chlorides and sulfates. As a result, fiber-reinforced concrete often shows improved performance under mechanical loads and environmental stress due to this crack-bridging effect. Among available fiber types, basalt fiber has emerged as an ideal material for GPC modification due to its high modulus of elasticity, good alkali resistance, low cost, and simple manufacturing process. Furtos *et al*. (2024) demonstrated that GPC specimens incorporating basalt fibers exhibit higher strengths compared to those without fiber reinforcement, making them particularly suitable for enhancing durability under marine-like exposure conditions.

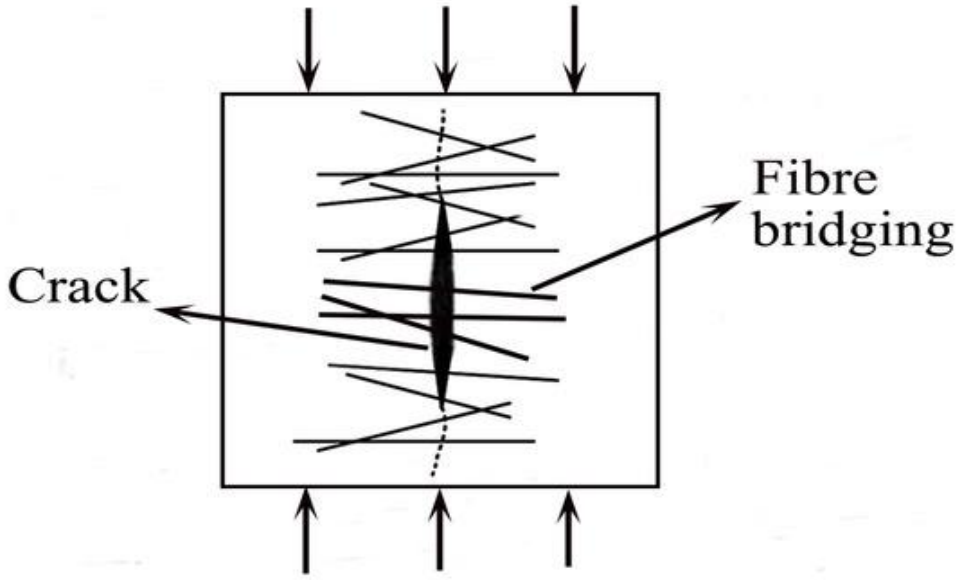
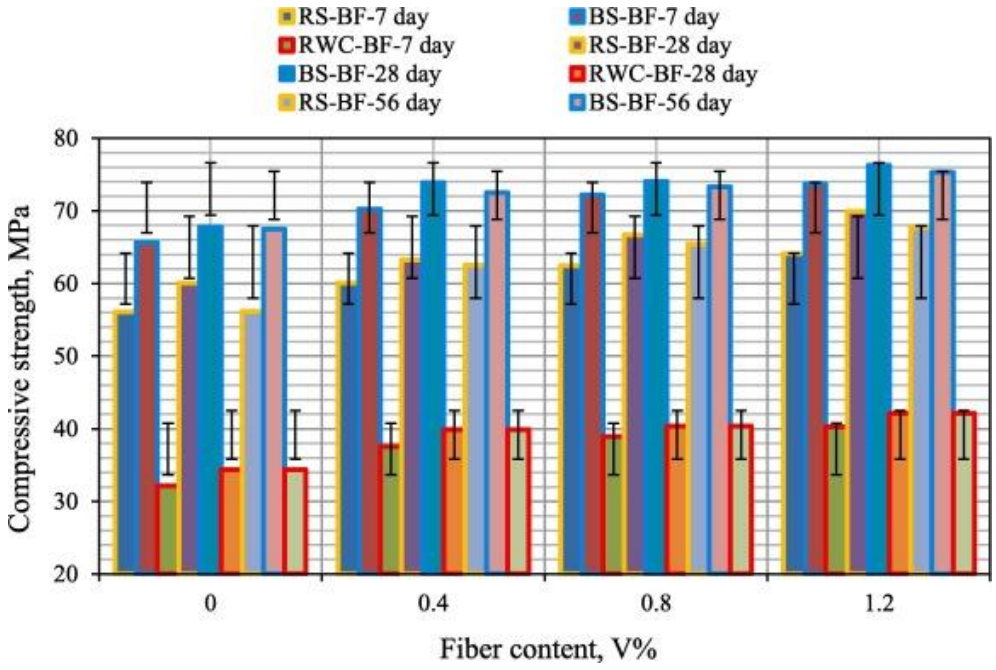


Figure 2.11 Bridging mechanism of fiber reinforcements (Jiang, 2024)

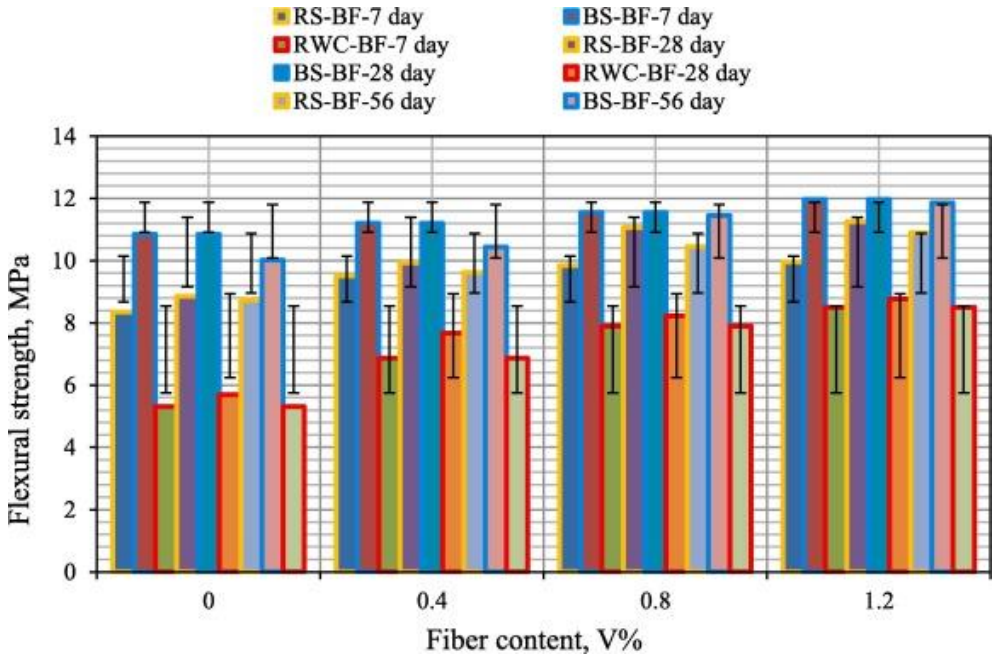
Basalt fiber is relatively new high-performance fiber, derived from the extrusion of molten basalt rock. It shares some similarities with glass fiber but generally has higher strength and is more alkali resistant. Basalt fibers have a tensile strength in the range of 2000-4800 MPa and a modulus of 70-90 GPa, making them suitable for reinforcing brittle matrices. They are also stable up to ~700-800 °C, giving them advantages in fire resistance compared to polymeric fibers. In cementitious systems, basalt fibers have shown promise in enhancing toughness, reducing cracking, and improving impact resistance.

Mechanical improvements resulting from the incorporation of basalt fibers into MK-based geopolymer mortars have been demonstrated in recent research. Şahin *et al*. (2021) showed that adding basalt fibers in the range of 0-1.2% by volume led to significant gains in both compressive and flexural strength. Specifically, compressive strength increased by up to 25%, while flexural strength improved by as much as 50%. These enhancements were most pronounced at optimal fiber contents around 0.5-1.0%, beyond which the performance gains tended to plateau or decline due to reduced workability and potential fiber agglomeration. The improved performance is primarily attributed to the crack-bridging effect of basalt fibers, which carry tensile stress across microcracks and inhibit their propagation. This mechanism increases the

strain capacity of the material and enhances its post-cracking toughness. The uniform dispersion of short basalt fibers within the geopolymer matrix also contributes to the formation of a multidirectional reinforcement network, which improves overall mechanical resilience and durability.



(a)



(b)

Figure 2.12 Effect of basalt fiber content on the mechanical properties of geopolymer specimens (a) compressive strength and (b) flexural strength

(Şahin, 2021)

Drying shrinkage is a critical factor affecting the durability of concrete, as it can induce microcracks that promote the penetration of aggressive agents. In GPCs—particularly those activated with high-alkali solutions—drying and autogenous shrinkage can be comparable to or even exceed that of OPC if not properly mitigated. Research by Yavana *et al*. (2022) demonstrated that incorporating PP fibers effectively reduces drying shrinkage by distributing internal strains and suppressing crack formation. Similarly, basalt fibers, owing to their higher stiffness, are expected to mitigate shrinkage by restraining matrix movement and limiting crack width. By minimizing shrinkage-induced cracking, basalt fibers help maintain a more monolithic structure, thereby enhancing long-term durability through reduced permeability.

Chemical resistance is another crucial aspect of durability. The inclusion of fibers can alter the degradation behavior of concrete under chemically aggressive conditions. In OPC systems, a key concern with basalt fibers is their potential degradation in highly alkaline environments (pH > 12.5) over extended periods. However, geopolymer matrices generally exhibit a less aggressive pore solution chemistry compared to OPC, depending on the alkali activator used. Although NaOH-activated geopolymers may initially have high pH levels (~13), their long-term chemical environment is more stable and less calcium-rich than OPC paste. Studies indicate that basalt fibers exhibit only minor tensile property reductions after prolonged exposure to alkaline solutions, demonstrating superior alkali resistance compared to conventional glass fibers.

This is supported by accelerated aging tests conducted by Wu *et al*. (2015) on basalt fiber-reinforced polymer (BFRP) bars immersed in hydroxide-rich solutions revealed that while an initial strength reduction occurred, the fibers retained over 60% of their original tensile strength after nine weeks at 55°C. Notably, the degradation rate significantly decreased after the initial exposure phase, underscoring the fibers' inherent resilience in alkaline conditions. Critically, even if long-term degradation occurs, basalt fibers primarily enhance early-stage durability by bridging microcracks and controlling crack propagation during initial environmental exposure. This crack-

restraining effect reduces the ingress of harmful agents, thereby improving the overall durability of the concrete matrix.

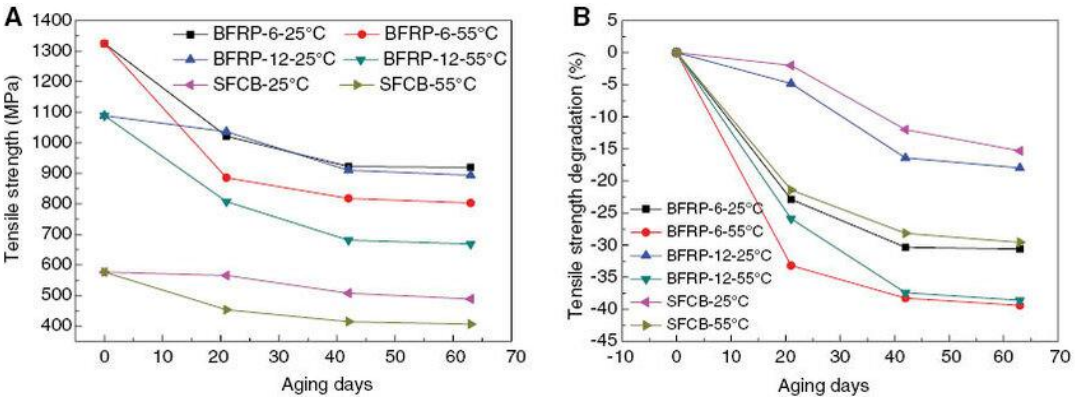


Figure 2.13 Tensile strength degradation of different FRP bars: (A) tensile strength degradation (B) tensile strength retention (Gang Wu, 2015)

Experimental studies demonstrate that fiber reinforcement significantly improves the durability of geopolymers by reducing permeability, mitigating crack propagation, and enhancing resistance to chemical attack. Mohseni *et al*. (2018) observed that incorporating PP fibers in lightweight geopolymer concrete (containing rice husk ash and nano-alumina) decreased chloride permeability (as indicated by reduced charge passed in RCPT) and improved acid resistance due to lower porosity and effective crack control. While their study focused on PP fibers, basalt fibers—with their superior mechanical properties—are expected to provide similar or greater permeation resistance.

Sulfate resistance is also enhanced by fiber reinforcement. Kantarci (2022) found that an optimal PP fiber dosage (0.5% vol) in fly ash-based GPC significantly reduced compressive strength loss after $MgSO_4$ exposure. SEM analysis revealed that fibers bridged microcracks, preventing their expansion under sulfate-induced stress. Given their higher modulus, basalt fibers should perform even more effectively in sulfate-rich environments. Supporting this, Langaroudi *et al*. (2025) reported that basalt fiber-reinforced geopolymer mortars (0.6% vol) exhibited reduced sulfate penetration depth and retained higher residual strength after prolonged exposure. The crack-bridging and stress-distribution capabilities of basalt fibers restrict microcrack

propagation, limiting aggressive ion ingress and accommodating expansive ettringite formation.

In marine environments, where chlorides and sulfates coexist, basalt fibers enhance durability by improving mechanical resilience and reducing crack formation. Since ion penetration is diffusion-controlled, maintaining a dense, crack-free microstructure slows deterioration. Unlike steel fibers, basalt fibers are corrosion-resistant, ensuring long-term reinforcement in saline conditions. Rahman *et al*. (2023) confirmed this, demonstrating that BFRGC retained ~87% of its load capacity after 12 months of accelerated tidal cycling and thermal exposure (50°C), with minimal chloride ingress and no visible cracking. These findings highlight the synergistic benefits of combining geopolymer binders with basalt fibers for durable marine structures.

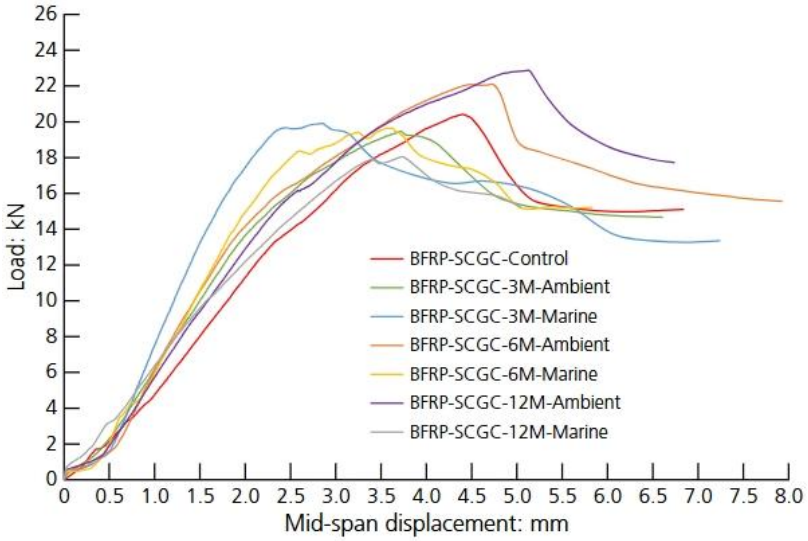


Figure 2.14 Load against mid-span displacement plot for 28 days,3 months, 6 months and 12 months BFRP-SCGC beam specimens (Rahman, 2023)

The length and content of fibers must be carefully optimized to ensure performance benefits without compromising workability. Excessive fiber content can reduce workability, leading to poor compaction and voids that diminish durability. Similarly, longer fibers are prone to clumping and uneven distribution. Kantarci (2022) found that 6 mm PP fibers improved sulfate resistance more effectively than 12 mm fibers due to their superior dispersion. While basalt fibers are available in lengths ranging

from 6 to 24 mm, a moderate length of 12 mm—combined with proper mixing techniques—ensures uniform distribution. Fiber content should typically remain between 0.5% and 1% by volume, as higher dosages may impair workability with diminishing returns. In this thesis, a 1 % basalt fiber content was selected to optimize crack control without sacrificing mix performance.

Basalt fiber reinforcement enhances the durability of GPC through multiple, interacting mechanisms: it raises resistance to crack initiation and propagation—thereby reducing permeability—improves tensile capacity under mechanical and thermal actions, and mitigates shrinkage-induced cracking. These mechanical gains complement the intrinsic chemical robustness of geopolymer binders, making the composite well suited to marine exposure. Empirical studies consistently report durability improvements in fiber-reinforced geopolymers; for example, the incorporation of waste fibers has been shown to markedly reduce chloride penetration, further strengthening an already durable binder system.

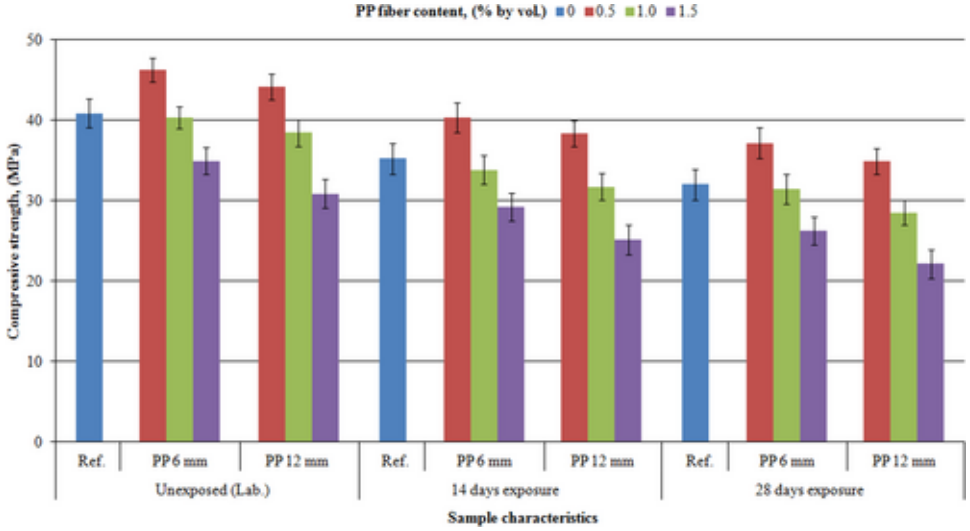


Figure 2.15 Compressive strength of GPC samples before and after sulfate exposure (Kantarci, 2022)

Despite the growing interest in BFRGC, systematic investigations into their durability under combined marine and sulfate exposure remain limited. Moreover, many existing studies do not fully decouple the fiber-related crack-control effect from coupled influences of matrix composition, curing quality, and exposure regime,

which can also govern durability evolution. This thesis addresses this research gap by evaluating the potential of basalt fibers to enhance the sulfate resistance of GPC subjected to wet-dry cycling and prolonged immersion. The hypothesis suggests that incorporating basalt fibers can mitigate strength degradation primarily through crack control and improved damage tolerance under the investigated exposure conditions. The experimental findings presented in Chapter 3 will critically evaluate this hypothesis and provide insights into the degradation mechanisms of BFRGC in aggressive environments.

2.4 Machine Learning Applications in Concrete and Geopolymer Research

Recent advances in ML and data-driven modelling have significantly enhanced the ability to analyse and optimise concrete performance. In civil engineering materials research, ML techniques have been increasingly employed to predict material properties from mix composition and curing-related descriptors, and to learn nonlinear, high-dimensional relationships that are difficult to represent using conventional empirical equations. While many studies have demonstrated the effectiveness of ML in estimating mechanical and fresh properties of conventional concrete and GPC, the application of ML to time-dependent performance evolution under well-defined exposure protocols remains comparatively limited. In particular, modelling compressive-strength degradation or retention as a function of mix-design variables and exposure duration is challenging due to data scarcity, reporting inconsistency across studies, and the coupled nature of degradation processes.

This thesis addresses this gap by leveraging ML models to improve the prediction of compressive-strength degradation under a standardized sulfate-exposure scenario. To mitigate the limitations associated with sparse and imbalanced datasets, a CVAE-based approach is introduced for feature representation learning and data augmentation, with the aim of improving predictive robustness within the scope of the compiled dataset. By integrating advanced ML techniques with experimentally grounded performance evaluation, this work contributes to a more systematic and data-driven framework for analysing the performance evolution of GPC under sulfate exposure.

2.4.1 Predicting Concrete Properties from Mix Design using ML

Traditionally, concrete mix design and performance prediction have relied on empirical equations or graph-based methods. One of the most well-known is the Abrams water-cement ratio law proposed by Abrams (1918), which establishes an inverse relationship between compressive strength and the water-to-cement ratio:

$$S = \frac{A}{B^{(w/c)}} \tag{1.1}$$

where S is the compressive strength of concrete at a given age, w/c is the water-to-cement mass ratio, and A, B are empirical constants. As illustrated in Figure 2.16, both Abrams' law and Bolomey's formula demonstrate that concrete strength decreases sharply with increasing water-cement ratio, assuming proper compaction. While these classical models have provided a foundational understanding of concrete behavior, they are inherently limited to low-dimensional systems with conventional mix compositions and curing regimes. Moreover, they lack the capacity to generalize to modern concretes incorporating fibers, supplementary cementitious materials, or alternative binders such as geopolymers.

To address this limitation, recent research has investigated the application of ML techniques to model compressive strength without relying on predefined functional relationships. ML algorithms can autonomously discern complex patterns and capture nonlinear interactions among multiple input variables, rendering them particularly effective for high-dimensional systems. In this thesis, while Abrams' law provides a foundational benchmark, strength prediction is approached through a data-driven framework employing advanced ML methodologies. This approach facilitates more precise and adaptable modeling of GPC behavior under sulfate exposure.

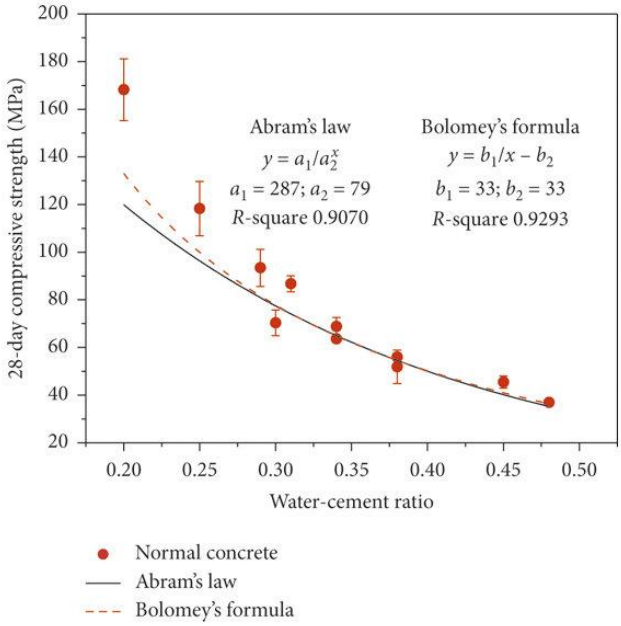


Figure 2.16 Simulations of Abrams' law and Bolomey's formula for normal concrete (Li, 2020)

Predicting the 28-day compressive strength of concrete based on mix proportions and curing conditions has been a prominent application of ML in civil engineering. Various ML techniques, including Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), decision tree-based ensembles (e.g., Random Forest, XGBoost), and genetic programming, have been employed to model the relationship between concrete ingredients (cement, water, aggregates, admixtures) and mechanical performance. These models are trained on historical datasets of mix designs and corresponding tested strengths, enabling them to predict the properties of new mixtures with high accuracy.

Compared to traditional regression-based formulas or code equations, ML models often demonstrate superior predictive performance due to their ability to capture nonlinear interactions among variables. For instance, they can effectively model synergistic effects, such as the combined influence of silica fume and superplasticizers on strength development. A seminal study by Yeh (1998) demonstrated the efficacy of ANNs using the UCI Concrete Compressive Strength Dataset, achieving a strong correlation between predicted and experimental results.

Subsequent research has further refined these models, enhancing their accuracy and generalizability.

The application of ML in GPC has gained increasing attention due to the material's complex chemistry and additional influencing factors. Unlike conventional concrete, geopolymers involve more variables, such as alkali activator concentration, curing temperature, and precursor composition (e.g., fly ash, slag). These complexities make GPC particularly suitable for data-driven modeling.

Rathnayaka *et al*. (2024) conducted a comprehensive review of ML approaches for predicting the compressive strength of fly ash-based GPC. Key input features identified include the SiO_2/Al_2O_3 ratio of the binder, NaOH molarity, activator-to-binder ratio, and curing conditions. The review emphasized that model performance heavily depends on feature selection and hyperparameter optimization. Ensemble methods and hybrid models have emerged as particularly effective, combining multiple algorithms to improve robustness.

For example, Wang *et al*. (2024) compared ANNs, Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and Gene Expression Programming (GEP) for strength prediction in fly ash/slag-based GPC, finding that GEP exhibited marginally better accuracy. Similarly, Dao *et al*. (2019) developed a hybrid ANFIS model optimized with particle swarm and genetic algorithms, achieving an R^2 exceeding 0.93 for GPC containing 100% recycled aggregates. These studies collectively demonstrate that ML can effectively model the intricate relationships between geopolymer composition, processing conditions, and mechanical properties, offering a valuable tool for optimizing mix designs.

Both conventional and GPC benefit from ML-based predictive modeling, which outperforms traditional empirical approaches by capturing complex, nonlinear relationships. While conventional concrete models focus on ingredient interactions, GPC models must additionally account for alkali activation and curing dynamics. Continued advancements in ensemble learning and hybrid techniques are expected to

further enhance predictive accuracy, supporting the development of high-performance and sustainable concrete mixtures.

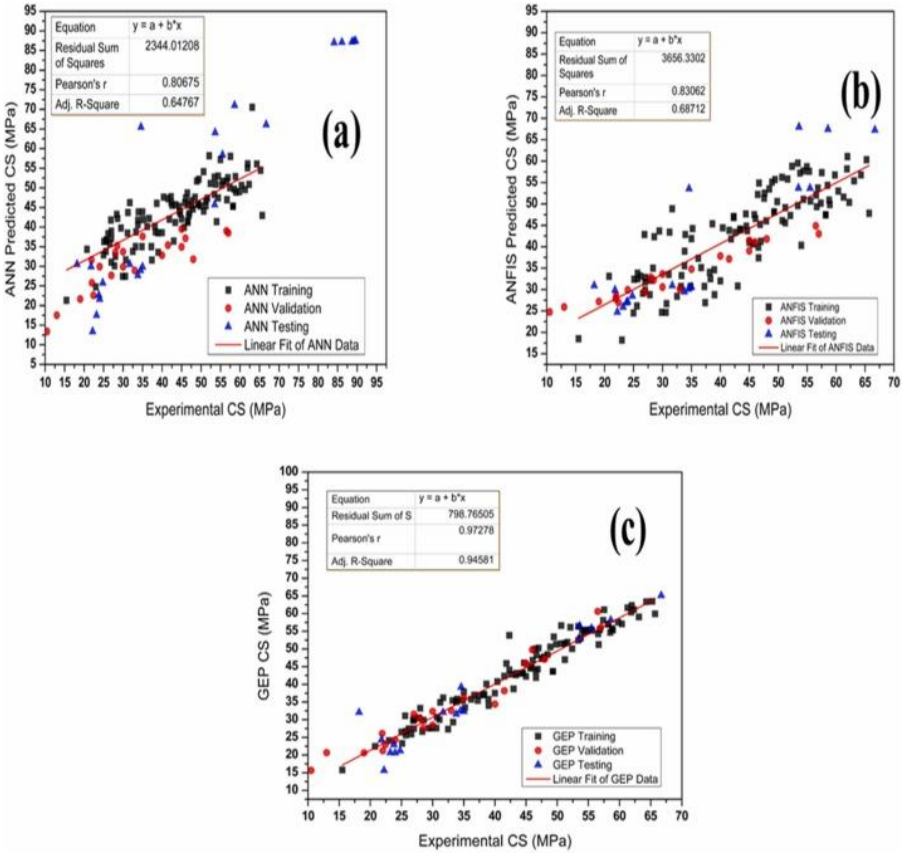


Figure 2.17 CS predicted: (a) ANN (b) ANFIS (c) GEP (Wang, 2024)

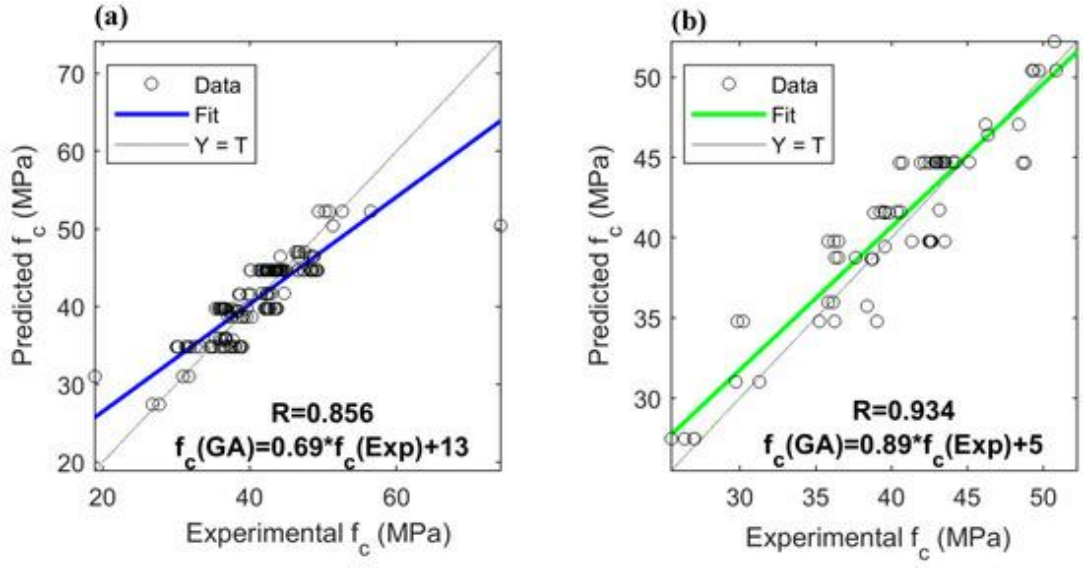


Figure 2.18 Correlation coefficients of different models: (a) training PSOANFIS, (b) testing PSOANFIS (Dao, 2019)

Beyond compressive strength, ML has been successfully applied to predict a wide range of fresh and hardened concrete properties, including workability (slump flow), setting time, elastic modulus, and durability indicators such as chloride diffusion coefficients. Alaneme *et al*. (2023) conducted a comprehensive review of AI techniques—including fuzzy logic and ANNs—for optimizing sustainable construction materials, particularly agro-waste-based geopolymers. Their findings highlight ML's potential in streamlining mix design optimization, curing condition selection, and material formulation to achieve desired performance characteristics while reducing environmental impact and experimental costs.

While most ML studies in concrete and geopolymers focus on mechanical properties, emerging research explores durability prediction. Krishna and Rao (2019) employed fuzzy logic to estimate both compressive strength and a durability indicator in GPC based on alkali activator molarity and mix composition. However, their definition of "durability" was limited, potentially referring to sorptivity or an inferred acid resistance metric derived from chemical ratios. This underscores the need for more rigorous and standardized durability assessment criteria in ML modeling.

The literature demonstrates that well-trained ML models can serve as highly accurate surrogate models, facilitating rapid virtual testing and multi-objective optimization. For instance, ML enables the identification of mix designs that maximize strength at a given cost or minimize carbon footprint while meeting specified performance thresholds. Such capabilities significantly reduce reliance on time-consuming and resource-intensive laboratory experimentation.

Despite these advantages, ML models are inherently data-dependent, meaning their accuracy and generalizability rely on the quality, quantity, and consistency of the training dataset. While conventional concrete benefits from extensive, well-documented datasets, geopolymer research often suffers from limited or inconsistent

data due to variations in precursor materials, curing regimes, and testing protocols. This data scarcity poses a major challenge for developing robust ML models in GPC applications.

ML has proven to be a powerful tool for predicting both mechanical and durability-related properties of conventional and GPC. Its ability to optimize mix designs and reduce experimental burdens makes it invaluable for sustainable material development. However, broader adoption, particularly in geopolymer research requires addressing data limitations through standardized testing protocols and expanded datasets. Future work should also focus on refining durability prediction models by incorporating well-defined performance metrics.

2.4.2 Machine Learning for Long-Term Performance and Degradation Modelling

While predicting the 28-day compressive strength of concrete is valuable, assessing its long-term performance—particularly under aggressive environmental conditions—is equally critical. This challenge is inherently more complex, as it involves time-dependent processes and often requires the coupling of chemical, physical, and mechanical phenomena. Traditional durability models, such as Fickian diffusion models for chloride ingress or empirical models for sulfate-induced expansion, have been widely used. However, these models often depend on numerous input parameters (e.g., diffusivity, threshold values, reaction rates) that are difficult to determine for emerging materials like geopolymers.

The application of ML to durability prediction remains an emerging field, though several notable studies have demonstrated its potential. Among these, Jin *et al*. (2025) investigated the use of ML to predict the residual compressive strength of concrete subjected to sulfate exposure. Their study compiled 591 data points from the literature, encompassing concrete specimens primarily OPC or blended cements—that underwent sulfate exposure for varying durations, with corresponding residual compressive strength measurements. The input variables included cement type, w/c

ratio, supplementary cementitious materials (e.g., fly ash or slag), sulfate concentration, exposure duration, and curing conditions.

Using these features, the authors developed three coupled ANN models, each trained via metaheuristic optimization algorithms, including particle swarm optimization. As illustrated in Figure 2.19, their best-performing model, an ANN optimized using the Sparrow Search Algorithm (SSA-BPNN) achieved superior predictive accuracy compared to a conventional backpropagation neural network. Notably, the study identified the initial 28-day compressive strength as the most influential factor in determining long-term residual strength after sulfate attack. This finding aligns with the well-established principle that denser, higher-strength matrices exhibit greater resistance to chemical deterioration.

This work underscores the potential of ML to enhance durability predictions, particularly when sufficient and well-curated datasets are available. Furthermore, the authors emphasize the utility of such models in supporting service life assessments under complex or aggressive environmental conditions. Their approach demonstrates that ML can complement traditional durability modeling, offering a data-driven pathway to overcome the limitations of conventional empirical and mechanistic methods.

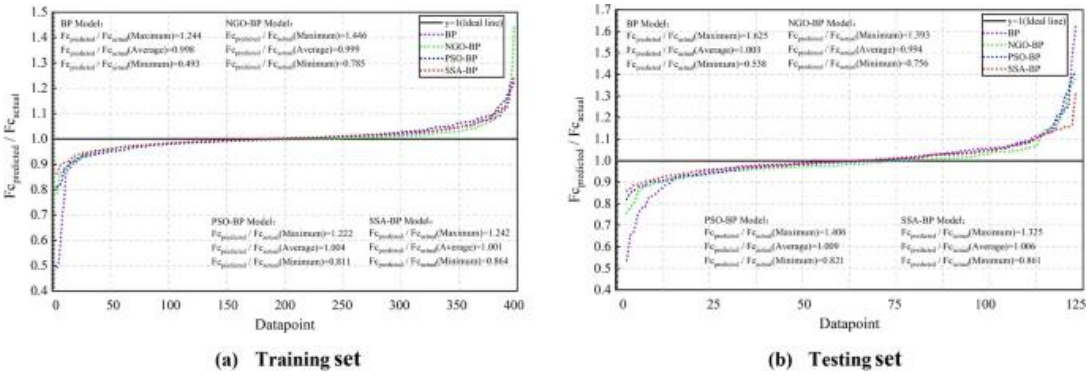


Figure 2.19 Performance comparison of four models (a) Training set (b) Testing set (Jin, 2025)

Benoît *et al*. (2023) - developed an interpretable ensemble ML model for predicting expansion of cement mortars under External Sulfate Attack (ESA). They combined

several tree-based models to predict expansion strain over time for mixes containing supplementary cementitious materials. Their model could predict the expansion behavior based on mix composition and was made interpretable by analyzing feature importance (e.g., showing that C_3A content and fly ash content had strong effects on reducing expansion). This kind of model could be extended to geopolymers to predict, say, length changes or cracking risk under sulfate exposure.

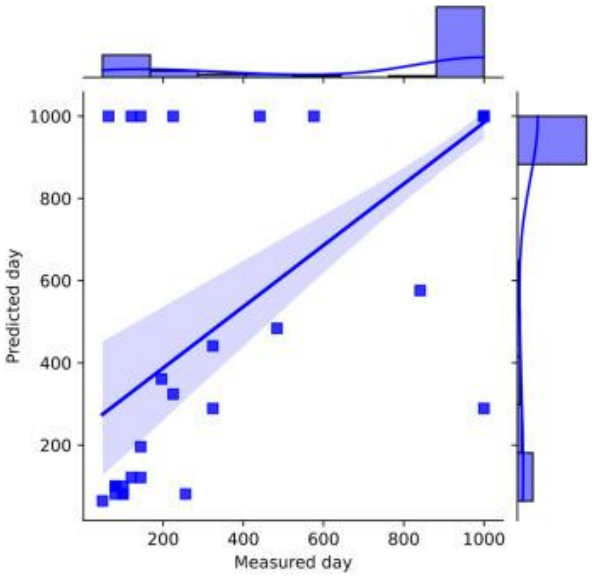


Figure 2.20 Comparison of XGB predicted and measured time needed to reach a 0.2% expansion with regression line and 95% confidence interval (Benoît, 2023)

Researchers have employed various ML techniques to predict chloride diffusion and corrosion performance in concrete. As illustrated in Figure 2.21, Zhou *et al*. (2023) utilized multilayer perceptron (MLP) and SVM models to estimate the chloride diffusion coefficient of supplementary cementitious material (SCM)-modified concrete, incorporating pore structure and mix design parameters. Their results demonstrated high predictive accuracy, highlighting the efficacy of these methods. Further advancing the field, Kechroud *et al*. (2025) demonstrated that ensemble methods, such as LightGBM, outperformed conventional modeling approaches in predicting chloride ion penetration. Their analysis identified the water-to-binder ratio and binder type as the most influential variables. These models can be applied either to directly estimate diffusion coefficients or to indirectly evaluate corrosion initiation time.

Additionally, advanced ML models have been developed to predict corrosion depth and rate under varying environmental exposure conditions, leveraging mix-related input features. Such models provide valuable tools for enhancing service life forecasting and durability assessments of concrete structures.

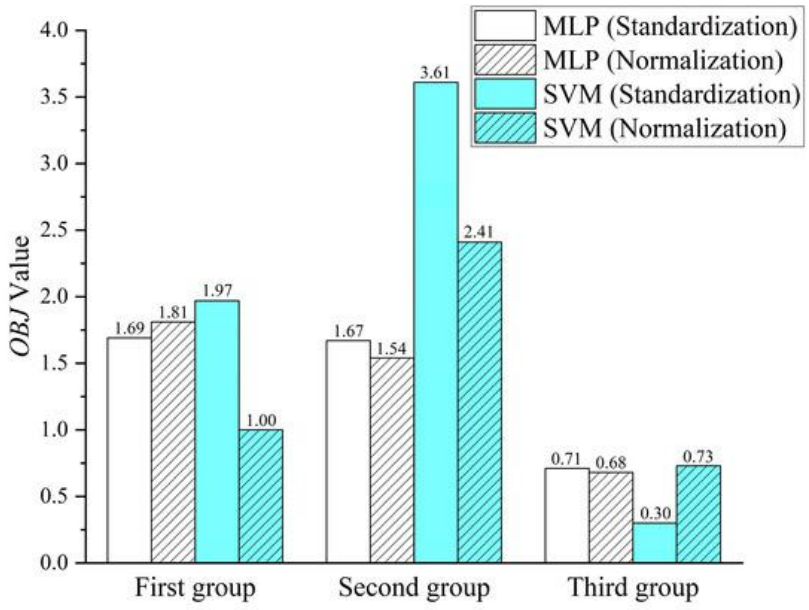


Figure 2.21 Comparisons of OBJ value obtained from MLP and SVM models with standardization and/or normalization preprocessing method (Zhou, 2023)

Carbonation depth prediction has increasingly been explored through ML approaches. Mehrdad *et al*. (2024) employed an ensemble of decision trees, random forest, and ANNs to model the carbonation depth of OPC, demonstrating strong predictive accuracy with a coefficient of determination (R^2) exceeding 0.99 when compared to experimental results. Similarly, Mansouri *et al*. (2025) utilized hybrid neural network ensembles to predict carbonation depth, achieving high correlation by incorporating parameters related to curing conditions, compressive strength, and environmental exposure. While carbonation is generally less significant in geopolymers due to the lack of portlandite, the established data-driven methodologies could be adapted to geopolymer systems once comprehensive datasets become available.

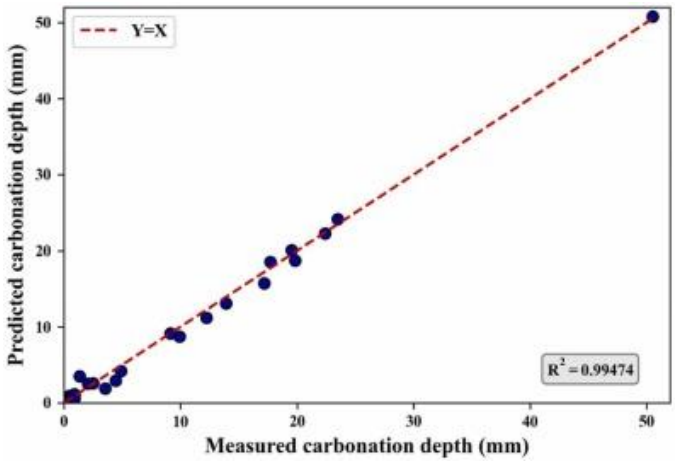


Figure 2.22 Scatter Plot of ANN Model for MOEA/D-ANN on Test Data. (Mehrdad, 2024)

However, within the specific context of GPC durability, there remains a notable paucity of research. Although Wong *et al*. (2022) conducted an extensive literature review on geopolymer durability, their work primarily focused on empirical studies and did not address ML based predictive approaches. This omission underscores the novelty of applying ML techniques in this domain. Unlike conventional concrete, geopolymers exhibit distinct durability-controlling factors, such as pore solution chemistry and the degree of polymerization, which are not adequately captured in existing datasets.

A critical gap in current research is the limited number of ML models that can represent the time-dependent evolution of GPC compressive strength under consistently defined sulfate-exposure conditions, particularly when data are collected from multiple sources with heterogeneous reporting practices. In response, this thesis develops a predictive framework that uses mix-design variables and exposure duration as inputs to estimate compressive strength at different ages under a standardized sulfate-immersion protocol. Rather than treating the problem as a single-point prediction (e.g., 28-day strength), the objective is to capture the evolution of strength over exposure time within the scope of the curated dataset.

An additional challenge is the scarcity of long-term data for geopolymer systems. To alleviate data sparsity and improve generalization in the downstream regressors, this

thesis introduces a CVAE-based strategy to enrich feature representation and support data augmentation in a manner consistent with the statistical structure of the compiled dataset. This methodology enhances model robustness and improves generalization capabilities.

2.4.3 CVAE for Data Augmentation and Enhanced Prediction

A Variational Autoencoder (VAE) is a deep learning model that learns a compressed latent representation of input data and reconstructs it to approximate the original distribution. The CVAE extends this framework by incorporating conditioning variables, allowing for controlled generation of data that adheres to specified input criteria. In this thesis, the conditioning variables consist of mix-design parameters and exposure duration, and the target output is compressive strength under sulfate exposure. The CVAE is utilised to learn a compact latent representation of the joint distribution of the input features and strength response, which can be leveraged to support feature enhancement and data augmentation for downstream regression. This approach enables the generation of realistic synthetic data that captures time-dependent deterioration under specific environmental and compositional conditions.

The key advantage of using a CVAE lies in its ability to mitigate data scarcity and improve the robustness of predictive models. When only limited experimental data are available, as is often the case with strength deterioration in geopolymer mixtures, traditional supervised learning methods risk overfitting may fail to generalize underlying deterioration trends. In contrast, a CVAE learns the joint probability distribution between input and output variables, enabling the generation of synthetic data that statistically aligns with the original dataset. These augmented samples can then be used to enhance downstream predictive models, either by integrating the CVAE directly into the modeling framework or by expanding the training dataset prior to model development.

Generative models have demonstrated promising potential in the design of concrete mixtures. For example, Ge *et al*. (2022) employed CVAE to generate novel concrete formulations with specified target properties. In their framework, the input conditions consisted of desired mechanical properties (e.g., compressive strength and slump)

along with sustainability constraints (e.g., reduced carbon footprint), while the output comprised optimized mixture proportions. The CVAE was trained on an existing dataset of concrete mixes, enabling it to propose new designs that satisfied the given constraints. Notably, the model-generated mixtures exhibited a 40-50% reduction in cement content—thereby lowering associated carbon emissions—while still meeting the required strength criteria upon experimental validation. This thesis underscores the efficacy of CVAEs in navigating complex material design spaces and generating viable solutions that extend beyond the bounds of the original training data.

This thesis employs CVAE to augment existing datasets by generating synthetic compressive strength data at untested exposure periods, rather than creating entirely new mix designs. By leveraging known mix proportions, the CVAE reconstructs continuous strength deterioration trajectories, inferring plausible values at intermediate time intervals (e.g., estimating 30- or 60-day strength from 0-, 90-, and 180-day experimental data). This approach smooths and densifies degradation curves while generalizing common deterioration patterns across different mixtures, even when long-term data are sparse. Consequently, it mitigates data discontinuity challenges in corrosion-related durability research and enhances the predictive robustness of ML models for GPC.

A key advantage of the CVAE is its ability to quantify predictive uncertainty. Unlike deterministic models that yield single-point estimates, the CVAE generates a distribution of possible outcomes, reflecting inherent material variability. The model's latent space encodes unobserved degradation factors, such as microstructural heterogeneity, enabling probabilistic sampling of strength values under identical input conditions. This feature is critical for durability assessments, where uncertainty significantly influences performance forecasting and design safety.

Moreover, data augmentation via CVAE improves deep learning regression performance by preserving underlying statistical relationships. Unlike random noise injection, which may distort data structures, the CVAE generates synthetic samples consistent with the learned joint distribution of input-output variables. Shmuel *et al*.

(2025) demonstrated that such advanced augmentation techniques can enhance model accuracy by over 10%, underscoring the efficacy of this approach.

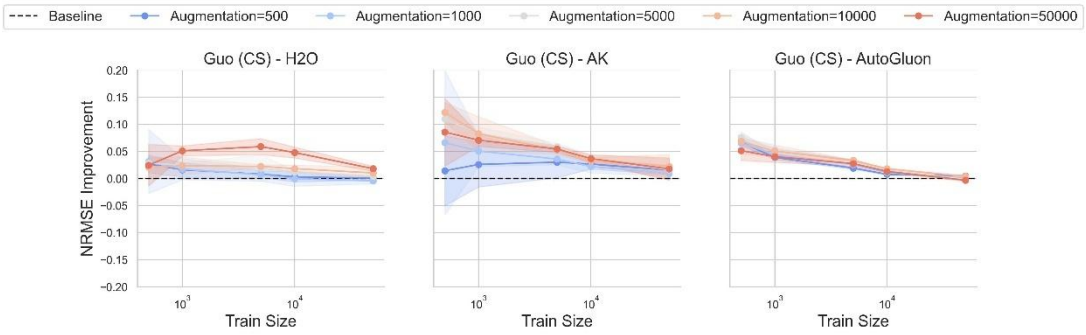


Figure 2.23 Relative RMSE improvement as a function of training size and number of augmented rows (Shmuel, 2024)

In this thesis, CVAE is employed to generate synthetic training data, enhancing the accuracy and generalizability of predictive models such as ANNs or ensemble methods. The CVAE is trained on a durability dataset, with input vectors comprising mix composition features and exposure time, and output corresponding to compressive strength. After training, the decoder can generate synthetic samples that are statistically consistent with the training distribution, supporting the enrichment of the training set (e.g., within the observed ranges of mix variables and exposure duration) and improving the robustness of downstream regression models. These augmented data can either supplement the training set of a separate predictive model or serve as a generative predictor via the decoder. This approach improves the modeling of GPC deterioration under aggressive environmental conditions.

The application of CVAE to model concrete durability presents a novel contribution to the field. While VAEs have been used in materials science for tasks such as generating hypothetical alloy compositions or polymer structures, their application in predicting time-dependent deterioration in cementitious materials remains limited. This thesis integrates mechanistic insights into geopolymer degradation with data-driven augmentation techniques, enhancing the robustness and generalizability of long-term performance predictions.

Model interpretability is important for engineering applications. Although deep-learning models can achieve high predictive accuracy, extracting meaningful insights into the influence of input variables is essential for practical use and for supporting mix-design decision-making. Prior studies have also emphasized the value of interpretability analyses in data-driven modeling of sulfate-exposed cementitious materials; for example, Jin et al. (2025) employed interpretability-oriented analysis and identified initial compressive strength as an influential factor for residual strength under sulfate exposure. Accordingly, this thesis applies SHAP-based interpretability analysis to examine how key mix-design variables reported in the compiled dataset (e.g., MK, GGBFS, fly ash, activator components, water content, aggregate contents, and exposure duration) influence the predicted compressive strength under sulfate immersion. The resulting importance rankings and directional effects are then discussed in relation to established understanding of geopolymer binder chemistry and strength evolution under sulfate exposure, thereby improving the transparency and physical plausibility of the ML-driven predictions.

2.5 Summary

Previous studies have provided valuable insights and established a solid foundation for the present research. Extensive experimental and computational investigations have demonstrated the favorable properties of GPC and the potential of ML techniques in predicting concrete properties. However, several critical research gaps and limitations remain, which this thesis seeks to address:

First, although numerous studies have examined the mechanical properties of GPC, there is still a paucity of systematic evidence regarding its long-term durability under marine-relevant exposure conditions. In particular, existing research has predominantly focused on early-age performance, whereas sustained assessments under sulfate-rich environments and cyclic exposure regimes remain comparatively limited.

Second, while fiber reinforcement is widely recognized as an effective strategy for crack control and damage tolerance enhancement, the specific influence of basalt

fibers on the durability-related performance evolution of GPC in marine-like exposure scenarios is still insufficiently explored. Current literature reports more frequently on polypropylene and other synthetic fibers, with limited quantitative evidence clarifying the role of basalt fibers in mitigating damage development under sulfate exposure.

Third, from a data-driven perspective, time-dependent prediction of strength degradation for GPC under sulfate exposure has been far less studied than conventional single-age strength prediction. Most existing ML frameworks focus on OPC systems or on estimating compressive strength at a fixed age, highlighting the need for approaches that can capture performance evolution over exposure duration, where strength retention can serve as one performance-related indicator.

Finally, limited and heterogeneous datasets remain a practical constraint for data-driven modeling of long-term exposure behavior. Advanced generative approaches, such as CVAEs, offer potential for learning complex feature representations and supporting data augmentation under controlled conditions, yet their application to time-dependent strength-degradation modeling of GPC remains scarce.

By addressing these gaps, this thesis aims to strengthen the experimental evidence base for GPC under marine sulfate exposure and to evaluate the feasibility of contemporary ML approaches for predicting time-dependent strength performance under standardized exposure conditions. The subsequent chapter presents the experimental methodology and key observations, providing a basis for the predictive modeling work in Chapters 4 and 5.

Chapter 3. EXPERIMENTAL INVESTIGATION OF GEOPOLYMER CONCRETE WITH AND WITHOUT BASALT FIBERS UNDER MARINE SULFATE EXPOSURE

3.1 Introduction

Marine environments present significant durability challenges for concrete structures due to the synergistic effects of sulfate attack, chloride ingress, and cyclic wetting-drying exposure. Conventional OPC concrete is susceptible to substantial degradation under such conditions, primarily due to the formation of expansive ettringite and gypsum. In contrast, GPC with its dense N-A-S-H/C-A-S-H gel microstructure and reduced permeability, demonstrates enhanced chemical resistance. However, its inherent brittleness can compromise its performance once cracking is initiated. Fiber reinforcement, particularly incorporating basalt fibers, has shown potential in enhancing crack resistance and mechanical properties. While extensive research has been conducted on fiber-reinforced OPC systems, limited studies have explored the long-term sulfate resistance of BFRGC under marine exposure conditions. This chapter examines the mechanical properties and porosity evolution of BFRGC in comparison to plain GPC when subjected to dry-wet cycles and prolonged immersion. Compression testing, UPV, and NMR analyses are employed to assess key durability indicators across varying exposure durations.

3.2 Experimental methods

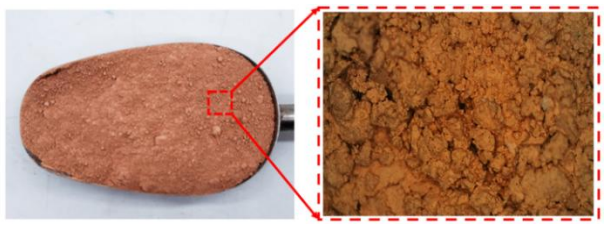
3.2.1 Materials

The constituent materials for BFRGC included MK, GGBFS, and basalt fibers; their mix proportions are detailed in Table 3.1. In this thesis, the MK used in the experiments was sourced from Henan Province, China. Figure 3.1(a) shows that the MK has a brick-red color, with particle sizes smaller than 13 μm and a uniform particle size distribution, with no visible agglomeration. The GGBFS used in this study is a byproduct of iron smelting, processed into industrial-grade solid waste. The selected GGBFS is classified as Grade S95, with a specific surface area of 800 m^2/g and an early strength index of 95 at 7 days. The density of the slag is 2.84 g/cm^3, with

an average particle size of 15.0 μm, meeting the requirements of ASTM C618. Figure 3.1(b) shows that the slag powder is silver-gray in appearance, with uniform particle sizes and no clumping phenomenon. The basalt fibers used in this study have a length of 9 mm, a diameter of 14.1 μm, a density of 2.69 g/cm^3, a tensile strength of 2.23 GPa, an elastic modulus of 85.8 GPa, and a fracture elongation of 2.9%. Figure 3.1(c) shows that the basalt fibers are slender, black to dark brown in color, with a smooth surface and a tough texture.

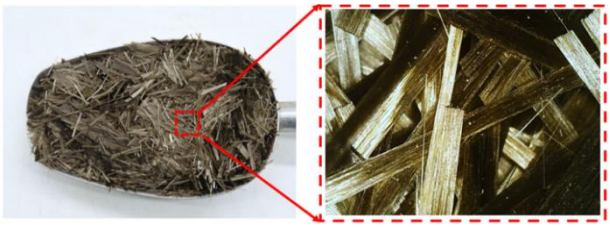
Table 3.1 Mix ratio design of the GPC.

Serial number	The percentage of BF (%)	Basalt fiber (kg/m^3)	Metakaolin (kg/m^3)	Slag (kg/m^3)	River sand (kg/m^3)	Limestone crushed stone (kg/m^3)	Sodium silicate solution (kg/m^3)
1	0.0	0.0	340.0	85.0	650.0	1300.0	337.0
2	1.0	26.4	340.0	85.0	650.0	1300.0	337.0



(a)

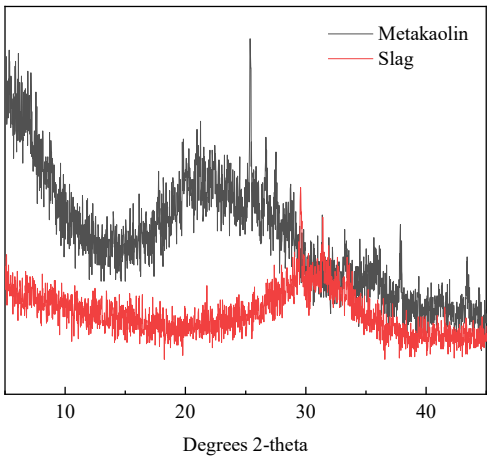
(b)



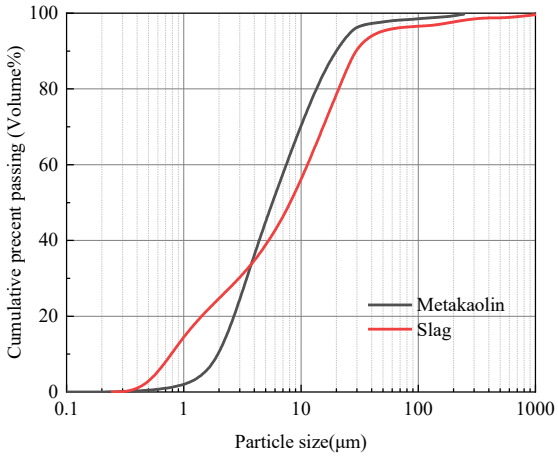
(c)

Figure 3.1 The materials for basalt fiber GPC. (a) MK. (b) slag. (c) basalt fiber.

Figure 3.2 (a) shows the XRD patterns of the MK and slag. The MK is nearly amorphous, exhibiting a broad hump feature in the 15° to 30° 2-theta range. This indicates that it is a highly reactive aluminosilicate cementitious material. Similarly, a broad hump feature in the 20° to 40° 2-theta range is observed in the XRD spectra of the slag. Figure 3.2 (b) shows the particle size distribution of MK and slag. As shown in Fig. 2(b), the MK exhibits a small particle distribution, which D_{10}, D_{50} and D_{90} are 2.23 µm, 5.81µm and 19.07 µm, respectively. Similarly, the D_{10}, D_{50} and D_{90} of slag are 0.62 µm, 8.47 µm and 29.26 µm, respectively.



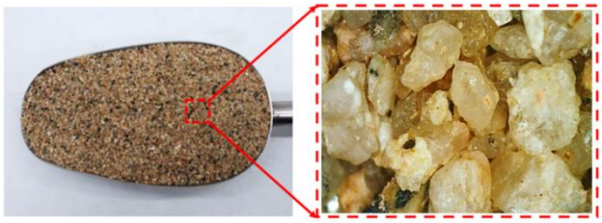
(a)



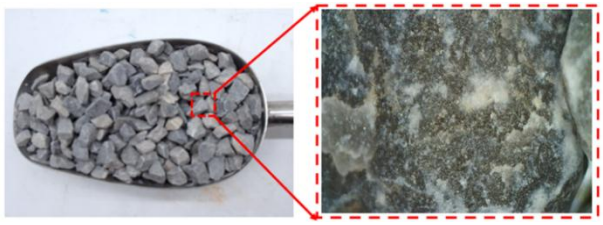
(b)

Figure 3.2 Materials characterization of MK and slag. (a) XRD pattern of MK and slag. (b) Particle size distribution of MK and slag.

Figure 3.3 shows the particle morphology of the aggregates used in this study. The aggregates employed for preparing BFRGC are river sand and limestone gravel. Figure 3.3 (a) shows the fine aggregate, which consists of river sand with a fineness modulus of 2.7. Figure 3.3 (b) shows the coarse aggregate, comprising limestone gravel with particle sizes ranging from 2.50 to 16.00 millimeters. The alkali activator consists of water glass, NaOH, and water. During the experimental process, water glass with a modulus of 3.24 was adjusted to a modulus of 1.3 by regulating the ratio of Na_2O to SiO_2, using an 8 mol/L NaOH solution to prepare the alkali activator.



(a)



(b)

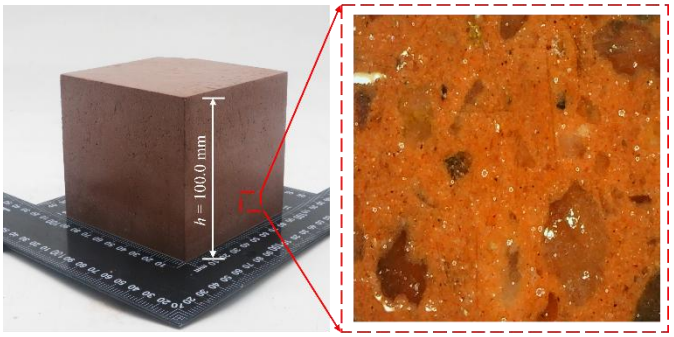
Figure 3.3 Aggregates for Basalt Fiber GPC. (a) River sand. (b) Limestone crushed stone.

3.2.2 Preparation of the specimens

To investigate the sulfate resistance of BFRGC, two simulated marine erosion regimes were established: continuous immersion and cyclic wetting-drying for a comparative study of the mechanical properties between BFRGC and GPC without fiber. Utilizing a 5.0% Na_2SO_4 aqueous solution as the simulated environment, as determined by ASTM C1012 standards, sulfate resistance experiments were conducted on GPC under immersion environment and dry-wet cycle conditions. After a 28-day curing period, specimens were subjected to testing at 0, 32, 64, 96, 128, 160, and 192 days of the erosion testing period to study the variations in static mechanical properties and pore characteristics of GPC without fiber and GPC with 1.0% basalt fiber content. For the immersion environment test, specimens were immersed in a 5.0%

Na_2SO_4 solution, with the solution replaced every 16 days to maintain a consistent ion concentration. Mechanical property and pore characteristic tests were conducted at the seven erosion testing periods (0, 32, 64, 96, 128, 160, and 192 days). The dry-wet cycle design employed in this experiment involved immersing the specimens in a 5.0% Na_2SO_4 solution for 48 hours, followed by drying at a constant temperature of 24°C for 48 hours, resulting in a 1:1 wet-dry ratio with a total cycle duration of 4 days. The solution was replaced every 16 days during the experiment. Mechanical property and pore characteristic tests were conducted at seven erosion testing periods (dry-wet cycle 0 days (0 cycles), 32 days (8 cycles), 64 days (16 cycles), 96 days (24 cycles), 128 days (32 cycles), 128 days (40 cycles), and 192 days (48 cycles)).

This experiment involved casting two types of specimens to investigate the mechanical properties and pore characteristics of GPC and BFRGC. Mechanical property tests were conducted using cubic specimens with dimensions of 100.0 mm × 100.0 mm × 100.0 mm, as shown in Figure 3.4 (a) and Figure 3.4 (b). Additionally, cylindrical specimens with a diameter of 50.0 mm and a height of 100.0 mm were prepared for LF-NMR tests, as shown in Figure 3.5 (a) and Figure 3.5 (b).



(a)

(b)

Figure 3.4 Cubic specimens of (a) GPC and (b) GPC with 1.0% addition of basalt fiber.

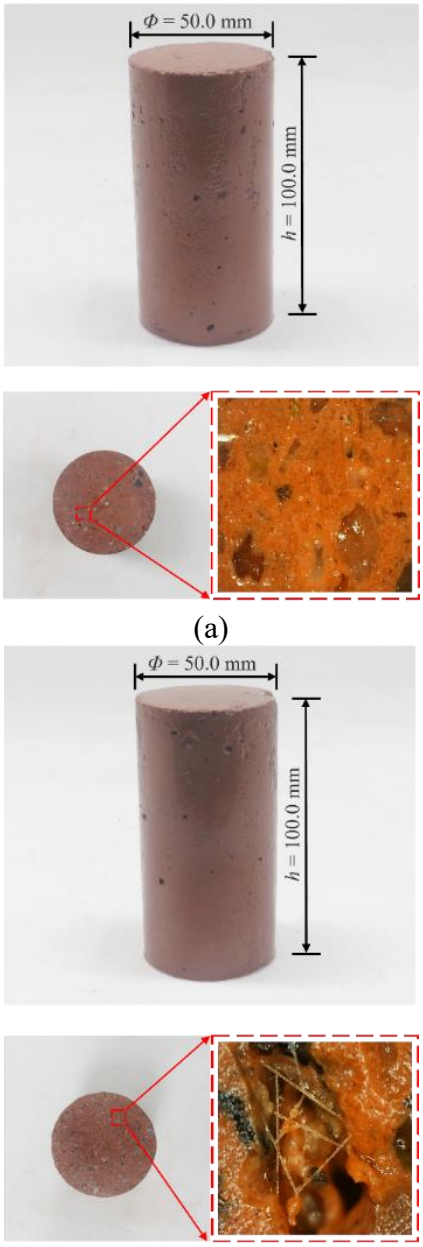


Figure 3.5 Cylindrical specimens of (a) GPC and (b) GPC with 1.0% addition of basalt fiber.

Figure 3.6 shows the specific preparation process of the specimens. Firstly, raw materials such as water, MK, GGBFS, basalt fibers, solid NaOH, river sand, and limestone aggregate were weighed according to the proportions. Secondly, the modulus ratio of Na_2SiO_3 was adjusted to 1.3 using an 8.0 mol/L NaOH solution. Subsequently, an alkaline activator was obtained by manual stirring and cooling to room temperature. Next, the basalt fibers (added in 1.0% BFRGC), MK, and GGBFS

were dry mixed in a mixing bowl for approximately 5.0 minutes to ensure thorough blending. Then, the alkaline activator was gradually added and stirred for about 5.0 minutes. Additionally, river sand and limestone aggregate were added and mixed until a homogeneous mixture was obtained. Finally, the fresh mixture was poured into molds and vibrated for 2.0 minutes on a vibrating table. The cast specimens were then sealed with plastic film and cured in a curing chamber at a temperature of 20.0±2.0°C and a humidity of 95.0% for 24.0 hours. After demolding, the specimens were further cured until the age of 28 days.

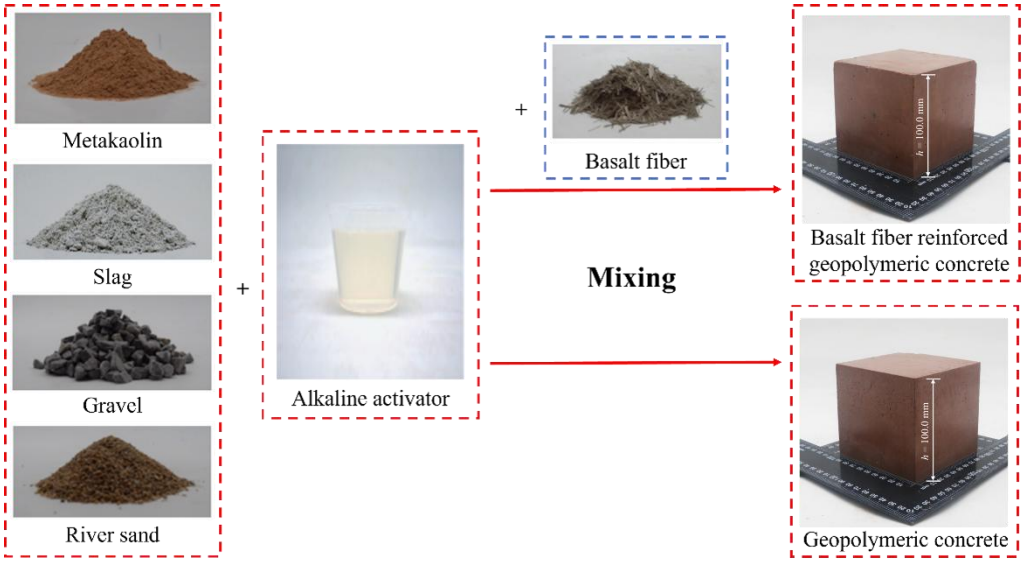


Figure 3.6 The preparation process of the specimens.

3.3. Methods

3.3.1 Density Change and Mass Loss Test

To track progressive changes, specimens were removed from the curing environment at 16-day intervals until the target test age was reached. After wiping the surfaces clean, the mass (m_d) and volume (V) of the specimens were measured to calculate the density changes and mass loss of GPC at different erosion testing periods. Observing and photographing the surface changes of the specimens provided a more detailed understanding of the surface deterioration of the GPC.

3.3.2 Ultrasonic pulse velocity test

Figure 3.7 shows the results of UPV tests conducted in accordance with ASTM C597 standards. The P-wave velocities of specimens with varying mix proportions were measured using the HS-YS2A ultrasonic velocity tester and ultrasonic pulse transmission technology. The samples were placed between the transmitter and receiver, with petroleum jelly applied to both ends to reduce wave attenuation. A consistent coupling force was applied to both sensors to ensure optimal acoustic contact with the specimens. The average longitudinal wave velocity of three specimens for each mix proportion was calculated to ensure result reliability.

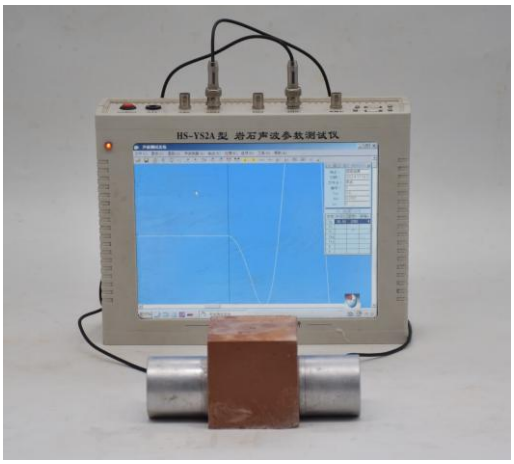


Figure 3.7 Testing the UPV with the HS-YS2A rock sonic tester.

3.3.3 Uniaxial compression test

Uniaxial compression tests were conducted on concrete cube specimens using an electro-hydraulic servo universal testing machine with a loading rate of 0.5 MPa/s. The displacement during the test was recorded by two dial gauges. The schematic diagram of the setup of the electro-hydraulic servo universal testing machine is shown in Figure 3.8. The concrete compression tests were conducted according to the standard of GB/T 50081. To ensure the reliability of the test results, three specimens for each test mix were tested.

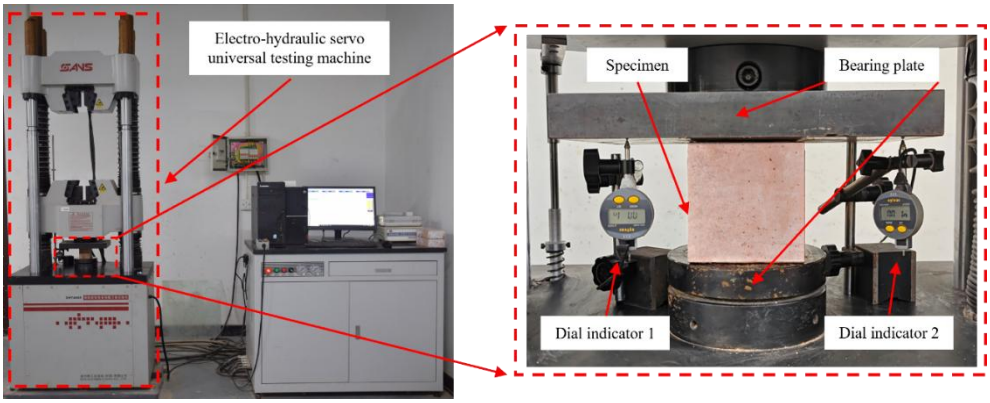


Figure 3.8 Microcomputer-controlled electro-hydraulic servo universal testing machine device used in the uniaxial compression test.

3.3.4 Nuclear magnetic resonance test

LF-NMR is a non-destructive technique that characterizes water-accessible pore structure in cementitious materials through ^1H T_2 relaxometry. In this study, cylindrical specimens were tested to track the evolution of pore-related indicators (T_2 distribution, apparent pore-size components, and porosity) of geopolymer concrete with different basalt fiber contents as the erosion duration increased. It is noted that interpreting T_2 spectra in terms of pore size relies on common surface-relaxation assumptions and, without dedicated surface-relaxivity calibration, the LF-NMR results are primarily used here for comparative assessment of pore-structure evolution rather than absolute pore-size quantification.

Prior to testing, specimens were conditioned using a standardized vacuum saturation procedure. The specimens were placed in a vacuum device for 2.0 h, followed by water saturation in a bath for 24.0 h. After saturation, the specimens were immediately sealed with cling film to minimize moisture loss and then subjected to LF-NMR testing. This conditioning protocol was applied consistently to all specimens to provide comparable moisture states for the NMR measurements. It is also recognized that incomplete saturation of the finest pores may lead to underestimation of NMR-derived porosity; therefore, the LF-NMR results are interpreted together with macroscopic durability-related indicators (mass/density change, UPV, and mechanical properties).

In this study, LF-NMR testing was conducted at a resonance frequency of 12.0 MHz with a probe coil diameter of 60.0 mm. The temperature during testing was maintained at 32.00 ± 0.02°C. Figure 3.9 shows the LF-NMR testing equipment.

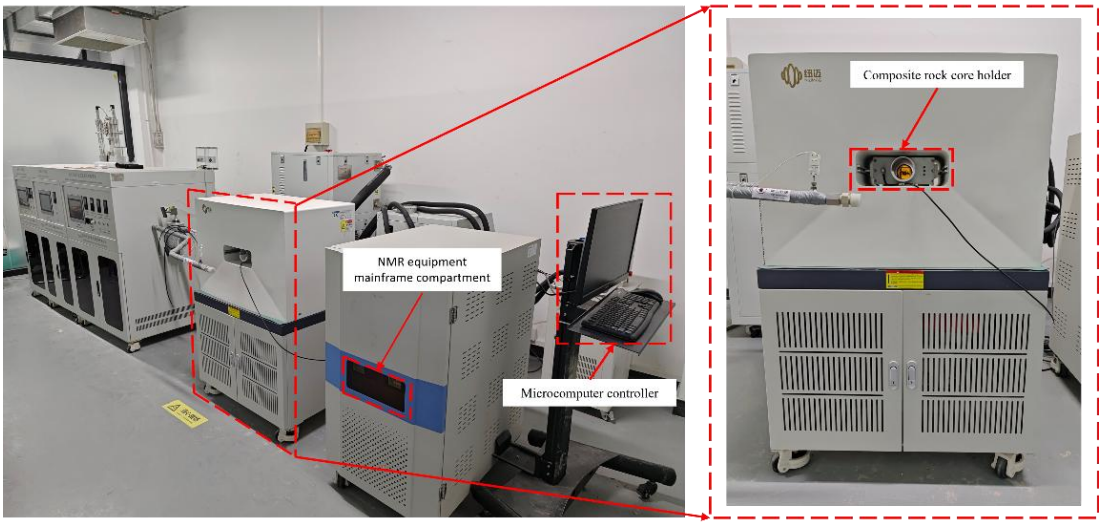
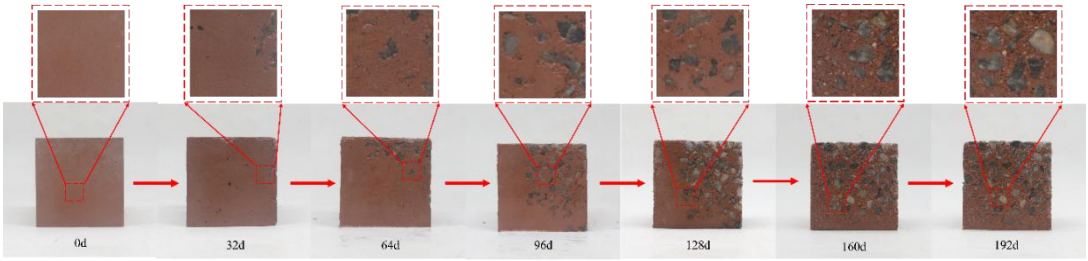


Figure 3.9 Equipment for LF-NMR testing.

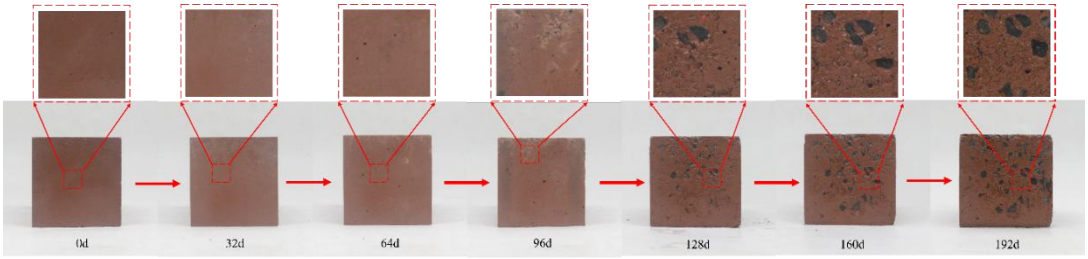
3.4 Results and Discussion

3.4.1 Apparent Changes in BFRGC under Erosion Exposure

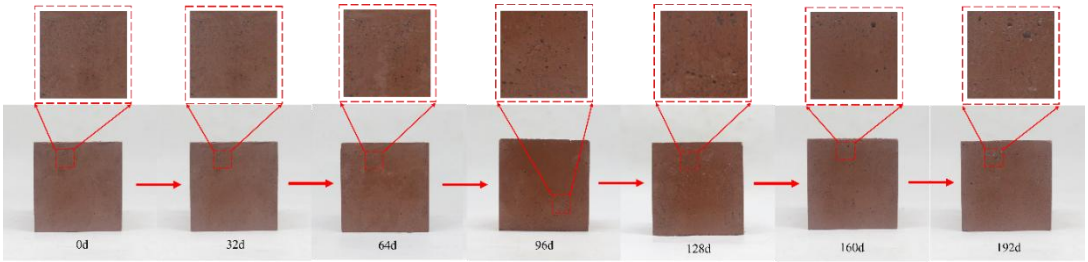
Figure 3.10 shows the surface changes of BFRGC specimens at different erosion testing periods. Figure 3.10 (a) shows that under W-D cycle conditions, specimens with 0% basalt fiber content displayed slight spalling at the edges at 32 days, with spalling expanding as the erosion testing age increased. By 192 days, most of the specimen surface exhibited spalling, exposing a significant number of aggregates. Figure 3.10(b) shows that specimens with 1.0% basalt fiber content exhibited signs of erosion spalling only after 96 days , with apparent aggregate exposure at 128 days, and considerable aggregate exposure by 192 days. Figure 3.10 (c) shows that specimens with 0% basalt fiber content exhibited minor surface changes during immersion treatment conditions. At 96 days, a slight spalling occurred on the specimen surface, and as the erosion testing time increased further, larger pores developed by 192 days. Figure 3.10 (d) shows that specimens with 1.0% basalt fiber content exhibited no significant surface changes as the erosion testing age increased. However, upon magnified observation, it was evident that tiny pores gradually increased with age.



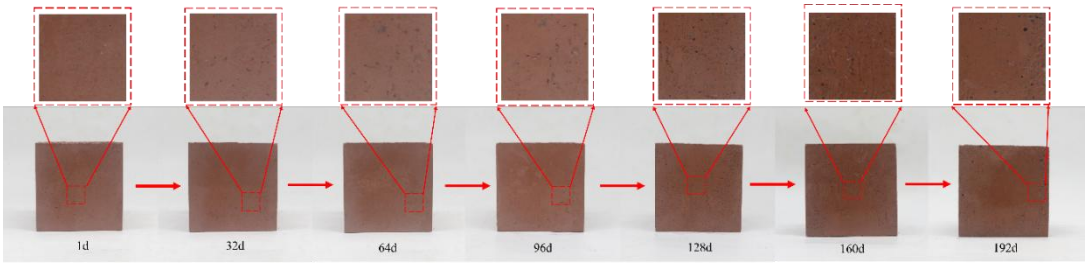
(a)



(b)



(c)



(d)

Figure 3.10 Apparent changes in BFRGC cubes under different erosion conditions. (a) W-D cycle, 0% fiber; (b) W-D cycle, 1.0% fiber; (c) Immersion, 0% fiber; (d) Immersion, 1.0% fiber.

Under wet-dry cycling, control specimens without fibers show markedly more severe surface scaling and aggregate exposure, whereas the fiber-reinforced specimens exhibit delayed and more uniform spalling. This contrast reflects the combined

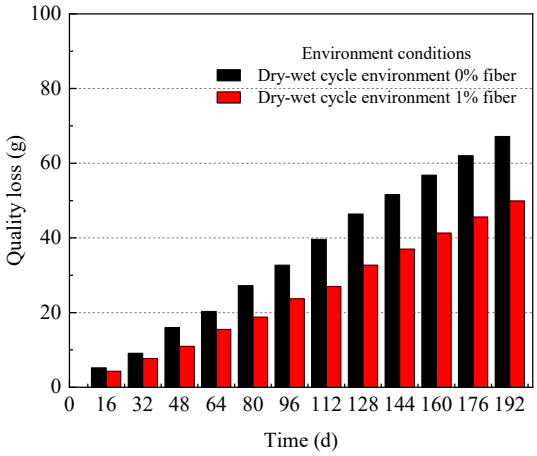
effects of exposure regime and material resistance. The damaging effect of wet-dry cycling is mechanistically linked to repeated imbibition and evaporation, which promotes salt crystallisation pressure near drying fronts, accelerates near-surface damage, and weakens the interfacial transition zone. In continuous immersion, the absence of pronounced crystallisation distress generally results in milder surface deterioration, and damage evolution is dominated by slower chemical interactions and transport processes.

The incorporation of short basalt fibers contributes to improved damage tolerance primarily through crack control. Fiber bridging can restrain microcrack opening and delay crack coalescence, which helps reduce the likelihood of spalling initiated from interconnected cracking. Within the investigated mix design and fiber dosage, the qualitative surface-damage ranking is: W-D/0% fiber > W-D/1% fiber ≫ Immersion/0% fiber ≳ Immersion/1% fiber.

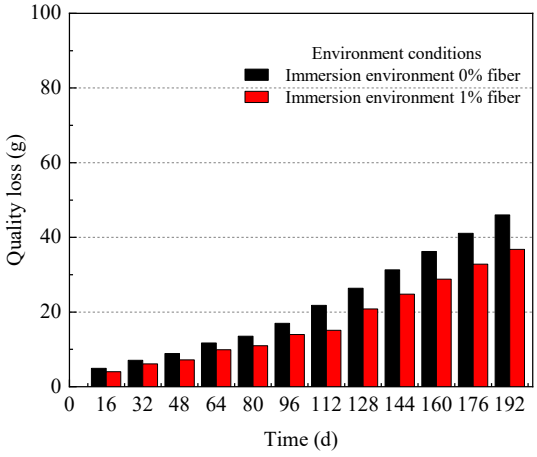
3.4.2 Mass and Density Changes of BFRGC Under Erosion Exposure

Figure 3.11 shows the effect of erosion testing periods on the mass of GPC under different erosion testing conditions. Figure 3.11 (a) shows that under W-D cycle conditions, at a erosion testing age of 16 days, the mass losses of specimens with 0% fiber content and 1 % fiber content were 5.2g and 4.3g, respectively, with a difference of 0.9g. With increasing erosion testing age, at a erosion testing age of 192 days, the mass losses were 67.2g and 49.9g, respectively, with a difference of 17.3 g, representing an 1822% increase in the mass loss differential compared to the 16-day measurement. Figure 3.11 (b) shows that under immersion environment conditions, at a erosion testing age of 16 days, the mass losses were 4.9g and 4.1g, respectively, with a difference of 0.8g. With increasing erosion testing age, at a erosion testing age of 192 days, the mass losses were 46.1g and 36.8g, respectively, with a difference of 9.3g, representing an increase of 1063% compared to 16 days. As the erosion duration increased in both environments, the mass loss of plain GPC (0% fiber) became more pronounced, indicating the effective inhibition of spalling in the specimens due to the incorporation of fiber. Compared to immersion environment, W-D cycle conditions

show a more significant mass loss, indicating a more pronounced deterioration effect on GPC under such conditions.



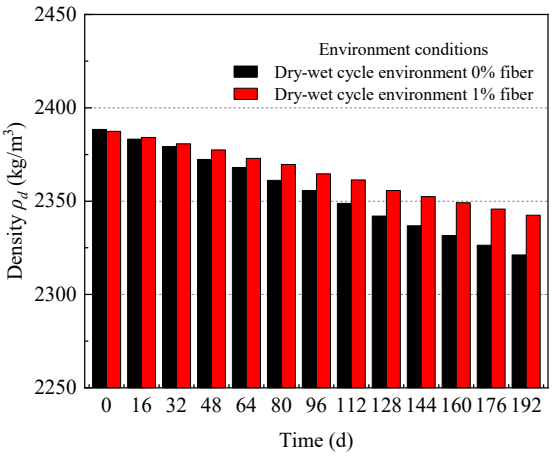
(a)



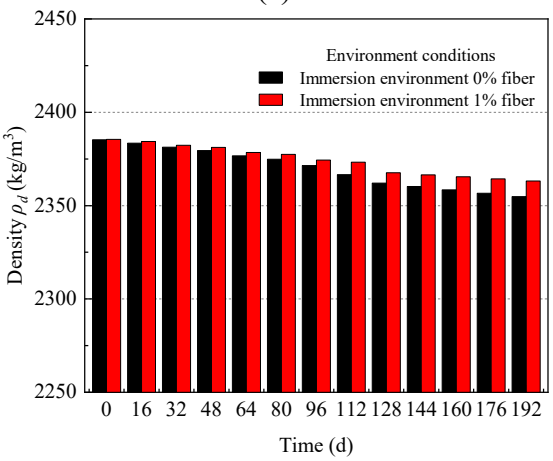
(b)

Figure 3.11 Mass Loss of BFRGC Under Different Erosion Conditions. (a) W-D cycle (b) Immersion environment

Figure 3.12 shows the effect of erosion testing periods on the density of GPC under different erosion testing conditions. Figure 3.12 (a) shows that under W-D cycle conditions, the density of specimens with 0% basalt fiber content decreased from 2388.4 kg/m^3 to 2321.2 kg/m^3, indicating a 2.8% decrease, while specimens with 1.0% basalt fiber content decreased from 2387.4 kg/m^3 to 2342.5 kg/m^3, showing a 1.9% decrease. Figure 3.12 (b) shows that under immersion environment conditions, the density of specimens with 0% basalt fiber content decreased from 2385.3 kg/m^3 to 2354.8 kg/m^3, with a decrease of 1.3%, while specimens with 1.0% basalt fiber content decreased from 2385.5 kg/m^3 to 2363.2 kg/m^3, showing a decrease of 0.9%.



(a)



(b)

Figure 3.12 Density variation of BFRGC under different erosion periods. (a) W-D cycle environment. (b) Immersion environment.

Under wet-dry (W-D) cycling, mass loss is a surface-controlled process. Repeated cycles of solution imbibition and evaporation generate destructive salt crystallization pressures within near-surface pores, directly accelerating edge scaling and aggregate exposure. These concentration-dilution cycles further amplify granular detachment (Flatt, 2002; Steiger, 2005). Consequently, the normalized mass-loss rate significantly exceeds that of continuous immersion, where damage is governed by slower chemical reactions in the absence of crystallization-induced spalling (Guo et al., 2019; Liu *et al*., 2021; Yang *et al*., 2025).

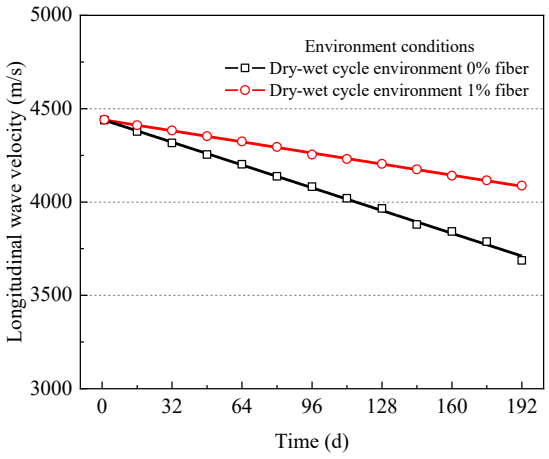
The observed reduction in bulk density provides a volume-scale measure of this deterioration. Microcrack networking and increased interfacial transition zone (ITZ)

porosity raise the fraction of open and connected pores, causing density to drop even when absolute mass loss is modest (ASTM C642, 2021).

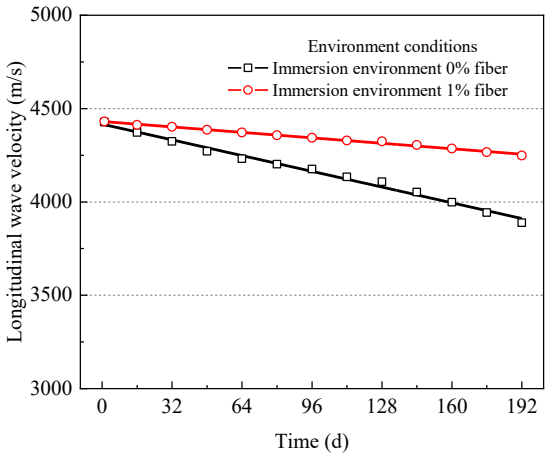
Consistent with the above surface-damage observations, the 1% basalt fibre mixes also exhibit reduced mass loss and a smaller density reduction. This trend indicates a lower propensity for fragment detachment, as fibre bridging and pull-out energy dissipation can moderate the size and frequency of spalled pieces (He et al., 2024; Hu et al., 2024).

3.4.3 The Ultrasonic Wave Velocity in Geopolymer Concrete at Different Erosion testing periods

Figure 3.13 shows the effect of erosion testing periods on the longitudinal wave velocity of GPC under different erosion testing conditions. In Figure 3.13 (a), after 192 days of W-D cycles, the wave velocity of plain GPC (0% fiber) decreased from 4439 m/s to 3687 m/s, a reduction of approximately 17%. For specimens with 1% basalt fiber, during the same period, the wave velocity decreased from 4441 m/s to 4088 m/s, with a reduction of about 7.9%. In Figure 3.13 (b), after 192 days of immersion treatment, the wave velocity of 0% fiber GPC decreased from 4427.0 m/s to 3888.0 m/s, with a reduction of approximately 12.2%. For specimens with 1.0% basalt fiber, during the same period, the wave velocity decreased from 4431 m/s to 4248 m/s, with a reduction of about 4.1%.



(a)



(b)

Figure 3.13 Longitudinal wave velocity of BFRGC under different erosion periods. (a) W-D cycle environment. (b) Immersion environment.

Under wet-dry (W-D) cycling, ultrasonic pulse velocity (UPV) degradation is markedly more pronounced. This occurs because elastic-wave speed is highly sensitive to the formation of microcrack networks and the loss of continuity along its path; even sub-millimetre cracking or interfacial transition zone (ITZ) debonding can trigger a disproportionate drop in velocity. The W-D mechanism itself drives this damage: repeated evaporation and imbibition of sulfate solution cause salt crystallization at drying fronts. This process actively opens and stabilizes microcracks while weakening the ITZ, thereby amplifying UPV loss. This stands in stark contrast to continuous immersion, where damage develops from slower chemical reactions in the absence of disruptive crystallization pressure (Rodríguez-Navarro et al., 2000; Wang *et al*., 2020).

The incorporation of 1% basalt fibres is associated with a smaller UPV decline, suggesting better preservation of acoustic continuity as damage accumulates. This behaviour is consistent with fibre-assisted crack control, which can delay microcrack growth and coalescence. Given that UPV is sensitive to cracking and microstructural discontinuities under the coupled actions discussed above, the reduced UPV loss should be interpreted as an integrated response of improved crack control and the underlying matrix resistance rather than an effect governed by fibre reinforcement alone. This interpretation aligns with established UPV-based condition assessment

standards for concrete (ASTM International, 2022) and is supported by recent reports on basalt-fibre-reinforced geopolymers under marine exposure (Zhang et al., 2024).

3.4.4 Mechanical Properties of Geopolymer Concrete under Erosion Exposure

Figure 3.14 shows the compressive stress-strain curves of GPC with different erosion testing periods under various erosion testing conditions. With an increase in the erosion testing period, the peak stress of the stress-strain curves for the GPC specimens gradually decreased, and the corresponding strain at peak stress increased, suggesting a gradual decline in the ultimate compressive stress and an increase in the ultimate compressive strain of GPC. A comparison of Figure 3.14 (a), Figure 3.14 (b), Figure 3.14 (c), and Figure 3.14 (d) reveals that reveals that the addition of basalt fiber has a less pronounced effect on the compressive strength of concrete. In the W-D cycle environment, with the increase of erosion testing age, the stress-strain curve of 1.0% fiber GPC shows a noticeable decreasing trend, with the ultimate compressive stress gradually decreasing. Compared to 0% fiber GPC, the reduction is relatively smaller. In the immersion environment, the stress-strain curve also exhibits a gradual decrease with the increase of erosion testing age.

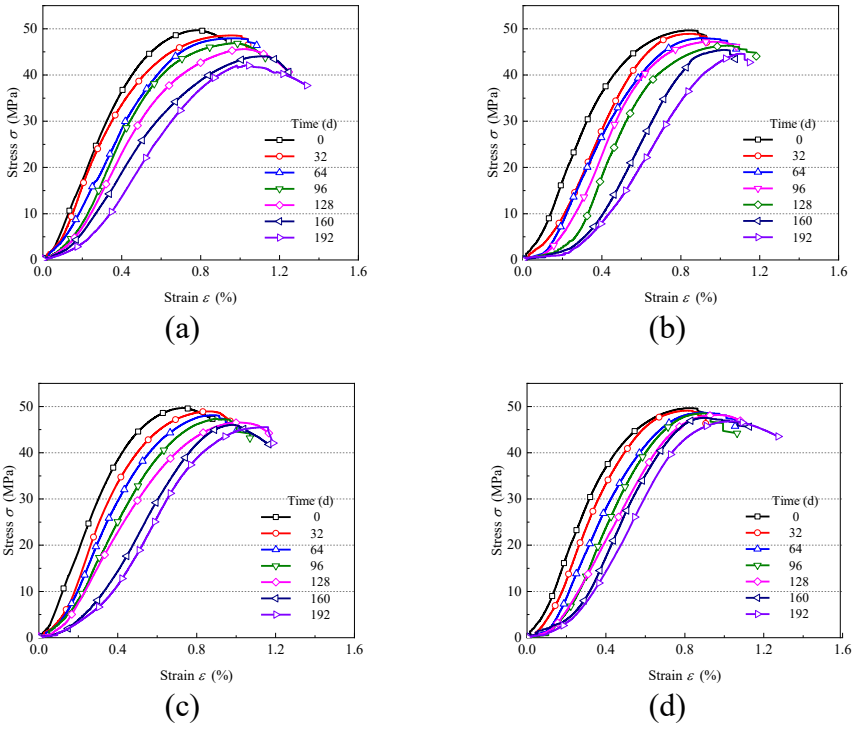
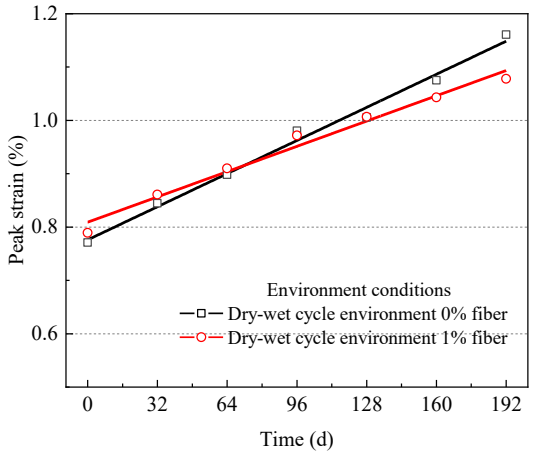
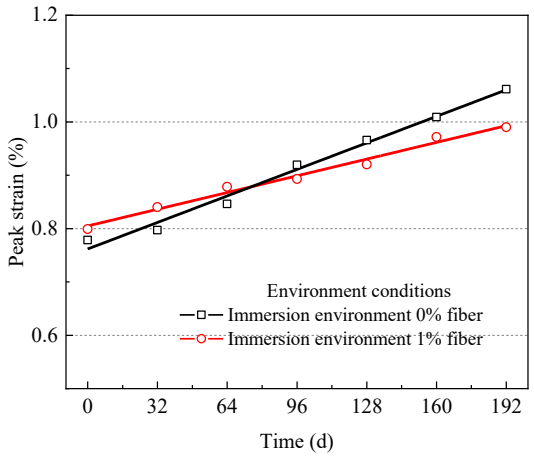


Figure 3.14 Stress-strain curves of BFRGC at different erosion periods under varied conditions. (a) W-D cycles, 0% fiber; (b) W-D cycles, 1.0% fiber; (c) Immersion, 0% fiber; (d) Immersion, 1.0% fiber.

Figure 3.15 shows the effect of erosion testing periods on the peak strain of GPC under different erosion testing conditions. In Figure 3.15 (a), with increasing age of erosion testing, the peak strain of 1.0% fiber GPC increased by 35.4% from 0.79% to 1.07% in W-D cycle conditions. While the peak strain of 0% fiber ground polymer concrete increased from 0.77% to 1.16%, an increase of 50.7%. In Figure 3.15 (b), the peak strain of 1.0% fiber ground polymer concrete increased by 23.7% from 0.80% to 0.99% in long term immersion environment. Whereas the peak strain of ground polymer concrete with 0% fiber increased from 0.77% to 1.06% with an increase of 37.6%. It can be observed that there is a difference of 11.6% between the peak strains of 1.0% fiber GPC in the immersed and wet and dry cycling environments, while there is a difference of 13.1% for 0% fiber GPC.



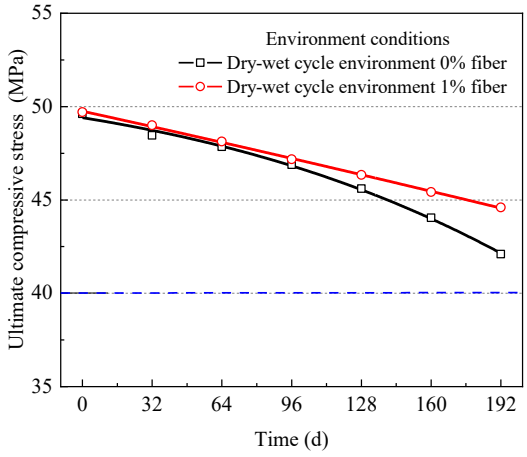
(a)



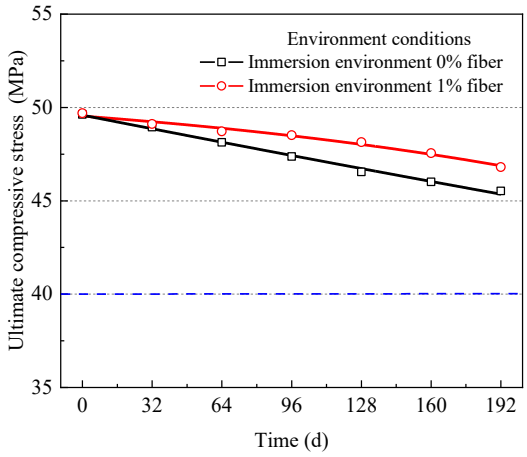
(b)

Figure 3.15 Effect of erosion periods on the peak strain of GPC under different conditions. (a) W-D cycle environment. (b) Immersion environment.

Figure 3.16 shows the effect of erosion testing periods on the ultimate compressive stress of GPC under different erosion testing conditions. In Figure 3.16 (a), under W-D cycle conditions, the ultimate compressive strength of 1.0% fiber GPC decreased from 49.6 MPa to 44.6 MPa with the increase of erosion testing age, representing a reduction of 10.1%. Similarly, for 0% fiber GPC, the ultimate compressive strength decreased from 49.7 MPa to 42.1 MPa, showing a reduction of 15.5%. In Figure 3.16 (b), under immersion environment conditions, the ultimate compressive strength of 1.0% fiber GPC decreased from 49.6 MPa to 46.8 MPa, indicating a reduction of 5.6%. Conversely, for 0% fiber GPC, the ultimate compressive strength decreased from 49.7 MPa to 45.5 MPa, with a reduction of 8.1%. Both types of GPC showed excellent compressive strength retention under the erosive conditions. After 192 days of immersion, the compressive strength of the GPC remained above 40 MPa, which is comparable to that of OPC concrete. This performance meets the European Standard for marine concrete strength requirements, indicating that BFRGC has potential as a substitute for C40 concrete in marine environments.



(a)



(b)

Figure 3.16 Effect of erosion periods on the ultimate compressive stress of GPC under different conditions. (a) W-D cycle environment. (b) Immersion environment.

Figure 3.17 shows the effect of erosion testing periods on the elastic modulus of GPC under different erosion testing conditions. In Figure 3.17 (a), under W-D cycle conditions, the elastic modulus of specimens with 1.0% basalt fiber content decreased from 9.56 GPa to 8.40 GPa with a decrease of 12.1%, while specimens with 0% basalt fiber content decreased from 9.58 GPa to 8.11 GPa, with a decrease of 15.4%. In Figure 3.17 (b), under immersion environment conditions, the elastic modulus of specimens with 1.0% basalt fiber content decreased from 9.55 GPa to 9.02 GPa, with a decrease of 5.5%, while specimens with 0% basalt fiber content decreased from 9.58 GPa to 8.78 GPa, with a decrease of 8.4%. The elastic modulus shows a linear decreasing trend in both environments.

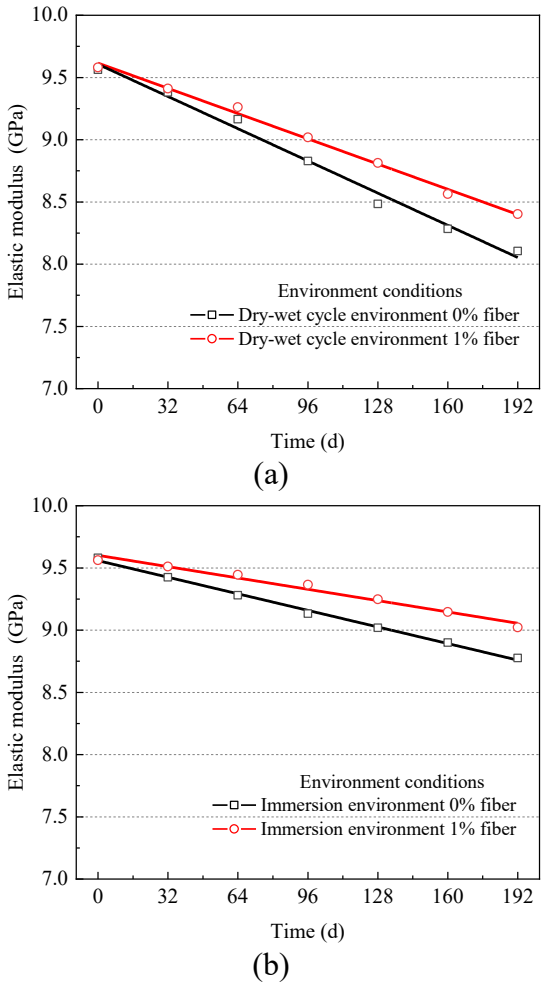


Figure 3.17 Effect of erosion periods on the elastic modulus of GPC under different conditions. (a) W-D cycle environment. (b) Immersion environment.

Under W-D cycling, repeated moisture exchange drives cyclic shrinkage and expansion, progressively initiating and growing microcracks and promoting debonding within the interfacial transition zone (ITZ). Simultaneously, dissolved salts concentrate near evaporation fronts, where repeated crystallization and dissolution further amplify near-surface damage, a process far more aggressive than the degradation under continuous immersion (Wang, 2021).

Consistent with this mechanism, specimens under W-D cycling exhibit greater losses in density, mass retention, and ultrasonic pulse velocity (UPV). These losses reflect a rapid increase in connected porosity and crack density, contrasting with sustained immersion, where damage results from slower chemical reactions in the pore solution (Guo *et al.*, 2021).

Notably, UPV highly sensitive to elastic continuity at sub-millimetre scales—declines disproportionately as microcracks multiply and coalesce. This heightened sensitivity explains why acoustic indicators often deteriorate earlier than bulk properties like mass or density in W-D environments (Hong *et al*., 2020).

In parallel, sulfate exposure contributes to matrix softening and pore-structure coarsening, and wet-dry cycling can accelerate these deleterious processes by facilitating ion transport and inducing repeated concentration-dilution cycles within the cover zone (Qin et al., 2024). Against this aggressive backdrop, the inclusion of 1% basalt fibre primarily enhances damage tolerance—reflected in a more stable post-peak response—rather than increasing peak compressive strength within the investigated mix design and exposure regimes. The fibres act through crack bridging and pull-out mechanisms that dissipate energy and moderate damage localisation, which is consistent with the reduced rates of strength and UPV decline observed in the fibre-reinforced mixes (Fan et al., 2024).

Overall, these observations indicate that the measured performance evolution reflects coupled chemical-physical actions under the prescribed exposure regimes, with fibre-related crack control providing a contributing, damage-mitigating effect.

3.4.5 Pore Structure Characteristics of Geopolymer Concrete at Different Erosion Periods

In low-field nuclear magnetic resonance (LF-NMR) testing, the signal intensity is proportional to the quantity of hydrogen protons within the sample, indicating a close correlation between the LF-NMR signal intensity and the moisture content of the specimen. Changes in the pore structure characteristics of the specimen can be characterized through variations in the LF-NMR T_2 spectrum. Figure 3.18 shows the effect of transverse relaxation time T_2 on the signal amplitude of GPC under different erosion testing periods. The transverse relaxation time (T_2) distribution for the concrete typically exhibited two predominant peaks. With increasing erosion testing

age, the signal amplitudes of the two peaks for different basalt fiber contents gradually increase under both simulated environments.

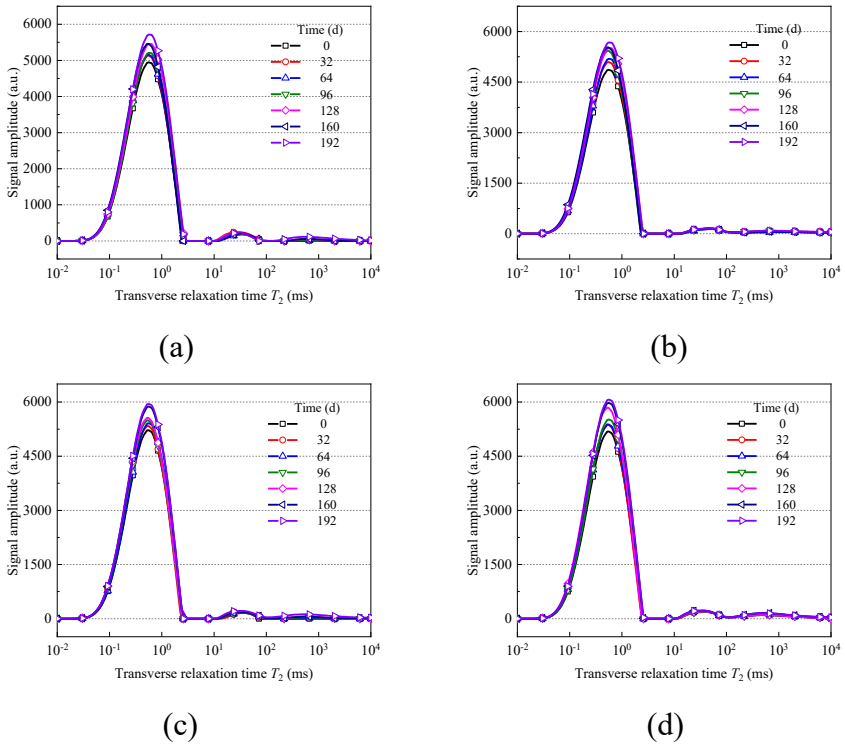


Figure 3.18 Effect of t T_2 on the signal amplitude of GPC under different erosion periods. (a) W-D cycle, 0% fiber. (b) W-D cycle, 1% fiber. (c) Immersion environment, 0% fiber. (d) Immersion environment, 1% fiber.

The transverse relaxation time T_2 from LF-NMR testing reflects the distribution of pore radii, where longer T_2 values indicate larger pore radii. Therefore, shorter T_2 values correspond to smaller pore radii. T_2 can be characterized as the distribution of pore radii. Figure 3.19 shows the effect of pore size on the pore size distribution of GPC under different erosion testing periods. The pore radius of GPC exhibits two peaks. The first peak concentrates in the range of 10^{-3} to 10^{-1} μm, while the second peak concentrates in the range of 10^{-1} to 10^{2} μm. With increasing erosion testing age, the amplitudes of both peaks gradually increase. At a erosion testing age of 96 days, a third peak in pore radius emerges, concentrated in the range of 10^{1} to 10^{2} μm, with a slight increase in amplitude over time, more prominently observed in the dry-wet cycle environment. The results of LF-NMR testing indicate that the pore structure of

GPC increases with increasing erosion testing age, with the appearance of larger pores over time.

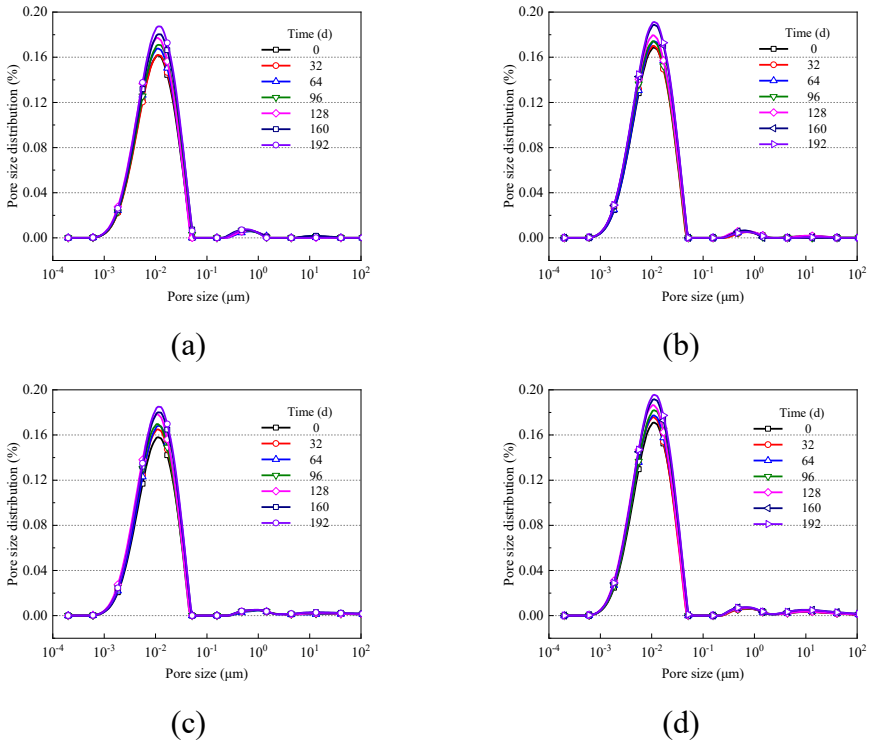
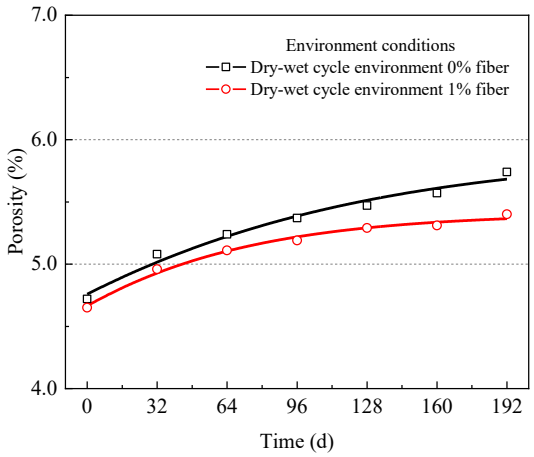
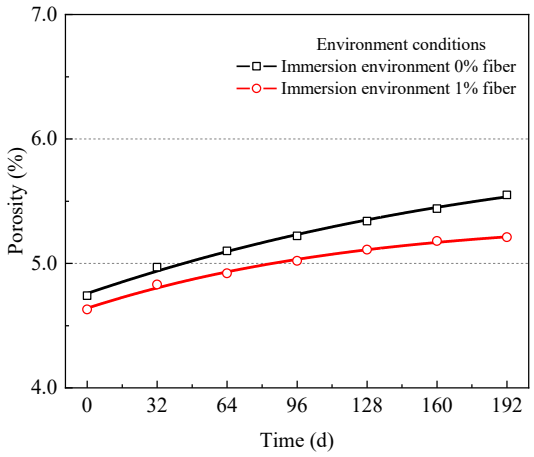


Figure 3.19 Pore size distribution of GPC under different erosion periods. (a) W-D cycle, 0% fiber. (b) W-D cycle, 1% fiber. (c) Immersion environment, 0% fiber. (d) Immersion environment, 1% fiber.

Fig 3.20 shows the effect of erosion testing periods on the porosity of GPC under different erosion testing conditions. As shown in Fig 3.20 (a), under W-D cycle conditions, the porosity of GPC without basalt fiber content increased from 4.7% to 5.7%, which represents a growth of 21.6%. Meanwhile, GPC with 1.0% basalt fiber content increased from 4.6% to 5.4%, with a growth of around 16.1%. As depicted in Fig 3.20 (b), under immersion environment conditions, the porosity of GPC without basalt fiber content increased from 4.7% to 5.5%, with an increment of approximately 17.5%, while GPC with 1.0% basalt fiber content increased from 4.6% to 5.2%, with an increment of about 12.7%.



(a)



(b)

Fig 3.20 Effect of erosion testing periods on the porosity of GPC under different erosion conditions. (a) W-D cycle environment. (b) Immersion environment.

In saturated cementitious media, LF-NMR T_2 relaxometry is often interpreted using a surface-relaxation-dominated framework, in which characteristic T_2 components are correlated with an effective pore-size scale. Under this interpretation, a right-shift of T_2 components and/or an increase in long-T_2 amplitudes is generally taken to indicate pore coarsening, reflecting a redistribution of porosity from finer pores toward larger capillaries and microcrack-related voids (Brownstein and Tarr, 1977). At the same time, deriving an absolute pore-size distribution from T_2 can be sensitive to the assumed surface relaxivity and may be affected by diffusion-related coupling effects. Accordingly, the present study uses LF-NMR primarily as a non-destructive indicator for comparative assessment of pore-structure evolution under a consistent

conditioning protocol, rather than for definitive absolute pore-size quantification across all durability-relevant pore ranges.

Under wet-dry cycling, recurrent moisture exchange and salt crystallisation at drying fronts and within the ITZ can promote crack opening and pore coalescence. This damage mechanism is reflected by the emergence of an additional long-T_2 component and a faster overall increase in NMR-derived porosity. In contrast, continuous immersion, lacking the repeated crystallisation pressure associated with drying, results in gentler spectral evolution over the 192-day period (Guo et al., 2020).

Mixes incorporating 1% basalt fibre exhibit reduced peak shifts and slower porosity growth compared with the fibre-free mixes. This behaviour is consistent with fibre bridging that restrains crack initiation and coalescence and thereby limits the development of larger pores and microcrack-related voids (Zeng et al., 2022). Since T_2 relaxometry does not directly measure pore connectivity, these microstructural indications are interpreted alongside independent macroscopic indicators, including the consistent deterioration trends observed in UPV degradation and stiffness/strength retention in this study (Espinosa et al., 2023).

3.5 Summary

This chapter presents an experimental study on the durability performance of BFRGC under simulated marine sulfate exposure conditions. The specimen preparation, exposure conditions, and testing procedures, including mechanical property evaluation and pore structure characterization**, are** described in detail. The key findings are summarized as follows:

1. The incorporation of basalt fibers contributed to improved durability-related performance of GPC under the investigated exposure regimes. Under W-D cycling conditions, with increasing exposure duration, the compressive strength of BFRGC with 1% fiber content decreases by 10.1%, whereas GPC exhibits a more substantial reduction of 15.5%. Similarly, the elastic modulus of BFRGC decreases by 12.1%, compared to a greater decline of 15.4% in GPC.

2. W-D cycling exacerbated the degradation of GPC compared to continuous immersion. Under such conditions, the compressive strength of 1.0% BFRGC declines by 10.1%, whereas in an immersion environment, the same mixture shows a lower strength reduction of 5.6%. Likewise, the elastic modulus of 1.0% BFRGC specimens decreases by 12.1% under W-D cycling, compared to only 5.5% under continuous immersion.

3. LF-NMR analysis suggests a slower microstructural deterioration in mixes incorporating basalt fibers, which is consistent with improved crack control under the investigated conditions. The expansion of pore size is less pronounced in 1.0% BFRGC specimens than in GPC. After 96 days of exposure, GPC exhibits a third pore-size peak that does not appear in BFRGC under identical conditions, indicating that basalt fibers mitigate the shift toward pore coarsening and microcrack-related void development.

4. LF-NMR results indicated that both 1.0% BFRGC and GPC maintained relatively stable pore-structure evolution under prolonged immersion conditions. The porosity of both 1.0% BFRGC and GPC increases only marginally with exposure duration. Furthermore, no third peak in pore size distribution emerges within 192 days, supporting the observation of a relatively stable microstructural evolution under immersion conditions.

Outlook

This experimental programme provides a coherent and well-controlled dataset that is sufficient to support an initial machine-learning prediction of sulfate-induced strength deterioration. To enhance the breadth and generalizability of the resulting models, this study augments the in-house measurements with a curated dataset from the literature, compiled under consistent experimental conditions (5% Na_2SO_4 solution; standard 28-day curing prior to exposure). This expanded corpus broadens coverage across diverse mix designs and exposure durations, mitigates overfitting risks, and supports the development of more robust and transferable deterioration

models for geopolymer concrete under sulfate immersion, while remaining aligned with the scope and constraints of the present experimental programme.

Chapter 4. METHODOLOGY FOR REMAINING STRENGTH PREDICTION USING CONDITIONAL VARIATIONAL AUTOENCODER ENHANCED MACHINE LEARNING ALGORITHM

4.1 Introduction

Building on the controlled sulfate-exposure dataset established in this thesis under standardized conditions, we now turn to data-driven prediction of GPC strength. The growing demand for low-carbon construction has positioned GPC as a credible alternative to OPC, offering advantages in mechanical performance, durability, and embodied-carbon reduction. However, reliable strength prediction remains challenging due to the variability of precursor materials and the high-dimensional, nonlinear relationships among mix parameters. Traditional empirical or analytical models struggle to represent these interactions, whereas ML can leverage experimental data to learn complex mappings from mix proportions to strength. Recent advances have produced a diverse toolbox of regressors—including ensemble methods (Random Forest, XGBoost), distance-based approaches (K-Nearest Neighbors), kernel models (SVR), neural networks (MLP), and newer learners such as Extra Trees and CatBoost—that have shown strong accuracy in concrete property prediction.

Against this backdrop, this chapter presents an advanced deep-learning framework centered on a Conditional Variational Autoencoder (CVAE) to enhance GPC strength prediction. By uniting probabilistic data generation with representation learning and coupling it with Bayesian hyperparameter optimization (Optuna), the CVAE pipeline improves predictive performance and robustness under limited data. We detail the theoretical foundations, algorithmic structures, and applicability of the baseline ML models and the CVAE-enhanced approach, establishing the methodological basis for subsequent model training and evaluation.

4.2 Theoretical Basis of Conventional Machine Learning Techniques

4.2.1 K-Nearest Neighbors

As shown in Figure 4.1, K-Nearest Neighbors (KNN) is a non-parametric method based on similarity measures among samples. For a given sample to be predicted, KNN first identifies its K nearest neighbors in the feature space according to a predefined distance metric (such as Euclidean or Manhattan distance). The predicted value for the sample is then obtained through a weighted average (or simple arithmetic mean) of the target values of these neighboring samples. The figure demonstrates the spatial distribution of different categories of samples (such as classes A, B, and C), and the relationship between the prediction point and its nearest neighbors. The KNN algorithm is straightforward to implement and effectively captures the intuitive principle of "similarity-based clustering"It is particularly suitable for regression modeling scenarios in material research, where datasets are limited, relationships are complex, and samples exhibit certain clustering features.

The fundamental principle of KNN regression is as follows: for an unknown sample x, KNN first identifies the K closest neighbor points from the training set based on a selected distance measure. The target variables (e.g., strength) of these neighbor points are averaged—often in a weighted manner—to obtain the predicted value of x.

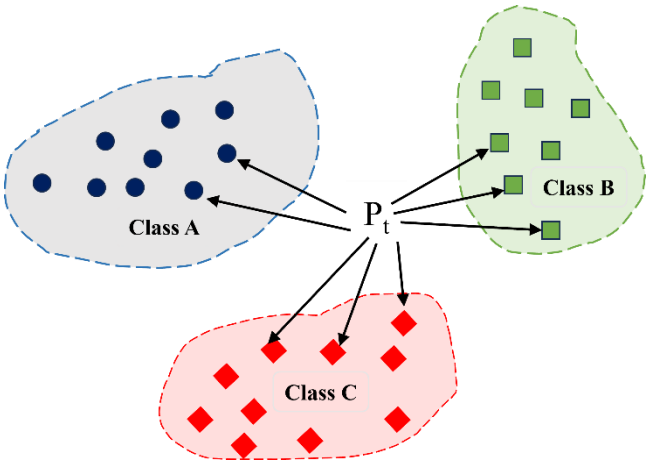


Figure 4.1 K-Nearest Neighbors

The commonly used definition of Euclidean distance is expressed as:

$$d(\mathbf{x}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^{p} (x_j - x_{ij})^2} \tag{3.1}$$

where x denotes the sample to be predicted, x_i represents the i^{th} training sample, p is the number of feature dimensions, and x_j and x_{ij} refer to the j^{th} feature components of x and x_i, respectively.

The predicted output of KNN regression can be formulated as:

$$\hat{y} = \frac{1}{K} \sum_{i \in \mathcal{N}_K(\mathbf{x})} y_i \tag{3.2}$$

where $\mathcal{N}_K(\mathbf{x})$ denotes the index set of the K nearest neighbors of x, and y_i is the target value corresponding to the i^{th} neighbor.

KNN regression provides an additional perspective rooted in physical similarity for concrete strength prediction. When traditional parametric models struggle to capture complex nonlinear relationships, the "similarity-based clustering" mechanism inherent in KNN can effectively identify locally optimal formulation combinations, thereby guiding rapid prototyping and optimization of novel materials.

4.2.2 Support Vector Regression

As shown in Figure 4.2, Support Vector Regression (SVR) is an extension of Support Vector Machines (SVM) applied to regression problems. The core concept of SVR is to identify an optimal hyperplane in a high-dimensional feature space such that the prediction error for most samples falls within a predefined tolerance margin (depicted as the "Margin" region in the figure). Simultaneously, by optimizing the position of support vectors—the data points closest to the hyperplane—SVR maximizes the model's generalization margin, thereby enabling highly accurate regression predictions. SVR demonstrates distinct advantages in handling small datasets, nonlinearity, high-dimensional feature spaces, and scenarios with strong inter-feature coupling, making it widely applicable in the precise prediction of material properties.

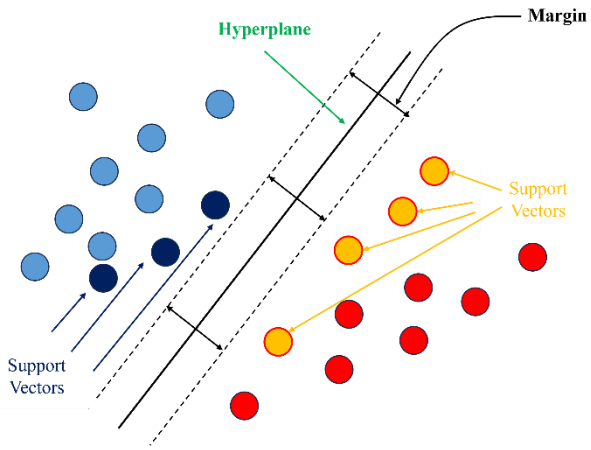


Figure 4.2 Support Vector Regression

SVR attempts to find a function *f(x)* in the input space such that its deviation from the target value *y* does not exceed a specified tolerance *ε*, while maintaining the function as flat (smooth) as possible. The general form of the regression function is expressed as:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \tag{3.3}$$

where *w* is the weight vector, *x* is the input vector, and *b* is the bias term, and ⟨w,x⟩ represents the inner product between vectors w and x.

The objective of SVR is to minimize the following loss function:

$$\min_{\mathbf{w},b,\xi_i,\xi_i^*} \frac{1}{2}\| \mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)^{A} \tag{3.4}$$

subject to the following constraints:

$$\begin{cases} y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \le \varepsilon + \xi_i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{3.5}$$

where *C* is the regularization parameter, and ξ_i, ξ_i^* are slack variables that allow certain samples to have prediction errors exceeding the tolerance *ε*. By introducing slack variables and a regularization term, SVR achieves a balance between prediction accuracy and model complexity, thereby effectively preventing overfitting.

SVR exhibits strong performance and generalization capability when handling nonlinear regression problems, complex feature relationships, and small sample datasets. Although SVR requires extensive parameter tuning and can be

computationally intensive, it remains one of the most competitive and theoretically grounded regression models in both engineering applications and scientific research, making it particularly well-suited for high-precision prediction tasks.

4.2.3 Random Forest

Random Forest (RF) is a nonlinear regression method based on ensemble learning, first proposed by Leo Breiman in 2001. It has been widely applied to data prediction tasks in engineering, materials science, and related fields. As illustrated in Figure 4.3, RF is an ensemble method consisting of multiple decision trees. Each decision tree is trained on a bootstrap sample—generated by randomly sampling with replacement from the original dataset—and serves as a "weak learner"

During the prediction phase, each decision tree produces an independent prediction, and the final output is obtained by aggregating the predictions through majority voting (for classification tasks) or averaging (for regression tasks). The core idea of RF regression is to construct multiple training subsets using the bootstrap sampling technique, with each subset used to train an individual decision tree. During the tree construction process, a random subset of features (rather than the full set) is selected at each node to determine the best split. This mechanism effectively reduces model variance and enhances generalization.

Such an architecture substantially mitigates the risk of overfitting associated with individual models, thereby improving robustness and predictive performance. RF is particularly well-suited for high-dimensional feature spaces, complex datasets, and scenarios involving missing or noisy data . It has become one of the most widely used algorithms in the field of material property prediction.

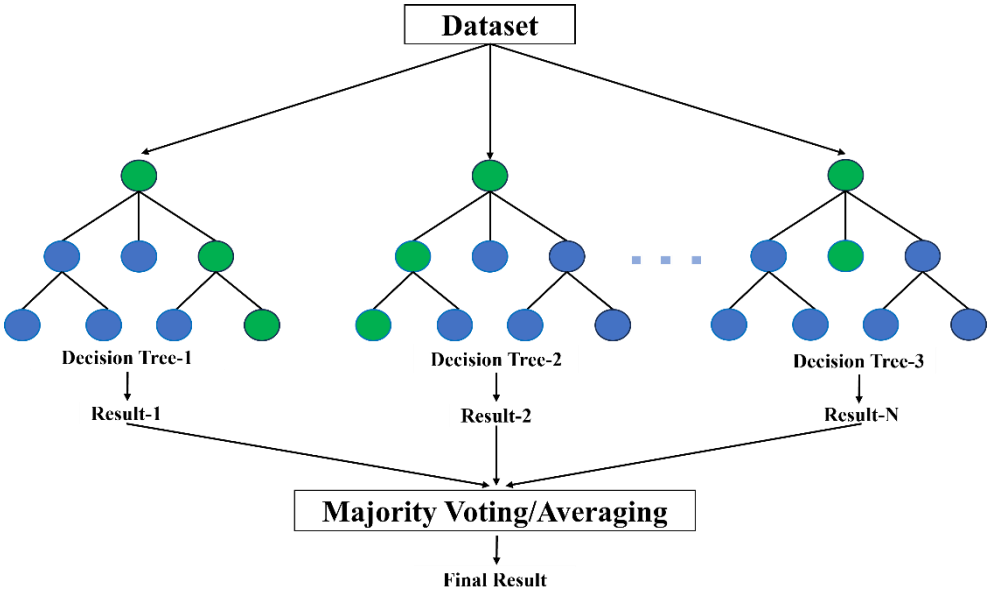


Figure 4.3 Random Forest

For an input sample x, the regression output of the i decision tree is denoted as $f_i(x)$. Given a total of N trees, the Random Forest prediction y is computed as the average of all tree outputs:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(\mathbf{x}) \tag{3.6}$$

The training process of each regression tree $f_i(x)$ essentially involves minimizing a loss function, such as the Mean Squared Error. For regression tasks, the prediction at each leaf node of a decision tree is typically the mean value of the target variables corresponding to the samples falling into that node. The output of a single regression tree can be expressed as:

$$f_i(\mathbf{x}) = \frac{1}{n_{i,\mathbf{x}}}\sum_{j=1}^{n_{i,x}} y_j \tag{3.7}$$

where $n_{i,x}$ denotes the number of samples falling into the leaf node, and y_j represents the true value of the corresponding samples.

The core idea of Random Forest can be summarized as: “an ensemble of weak learners with low correlation and high diversity can significantly improve prediction performance”. Its main advantages include the ability to handle high-dimensional features, strong resistance to overfitting, and robustness to outliers and missing values.

However, its disadvantages lie in the relatively high computational cost during training and prediction, as well as the difficulty in interpreting the decision logic of individual trees within the ensemble.

4.2.4 XGBoost

As illustrated in Figure 4.4, Extreme Gradient Boosting (XGBoost) is an efficient ensemble algorithm based on the concept of gradient boosting. XGBoost builds an additive model by iteratively combining multiple weak learners (represented in the figure as regression trees), where each successive model is trained to fit the residuals of the previous iteration, thereby improving the overall prediction accuracy. The illustration highlights that the outputs of each weak learner are combined in a weighted manner to form a strong predictive model.

Moreover, XGBoost incorporates second-order derivative (Hessian) information and regularization terms for model complexity, which helps prevent overfitting and enhances computational efficiency. Due to its high accuracy and scalability, XGBoost has become one of the most widely used algorithms in large-scale data mining and material property prediction applications.

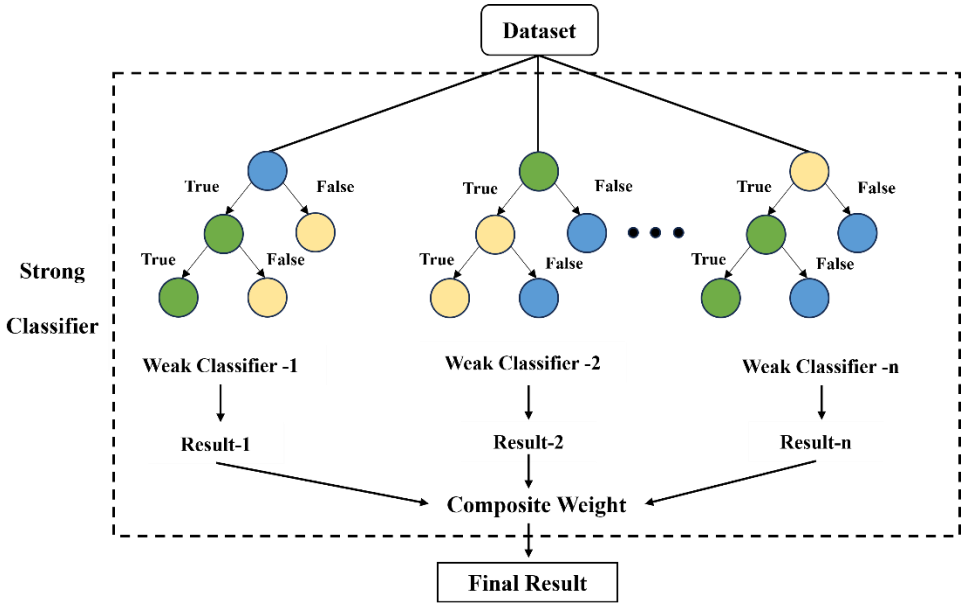


Figure 4.4 XGBoost

XGBoost is essentially an additive model that incrementally adds new trees (regression trees) to the existing ensemble in order to progressively minimize the overall loss function. Given a training dataset $\mathcal{D} = (\mathbf{x}_i, y_i)i = 1^n$, where x_i denotes the input features and y_i represents the target variable, suppose the model has been trained for t−1 iterations. The prediction at the t^{th} iteration is given by:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(\mathbf{x}_i), \quad f_k \in \mathcal{F} \tag{3.8}$$

where $\mathcal{F}$ denotes the set of all regression trees, and f_k represents the regression tree generated in the k iteration.

The objective function of XGBoost consists of two components: the loss function and the regularization term, which can be expressed as:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t) \tag{3.9}$$

where l is the specified loss function (e.g., mean squared error), and Ω is the regularization term that penalizes model complexity to prevent overfitting.
A commonly used form of the regularization term is:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{3.10}$$

where T denotes the number of leaf nodes in the tree, w_j is the weight of the j leaf node, and γ and λ are hyperparameters.

The XGBoost algorithm employs a second-order Taylor expansion to approximate the loss function and utilizes Newton's method for optimization. This approach not only improves computational efficiency by converting the problem into a quadratic minimization with respect to the leaf node weights, but also effectively controls model complexity and prevents overfitting using second-order information and regularization terms.

4.2.5 Multi-Layer Perceptron

As shown in Figure 4.5, the Multi-Layer Perceptron (MLP) is a typical feedforward neural network structure consisting of an input layer, one or more hidden layers, and

an output layer. The MLP Regressor is a regression model based on the backpropagation algorithm. By applying nonlinear transformations between neurons, it effectively captures the complex nonlinear relationships between input features and the target variable.

Compared with traditional linear regression and models based on distance or tree structures, MLP demonstrates superior fitting and generalization capabilities in high-dimensional, strongly coupled, and nonlinear data scenarios. This makes it particularly well-suited for predicting the compressive strength of novel materials such as GPC.

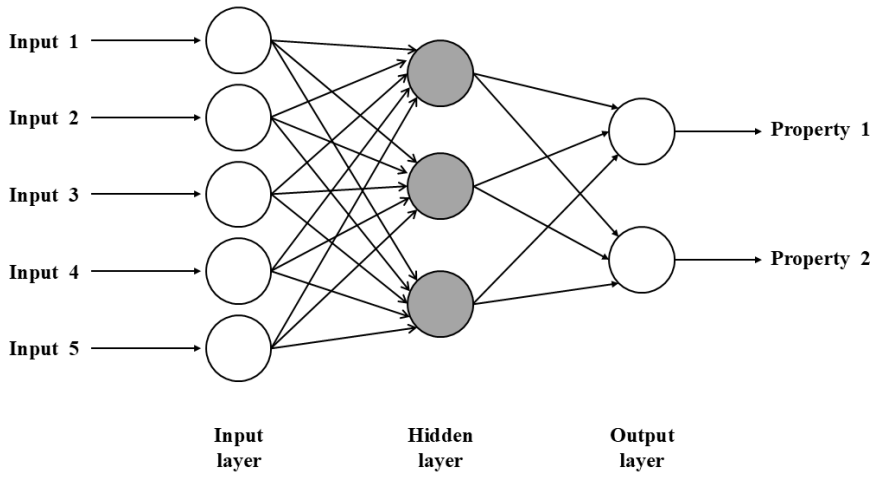


Figure 4.5 Multi-Layer Perceptron

Each layer in a MLP consists of multiple neurons, which are interconnected via weights w. For a given input sample x, the output $h_j^{(l)}$ of the j^{th} neuron in the 1^{th} layer is expressed as:

$$h_j^{(l)} = f\left(\sum_{i=1}^{n_{l-1}} w_{ji}^{(l)} h_i^{(l-1)} + b_j^{(l)}\right) \tag{3.11}$$

where $h_i^{(l-1)}$ denotes the output of the i^{th} neuron in the previous layer, $w_{ij}^{(l)}$ is the weight connecting the i^{th} neuron in layer $l-1$ to the j^{th} neuron in layer l, $b_j^{(l)}$ is the bias term, f is the activation function (e.g., ReLU, tanh, or sigmoid), and n_{l-1} is the number of neurons in the previous layer.

The final output of the MLP Regressor (the predicted value) is:

$$\hat{y} = g\left(\sum_{j=1}^{n_L} w_j^{(L+1)} h_j^{(L)} + b^{(L+1)}\right) \tag{3.12}$$

where L is the total number of hidden layers, and g is typically a linear activation function to suit regression tasks.

The MLP Regressor learns the complex relationship between input features and the target variable through multiple layers of nonlinear transformations and the backpropagation algorithm. It is capable of capturing high-order nonlinearities and feature interactions, making it highly effective in predicting material properties under conditions of structural complexity and strong feature coupling.

Its primary strengths lie in its powerful representational capacity and adaptability, allowing it to uncover hidden feature relationships. However, MLP models also have notable limitations, such as sensitivity to the size and quality of training data, heavy dependence on hyperparameters and network architecture, and potential issues during training including vanishing gradients and unstable convergence.

4.2.6 Extra Trees Regressor

The Extremely Randomized Trees Regressor (ET Regressor) is a nonlinear regression model based on ensemble learning. Similar to the RF algorithm, ET enhances the overall generalization ability of the model by aggregating the outputs of multiple decision trees. However, it introduces an even higher degree of randomness in both feature selection and split point determination. Specifically, Extra Trees not only randomly select a subset of features at each split node but also randomly generate split thresholds for each selected feature. This strategy accelerates tree construction and reduces model variance, effectively mitigating overfitting while maintaining predictive accuracy.

Unlike traditional decision trees, ET performs the following procedure at each split node: a subset of features is randomly selected (given a total of p input features), and

for each selected feature, several candidate split points are randomly generated. The algorithm computes the split gain for each candidate and selects the one with the highest gain to divide the node. This process substantially improves modeling efficiency and ensemble diversity.

For a given input sample x, the prediction output $f_i(x)$ from the i^{th} tree is computed as the average of the target values of the training samples falling into the corresponding leaf node:

$$f_i(\mathbf{x}) = \frac{1}{n_{i,\mathbf{x}}} \sum_{j=1}^{n_{i,x}} y_j \tag{3.13}$$

where $n_{i,x}$ is the number of training samples in the leaf node, and y_k is the true target value of sample x_k.

The final regression output from the ET ensemble with N trees is the average of predictions from all individual trees:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} f_i(\mathbf{x}) \tag{3.14}$$

By introducing increased randomness in both feature and split point selection at each node, the ET Regressor reduces correlations among decision trees, thereby enhancing the model's generalization ability and robustness. The core idea is that "extreme randomness and the integration of diverse weak learners can enhance nonlinear modeling capability". Extra Trees are capable of efficiently handling high-dimensional and noisy datasets, are less prone to overfitting, and place no special requirements on input features. However, the model's primary limitation lies in its relatively weaker interpretability compared to individual decision trees. Additionally, the high degree of randomness may result in underutilization of certain important features, especially in cases with very small datasets or extremely noisy environments.

4.2.7 CatBoost

As shown in Figure 4.6, CatBoost is a high-efficiency gradient boosting regression model developed by the Yandex team, specifically designed for the automatic handling of categorical features. CatBoost belongs to the Gradient Boosting Decision

Tree (GBDT) family and constructs an ensemble of regression trees to iteratively correct residuals, enabling precise modeling of complex nonlinear relationships.

Compared with algorithms such as XGBoost and LightGBM, the most prominent advantage of CatBoost lies in its native support for categorical features and its strong resistance to overfitting. In each iteration, CatBoost employs a symmetric tree structure (also known as an Oblivious Tree) and introduces techniques such as Ordered Target Statistics (a permutation-based target encoding method) and a permutation-driven randomization strategy. These mechanisms not only improve training speed and model stability but also effectively mitigate the risk of overfitting caused by high-cardinality categorical features. As a result, CatBoost demonstrates exceptional applicability and robustness in practical engineering scenarios—particularly in strength prediction tasks involving material mix designs, where categorical variables and high-dimensional features are common.

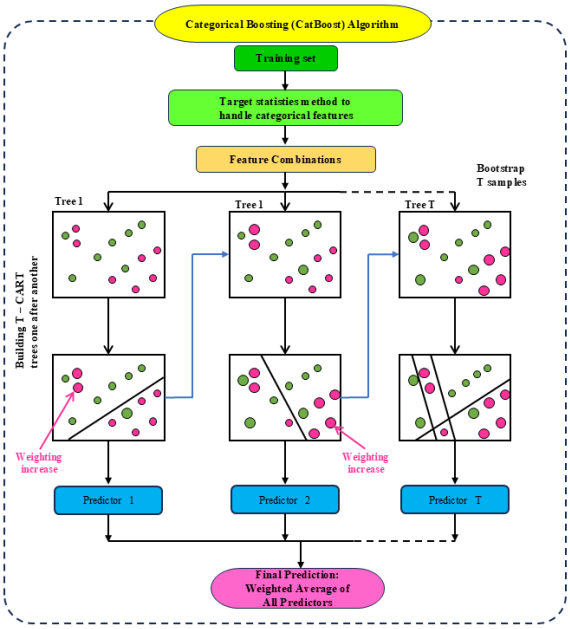


Figure 4.6 CatBoost

Let the training dataset be denoted as $(xi,y_i)i=1^n$, where x_i represents the input features and y_i the corresponding target variable. CatBoost adopts an additive modeling strategy, iteratively constructing a series of base regression trees, each designed to fit the residuals of the current model.

The predicted output at the ttht^{th}tth iteration is defined as:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \nu f_t(\mathbf{x}_i) \tag{3.15}$$

where v is the learning rate, and f_t is the regression tree generated in the t^{th} iteration. The final prediction after aggregating all T regression trees is given by:

$$\hat{y}_i = \sum_{t=1}^{T} \nu f_t(\mathbf{x}_i) \tag{3.16}$$

The CatBoostRegressor utilizes a gradient boosting framework and incorporates innovative mechanisms such as ordered target encoding, symmetric tree structures, and permutation-based randomization to efficiently handle categorical variables natively. This algorithm controls model complexity and mitigates overfitting while enhancing training efficiency and generalization capability, making it particularly suitable for data scenarios with numerous categorical features or high-dimensional heterogeneous inputs.

Its main advantages include automatic handling of categorical variables, strong robustness, and stable predictive performance. However, it also presents certain limitations, such as sensitivity to hyperparameter settings and feature distributions, relatively complex underlying structure, and potentially high memory consumption in cases involving extremely sparse categorical data.

4.3 CVAE-Based Data Enhancement for Remaining Strength Prediction

4.3.1 Evaluation Metrics and Interpretability of Machine Learning Models

For the task of concrete strength prediction, a comprehensive evaluation of regression model performance typically involves multiple statistical metrics to quantify the goodness of fit, error magnitude, and robustness between the predicted and actual values. Common evaluation metrics include the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE). Additional metrics such as median absolute error (Median AE), maximum absolute error (Max AE), and mean absolute percentage error (MAPE) are also employed to provide a more complete assessment of model accuracy. The specific definitions and calculation formulas for these metrics are as follows:

R^2 reflects the proportion of variance in the observations that is explained by the model. Its value ranges from 0 to 1, with values closer to 1 indicating better model fit. The formula is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.17}$$

where n denotes the total number of samples, y_i represents the actual value of the i^{th} sample, and $\hat{y}_i$ denotes the corresponding predicted value, $\bar{y}$ denotes the mean of all true target values.

RMSE is the square root of the Mean Squared Error (MSE), and shares the same unit as the predicted target variable. It intuitively reflects the average deviation between predicted and actual values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.18}$$

MAE measures the average magnitude of the absolute differences between predicted and actual values. It is less sensitive to extreme outliers, computationally simple, and offers good interpretability for engineering applications. A lower MAE indicates smaller overall prediction bias of the model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.19}$$

Mean Squared Error (MSE) measures the average of the squared differences between predicted and actual values, reflecting the overall bias and error level of the model. It is calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.20}$$

Mean Absolute Percentage Error (MAPE) reflects the average percentage error of predictions relative to the true values, providing an intuitive measure of the model's relative error level. It is calculated as:

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3.21}$$

Samples where y_i=0, must be excluded or handled using a specific strategy to avoid division by zero.

Median Absolute Error (Median AE) measures the median of the absolute prediction errors across all samples, reflecting the model's robustness to typical observations and being less sensitive to outliers. It is calculated as:

$$\text{Median AE} = \text{median}\left(|y_1 - \hat{y}_1|, |y_2 - \hat{y}_2|, \ldots, |y_n - \hat{y}_n|\right) \tag{3.22}$$

Maximum Absolute Error (Max AE) represents the largest absolute prediction error among all samples and is used to evaluate the model's worst-case performance. It is calculated as:

$$\text{Max AE} = \max_{1\le i\le n} |y_i - \hat{y}_i| \tag{3.23}$$

In this thesis, R^2 is used to assess the overall goodness-of-fit of the model, while RMSE primarily reflects the dispersion and bias of the prediction results. MAE quantifies the mean prediction error, and Median AE represents the median absolute error, indicating the model's robustness to typical samples. MAPE evaluates the relative error level, and Max AE characterizes the worst-case prediction error. These multiple indicators collectively provide a comprehensive evaluation of the model's accuracy and stability in predicting the compressive strength of GPC, offering a reliable basis for the comparative analysis of different models and feature engineering approaches.

To enhance the interpretability of the machine learning models, the Shapley Additive exPlanations (SHAP) method is introduced to explain the model outputs. SHAP is a model-agnostic interpretability tool grounded in cooperative game theory. Its core idea is to compute the average marginal contribution of each feature across all possible combinations of input features, thereby quantifying the impact of each feature on the prediction.

Unlike traditional feature importance analyses, SHAP provides both global explanations of the model behavior and local explanations for individual prediction. It reveals the direction (positive or negative) and magnitude of each feature's contribution for specific samples. Fundamentally, SHAP decomposes the prediction of a machine learning model into the sum of the baseline value and the weighted contributions of each input feature using an additive model. A schematic SHAP interpretation for three sample inputs is illustrated in Figure 4.7. Its mathematical expression is as follows:

$$g(x') = \phi_0 + \sum_{j=1}^{M} \phi_j x'_j \tag{3.24}$$

where $g(x')$ represents the SHAP additive explanation model, ϕ_0 is the average model output across all samples (i.e., the baseline value), ϕ_j is the Shapley value of the j^{th} feature, indicating its marginal contribution to the prediction, and x_j' denotes the presence (1) or absence (0) of the feature. M is the total number of features.

By decomposing SHAP values across all input features, it is possible not only to obtain a global ranking of feature importance but also to intuitively analyze the direction and magnitude of each feature's influence on individual sample predictions. In addition, SHAP is model-agnostic and can be applied to various mainstream machine learning algorithms, such as XGBoost and Random Forest. Its advantages lie in providing both quantitative and interpretable feature-level explanations for engineering material property prediction (e.g., concrete), while also offering intuitive visualizations. These characteristics make SHAP a valuable tool for model optimization and for guiding material mixture design with theoretical and practical insights.

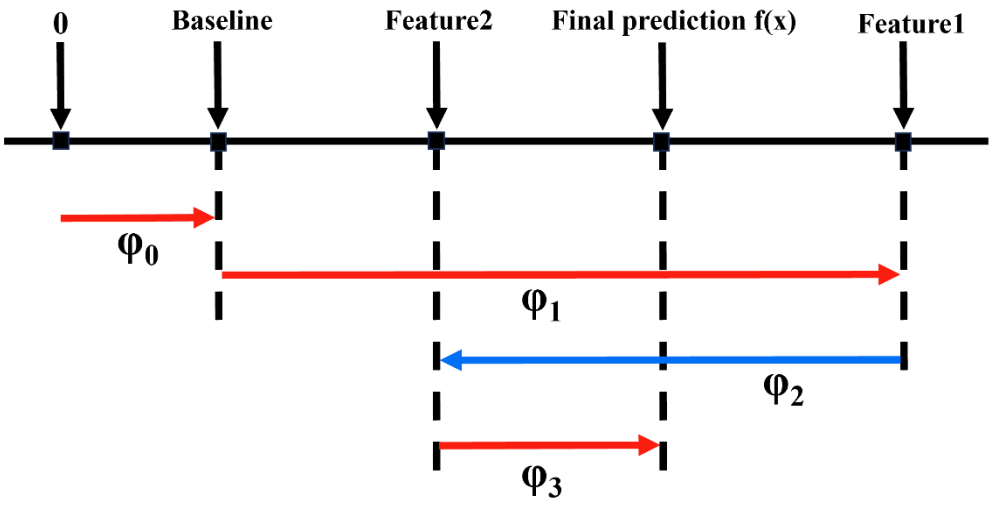


Figure 4.7 SHAP-Based Visualization of Input Feature Contributions

4.3.2 Conditional Variational Autoencoder

In civil engineering, accurately predicting and optimizing material performance, such as the compressive strength of concrete, relies heavily on the expressive power of data features. The relationship between concrete mix proportions and their mechanical properties is highly nonlinear and involves complex multivariate interactions. Traditional methods, including linear dimensionality reduction (e.g., Principal Component Analysis) or manual feature engineering, often fail to capture these intricate patterns.

To address this limitation, this thesis employs CVAE (Sohn, 2015) as a deep feature enhancement tool for complex engineering datasets. As a generative deep learning model, CVAE encodes raw features; such as mix proportions and curing age into a low-dimensional latent space conditioned on known parameters. By reconstructing enhanced features from this latent representation, CVAE effectively captures hidden multivariate relationships, improving prediction accuracy and generalization in concrete strength modeling.

The basic architecture of the CVAE consists of an encoder and a decoder. The encoder takes as input x (original standardized features) and c (conditional variables), and uses a multilayer perceptron to produce the mean and log-variance vectors of the

latent variable, denoted as μ and $log\sigma^2$, respectively. Using the reparameterization trick, a latent vector z is sampled. The decoder then takes (z,c) as input and reconstructs the original feature space through a neural network.

During model training, the objective is to maximize the variational lower bound of the conditional likelihood (ELBO). The loss function is formulated as:

$$\mathcal{L}_{\text{CVAE}} = \mathrm{E}_{\mathrm{q}_\phi(\mathbf{z}|\mathbf{x},\mathbf{c})}[\log \mathrm{p}_\theta(\mathbf{x} \mid \mathbf{z},\mathbf{c})] - \mathrm{D}_{\mathrm{KL}}(\mathrm{q}_\phi(\mathbf{z} \mid \mathbf{x},\mathbf{c}) \tag{3.25}$$

Where θ and ϕ represent the parameters of the decoder and encoder, respectively. The term $q_\phi(z|x,c)$ denotes the encoder's approximation of the posterior distribution of the latent variable, which is assumed in this thesis to follow a conditional Gaussian distribution, with the mean and log-variance output directly by the neural network.

The term $p_\theta(x|z,c)$ models the conditional probability distribution of the decoder over the reconstructed features, which aims to accurately reconstruct the original input given the latent variable z and the condition c. The prior distribution $p(z|c)$ is typically assumed to be a conditionally independent standard Gaussian distribution.

The term D_{KL} represents the Kullback-Leibler (KL) divergence, which measures the difference between the approximate posterior and the prior distribution. It serves to regularize the latent space, ensuring that it remains well-structured and capable of generating meaningful samples.

The model was implemented using the PyTorch framework, with both the encoder and decoder structured as multilayer perceptrons (MLPs). During training, the reconstruction loss is computed by minimizing the Mean Squared Error (MSE), while the KL divergence term is calculated using the closed-form expression for Gaussian distributions. The final total loss function is expressed as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda_{\text{KL}} \cdot \mathcal{L}_{\text{KL}} \tag{3.26}$$

where $\mathcal{L}_{\text{recon}}$ denotes the reconstruction error, $\mathcal{L}_{\text{KL}} = -\frac{1}{2}\sum_{j=1}^{d}\left[1+\log\sigma_j^2-\mu_j^2-\sigma_j^2\right]$ is the KL divergence term, and λ_{KL} is a gradually increasing scaling factor used to balance the two components during training.

Through this approach, the model not only achieves effective reconstruction of the original features during training but also captures the complex influence patterns of mix proportions, testing durations, and other variables on concrete strength under the constraint of conditioning information. After training, the latent variables z produced by the encoder are treated as enhanced features and concatenated with the original mix design parameters to serve as inputs for the subsequent regression modeling. By combining the CVAE-generated latent representations with the original features, the regression model is provided with a more informative and structurally optimized feature set, thereby improving the prediction of concrete compressive strength.

In this thesis, the latent vector z is not interpreted as a direct measurement of a single microstructural parameter. Instead, it serves as a compact, data-driven representation of higher-order interactions among mix constituents, exposure age, and compressive strength under the specified sulfate-exposure protocol. From a materials perspective, the latent dimensions can be viewed as aggregating coupled effects that govern strength evolution in alkali-activated systems, such as the effective binder reactivity (MK/GGBFS chemistry), reaction progress and gel formation, and the development of damage-related voids and microcracking with exposure duration. Accordingly, although no one-to-one mapping is imposed, the latent space remains physically plausible because it is learned from physically grounded inputs and constrained by reconstruction of the observed data. To limit unrealistic extrapolation, synthetic samples are generated by sampling within the learned conditional latent distribution and are used only for training-set augmentation.

In the implementation, the CVAE model was built using the PyTorch framework. The network architecture and hyperparameter settings are as follows: the encoder consists of four fully connected layers with 128, 64, 32, and 32 neurons, respectively.

Batch normalization and LeakyReLU activation functions are applied to the first two layers to enhance non-linear representation and training stability. Additionally, a dropout rate of 0.15 is used to mitigate overfitting. The decoder adopts a symmetric structure, taking as input the concatenated latent and conditional variables. The decoder maps this input through three hidden layers with 64, 128, and 64 neurons, respectively, and employs a combination of ReLU and LeakyReLU activations.

The latent space dimension is set to 8. The conditional variables include metakaolin and GGBFS, selected as core influencing factors of the mixture, to effectively capture the coupling relationship affecting compressive strength in GPC.

As shown in Table 4.1, the training process uses the Adam optimizer with a learning rate of 1e-3, a batch size of 32, and a total of 1000 training epochs. The loss function comprises the reconstruction error (mean squared error) and KL divergence, with the KL weight (λ) linearly increased during training (up to a maximum of 0.1). This progressive weighting allows the model to focus on accurate reconstruction in early training stages, while gradually enforcing structure regularization on the latent space in later stages, thereby producing more interpretable and generalizable latent representations.

Table 4.1 CVAE Hyperparameter Configuration

Parameter	Value/Type	Description
Optimizer	Adam	Optimizer used for training deep networks
Learning Rate	0.001	Controls the speed of parameter updates
KL Weight	Linearly increasing to 0.1	Balances reconstruction accuracy and latent space regularization
Loss Function	MSE + λ × KL	Combines reconstruction error and KL divergence
Batch Size	32	Number of samples per training batch
Number of Epochs	1000	Total number of training iterations

CVAE-Regression is an innovative hybrid optimization learning framework that combines Conditional Variational Autoencoder (CVAE) with Optuna-based

Bayesian automatic optimization to enhance feature representation and perform hyperparameter tuning, thereby improving regression model performance. The overall structure is illustrated in Figure 4.8.

First, the CVAE is utilized to perform deep feature enhancement on the original concrete mix design data, enabling effective reconstruction and representation of the nonlinear coupling relationships inherent in the material properties. The generated latent variables further enrich the original feature space. Then, Optuna is introduced for automatic feature subset selection and hyperparameter optimization. Leveraging a Bayesian optimization strategy, Optuna automatically searches for the optimal combination of input features and model parameters by minimizing cross-validation loss. Through large-scale and efficient sampling and iteration, this approach enables comprehensive exploration of the feature and parameter space, achieving globally optimal solutions in complex engineering data scenarios.

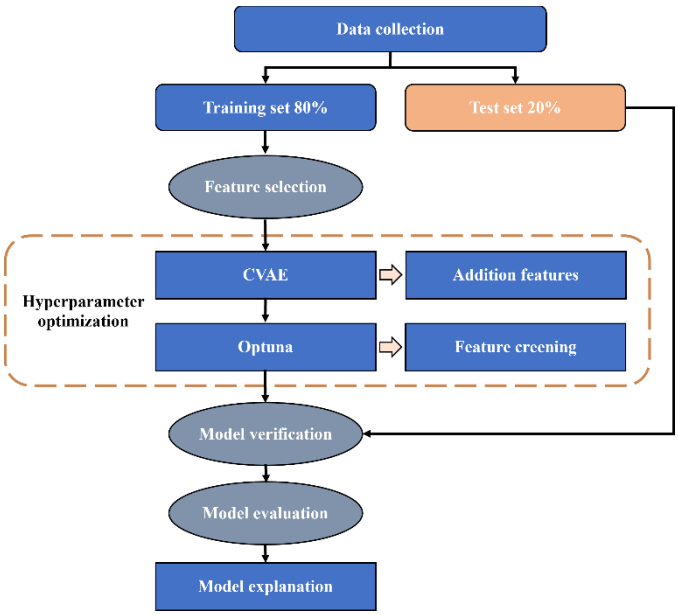


Figure 4.8 CVAE-Regression Model Regression Flow Diagram

4.4 Summary

In this chapter, the intelligent prediction of compressive strength for GPC was systematically investigated. The main conclusions are summarized as follows:

1. A range of mainstream machine learning regression algorithms—including KNN, SVR, RF, XGBoost, MLP, ET, and CatBoost—were reviewed and analyzed in terms of their theoretical foundations, mathematical formulations, and suitability for complex material datasets. Their strengths and limitations were discussed in the context of high-dimensional, nonlinear, and strongly coupled relationships inherent in GPC data.

2. The SHAP (SHapley Additive exPlanations) method was applied to interpret model outputs both globally and locally, revealing the physical significance of key features and improving the transparency and reliability of model predictions.

3. To overcome the limitations of traditional models, a hybrid strategy was proposed by integrating Conditional Variational Autoencoder (CVAE) for deep feature enhancement and Optuna for automated hyperparameter and feature subset optimization. This framework enables the model to capture complex nonlinear relationships more effectively and supports efficient optimization in high-dimensional engineering scenarios.

4. The proposed CVAE-Optuna modeling architecture lays a theoretical and methodological foundation for improving predictive performance in subsequent chapters. It also offers a novel technical route for intelligent mix design, high-precision performance forecasting, and the data-driven development of sustainable construction materials.

Chapter 5. MODEL DEVELOPMENT, TRAINING, AND PERFORMANCE EVALUATION FOR PREDICTING COMPRESSIVE STRENGTH

5.1 Introduction

Building on the theoretical framework and algorithm selection in Chapter 4, this chapter details the practical implementation of machine learning models to predict the compressive strength degradation of GPC exposed to marine sulfate. The goal is to develop a robust, data-driven approach that accurately captures strength trends over time using mix design parameters and exposure duration as inputs. In this thesis, the modelling task focuses on strength degradation under sulfate exposure, where strength retention is used as a durability-related performance indicator to support trend prediction over time. The chapter first outlines the dataset, including its sources, feature composition, and preprocessing steps (normalization, encoding, and cleaning). Next, the model development process is described, covering training-validation splits, hyperparameter optimization (using Optuna), and performance evaluation metrics.

A key focus is the CVAE framework introduced in Chapter 4. This section explains CVAE training, synthetic sample generation for data augmentation, and the resulting improvement in regression model generalization. Comparative analyses between baseline models (RF, XGBoost, MLP) and their CVAE-augmented counterparts demonstrate performance enhancements.

Model effectiveness is illustrated through predicted vs. actual strength plots, residual distributions, and degradation curves. Further insights are provided via feature importance analysis and robustness assessments. The chapter concludes with a comprehensive evaluation of the modeling framework, demonstrating how advanced ML, particularly CVAE enhances predictive accuracy for GPC strength degradation, even under the sulfate-exposure conditions represented in the dataset and limited data availability.

5.2 Dataset Description and Preprocessing

5.2.1 Dataset Sources and Selection

To accurately predict compressive strength changes in GPC under erosive conditions, a robust and high-quality experimental database is essential. This thesis compiled a database of 369 mix designs from 36 peer-reviewed journal papers, ensuring consistency by selecting only data where specimens underwent standard 28-day curing at ambient temperature before exposure to 5% Na_2SO_4 solution. This standardized approach guarantees comparability and reliability across datasets. The data were randomly split into an 80% training set and a 20% testing set, with the latter kept entirely separate from model development to enable unbiased evaluation of predictive generalization on unseen data.

Figure 5.1 presents the distributions of the input features, showing substantial variability in both mix compositions and exposure ages across the compiled studies. This spread indicates that the dataset spans a broad spectrum of reported mix proportions and test durations under the selected 5% Na_2SO_4 exposure protocol, supporting model development within the evidence base currently available. Future studies may further enhance generalizability by incorporating additional exposure descriptors when they are consistently reported.

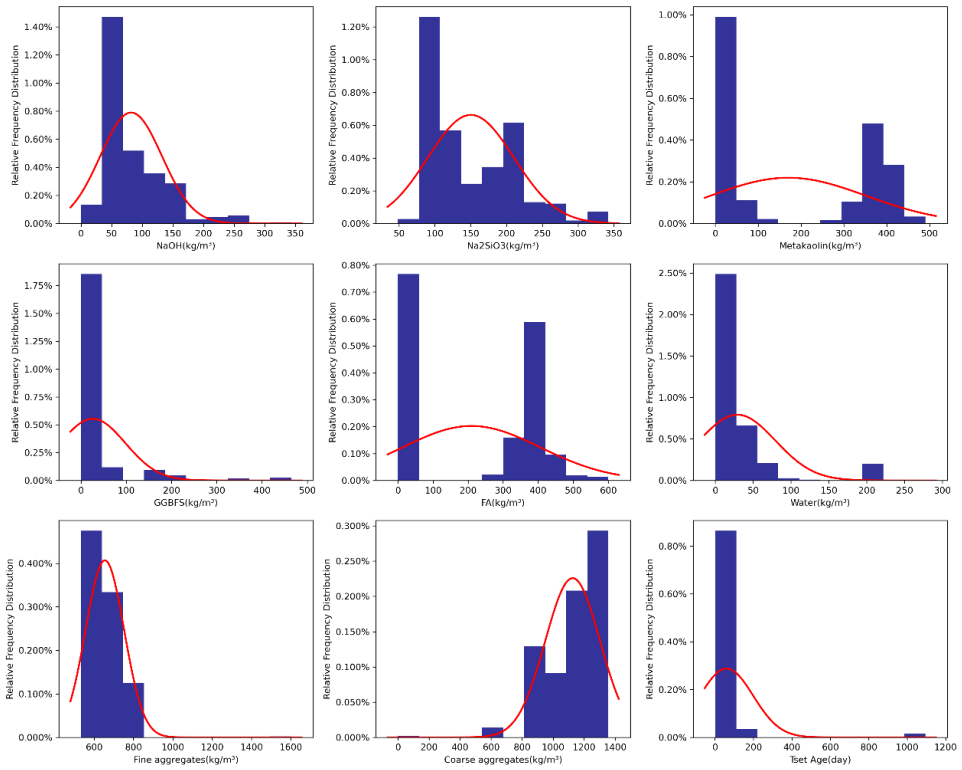


Figure 5.1 Distribution of Key Input Features in the Dataset

Table 5.1 summarizes the parameters included in the dataset along with their respective ranges and mean values.

Table 5.1 Descriptive statistics of the input and output parameters for predicting compressive strength

Variable	Minimum	Maximum	Mean	Category
NaOH (kg/m³)	0	343	82.04	Input
Na_2SiO_3 (kg/m³)	48.5	342	150.31	Input
Metakaolin (kg/m³)	0	490.3	170.76	Input
GGBFS (kg/m³)	0	464	25.97	Input
FA (kg/m³)	0	600	208.92	Input
Water (kg/m³)	0	277.5	29.02	Input
Fine aggregates (kg/m³)	533	1600.8	655.05	Input
Coarse aggregates (kg/m³)	0	1356	1125.49	Input
Test Age (day)	1	1095	59.64	Input
Fc (MPa)	6	89	37.27	Output

5.2.2 Feature Selection

Engineering datasets are often characterized by high dimensionality and limited sample sizes. This inherent discrepancy presents significant challenges to the interpretability of machine learning outcomes, necessitating robust feature selection methodologies. Effective feature selection facilitates the removal of redundant or non-informative variables while identifying those most strongly correlated with the target property. Consequently, this process enhances model predictive accuracy, generalizability, and overall interpretability.

Table 5.2 presents the Pearson correlation coefficients between each feature and the target variable. The Pearson correlation coefficient ranges from −1 to 1, where −1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation. The Pearson correlation coefficient is calculated as follows:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.1}$$

where x_i and y_i denote the feature value and the target value of the i sample, and $\bar{X}$ and $\bar{Y}$ are their respective means.

Based on the correlation analysis results, the nine most relevant features were selected as model inputs: NaOH, Na_2SiO_3, Metakaolin, GGBFS, FA, Water, Fine Aggregates, Coarse Aggregates, and Test Age. These features carry rich information, better reflect the variation of the target variable, and help reduce the interference from irrelevant features, thus improving the training efficiency and prediction accuracy.

The selected features and their correlations are visualized in the correlation matrix heatmap shown in Figure 5.2. The closer the correlation coefficient is to 1 or −1, the stronger the relationship between the two variables, as indicated by a darker color. Conversely, weaker correlations are represented by values closer to 0 and lighter colors.

Table 5.2 Pearson correlation coefficient between features and targets

Features	Pearson's correlation coefficient
NaOH	-0.06
Na_2SiO_3	-0.19
Metakaolin	-0.37
GGBFS	0.44
FA	0.34
Water	-0.30
Fine aggregates	-0.09
Coarse aggregates	-0.10
Tset Age	-0.10

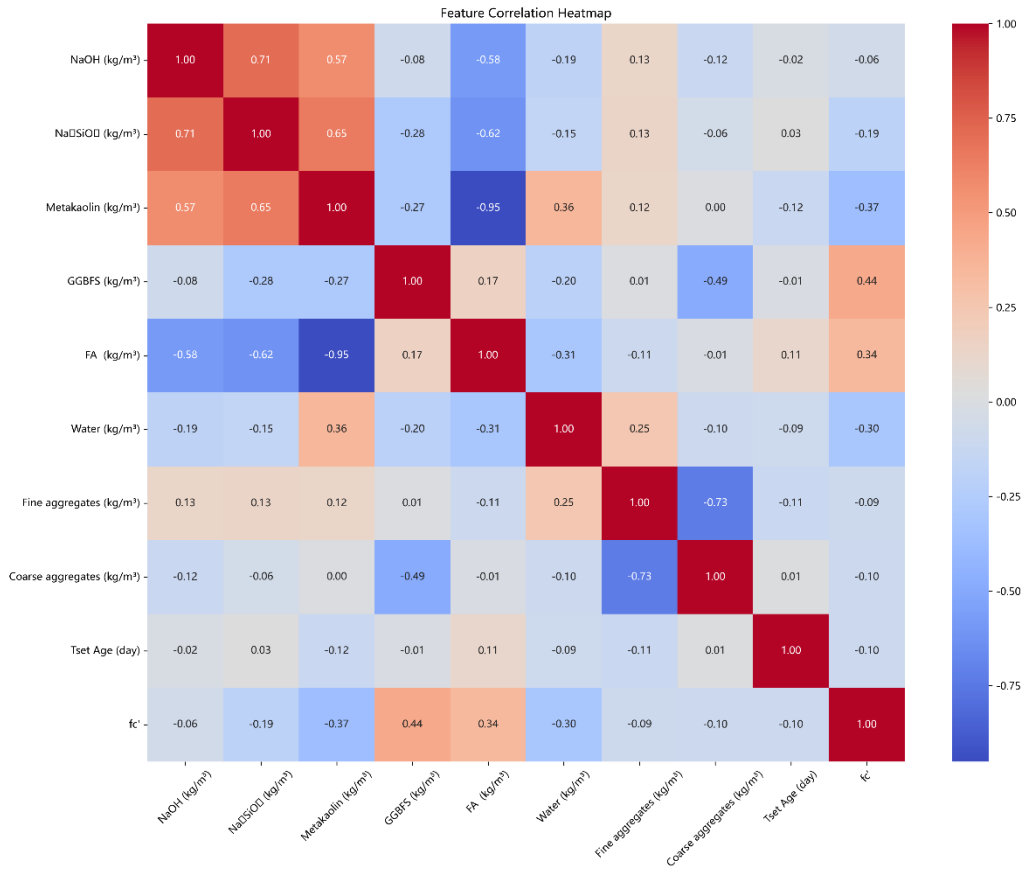


Figure 5.2 Feature Correlation Heatmap

5.3 Model Evaluation and Comparison Based on CVAE

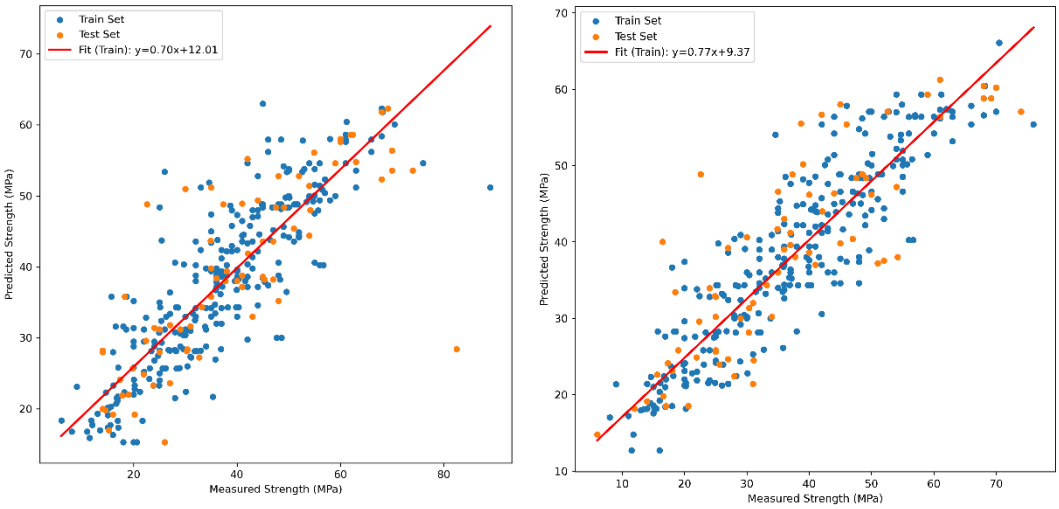
All model training and evaluation tasks were conducted on a workstation equipped with a 13th Gen Intel® Core™ i7-13700KF CPU and an NVIDIA® GeForce RTX 4090 GPU. This high-performance hardware configuration ensured efficient execution of deep learning-based feature augmentation (CVAE), large-scale cross-validation, and Bayesian optimization (Optuna) during the model development process.

5.3.1 K-Nearest Neighbors model

KNN is a typical instance-based regression algorithm that predicts the compressive strength of GPC by measuring the distance relationships between samples. To visually compare the impact of different feature engineering approaches on the predictive performance of the KNN model, this thesis presents a side-by-side comparison of prediction scatter plots and point-to-point prediction curves for the conventional KNN method and the enhanced KNN model based on CVAE and Optuna, as shown in Figure 5.3.

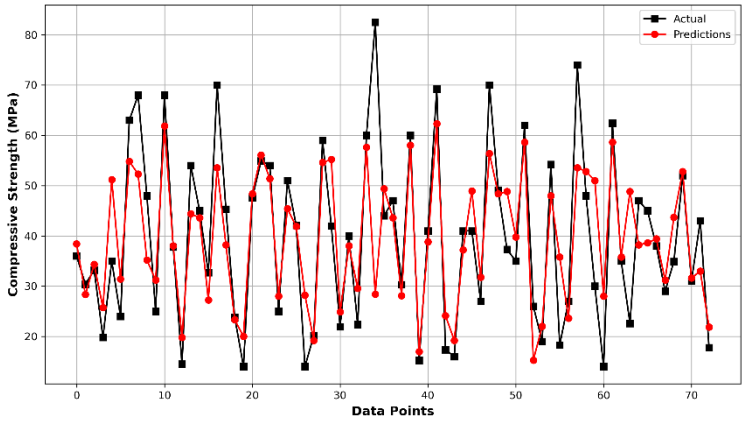
In the scatter plot comparison, the conventional KNN model exhibits a more dispersed distribution of data points with a tendency to underestimate compressive strength in the high-strength region. The fitted regression line presents a slope notably below 1, indicating a systematic bias in prediction. In contrast, the CVAE+Optuna-enhanced KNN model yields predictions that align more closely with the ideal diagonal line, characterized by a slope closer to 1 and a smaller intercept. This improvement reflects better agreement with actual compressive strength values and improved consistency in prediction.

As illustrated in Figure 5.4, the point-to-point prediction curve of the traditional KNN model deviates from the actual strength trend, particularly around extreme values, and shows higher overall prediction fluctuations. In comparison, the CVAE+Optuna-enhanced KNN model yields a prediction curve that closely follows the actual curve, capturing subtle variations in strength data more effectively and improving both fitting accuracy and robustness.

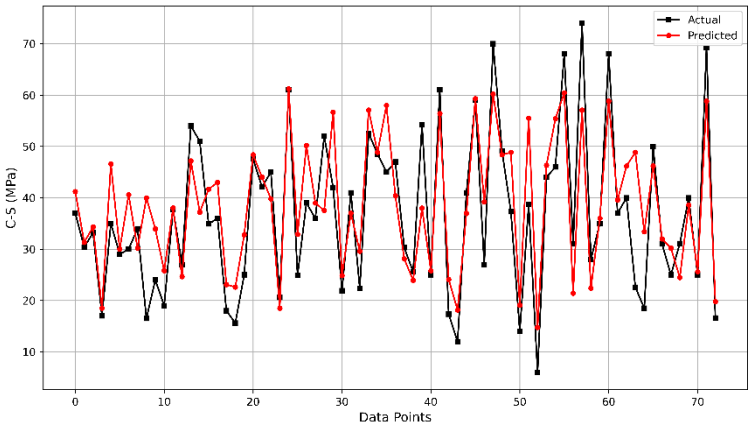


a. KNN b. CVAE+KNN

Figure 5.3 Scatter Plot of KNN Prediction Results



a. KNN



b. CVAE+KNN

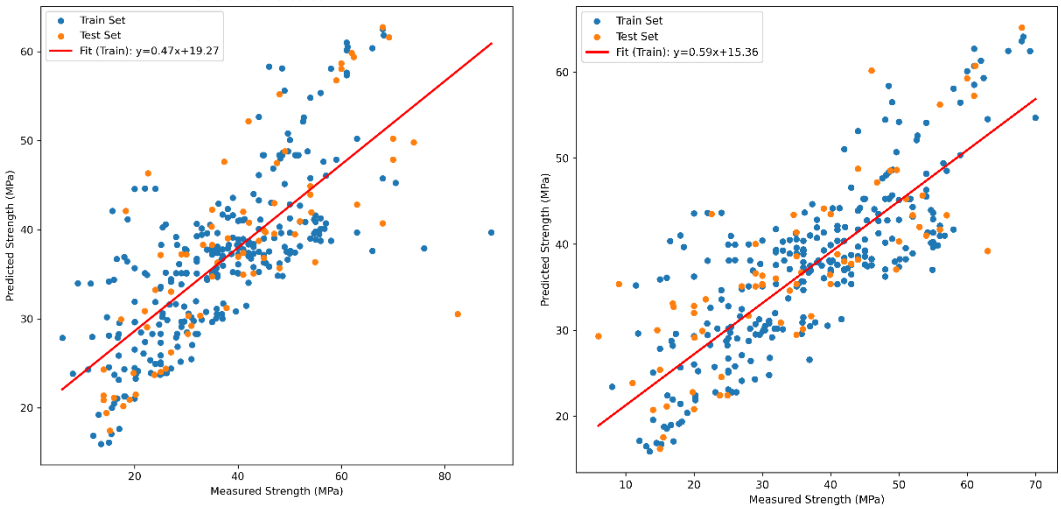
Figure 5.4 Comparison of Prediction Curves for KNN Models

5.3.2 Support Vector Regression model

SVR is a regression method based on the theory of Support Vector Machines, particularly effective in handling small-sample and high-dimensional datasets. In this thesis, the prediction performance of the conventional SVR model and the enhanced SVR model integrated with CVAE and Optuna was compared to demonstrate their applicability in predicting the compressive strength of GPC.

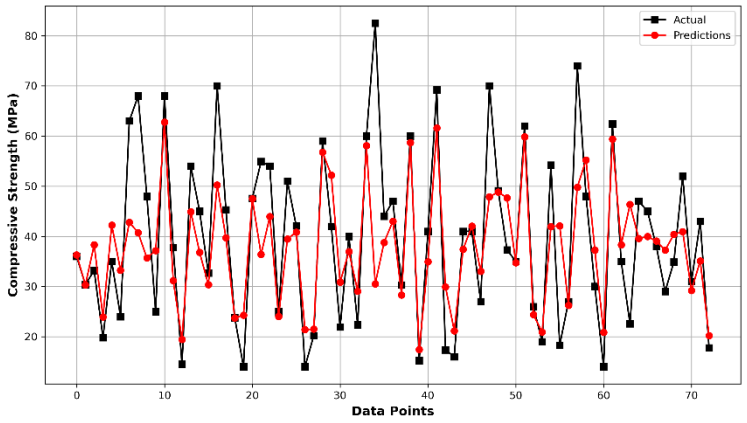
As shown in Figure 5.5 and Figure 5.6, both the scatter plots and the point-to-point trend curves reveal that the conventional SVR model exhibits a relatively dispersed distribution of predicted values. The fitted regression line has a low slope ($y=0.47x+19.27$), and underestimation is evident, especially in the high-strength range. Many points in the test set deviate noticeably from the ideal diagonal line, indicating limited overall fitting capability. In the point-to-point prediction curves, the conventional SVR model displays considerable fluctuation compared to the actual values and fails to effectively capture extreme values and trend variations.

In contrast, the SVR model enhanced by CVAE and Optuna shows significant improvements. In the scatter plot, the predicted points are more tightly clustered along the diagonal line, with a higher regression slope ($y=0.59x+15.36$), indicating that the model more accurately captures the true compressive strength and exhibits improved generalization ability. In the trend comparison, the enhanced SVR model closely follows the actual strength curve, effectively capturing both extreme values and fluctuation patterns. Overall prediction error is reduced, reflecting superior fitting accuracy and robustness.

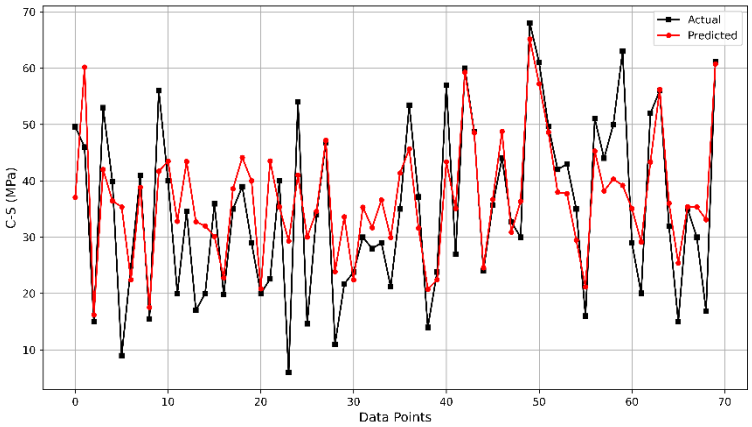


a. SVR b. CVAE+SVR

Figure 5.5 Scatter Plot of the SVR Model



a. SVR



b. CVAE+SVR

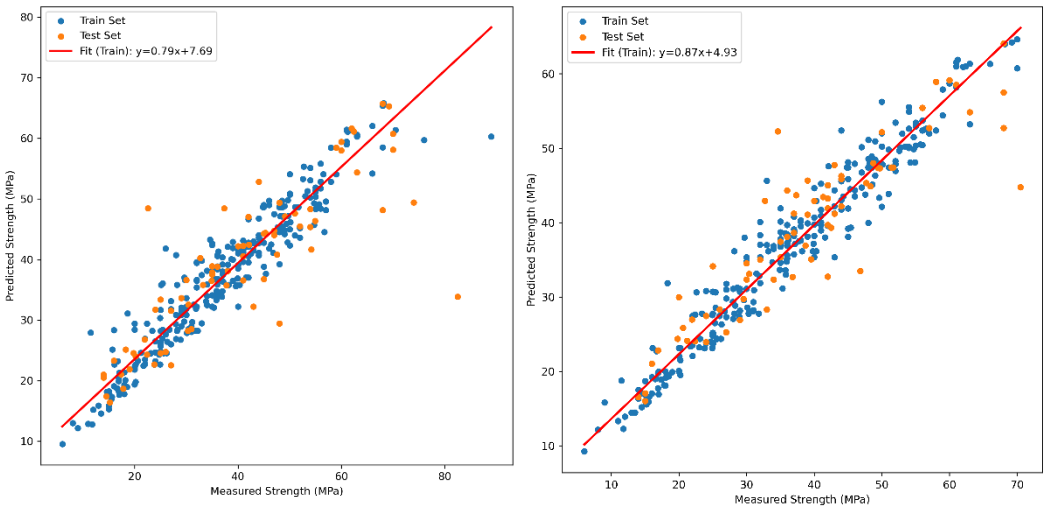
Figure 5.6 Comparison of Prediction Curves for SVR Models

5.3.3 Random Forest

RF is a nonlinear regression algorithm that integrates multiple decision trees and is known for its strong nonlinear fitting capability and robustness to noise. To evaluate the impact of different feature engineering methods on the predictive performance of the RF model, this thesis compares the prediction scatter plots and point-to-point trend curves of the conventional RF model and the enhanced RF model optimized with CVAE and Optuna. Figure 5.7 and Figure 5.8 show the prediction scatter plot and trend curve for the RF model, respectively.

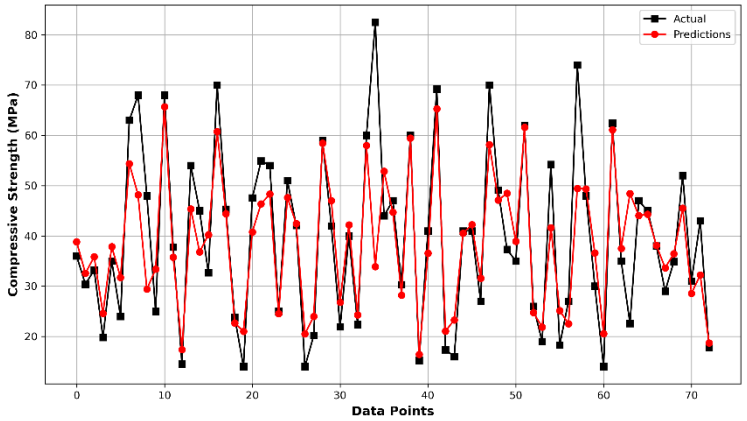
According to the comparison results, the scatter plot of the conventional RF model shows that the predicted points are generally distributed around the diagonal line. However, its fitting performance in the high-strength region is moderate, with a fitted regression line slope (y=0.79x+7.69) slightly lower than ideal. Some test points deviate from the fitted line, indicating a lack of precision in certain regions. In the point-to-point trend curve, the conventional RF model can largely follow the variation in actual compressive strength but still exhibits some deviation in extreme values and localized fluctuations, suggesting that its fitting performance in complex intervals needs improvement.

In contrast, the RF model enhanced with CVAE and Optuna demonstrates significantly better performance. The scatter plot shows a more concentrated and tightly distributed point cloud near the diagonal line, with a noticeably improved regression slope (y=0.87x+4.93). This indicates higher consistency and accuracy in the predictions. In the trend curve, the enhanced RF model more precisely tracks the fluctuations in actual strength, better captures extreme values and detailed variations, and further reduces prediction errors.

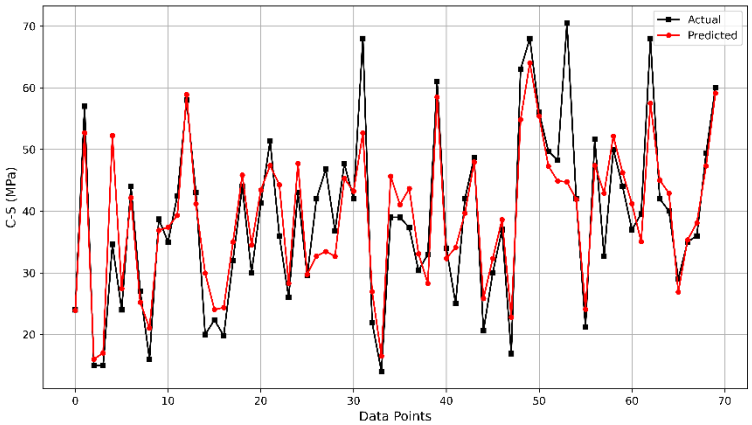


a. RF b. CVAE+RF

Figure 5.7 Scatter Plot of the RF Model



a. RF



b. CVAE+RF

Figure 5.8 Comparison of Prediction Curves for RF Models

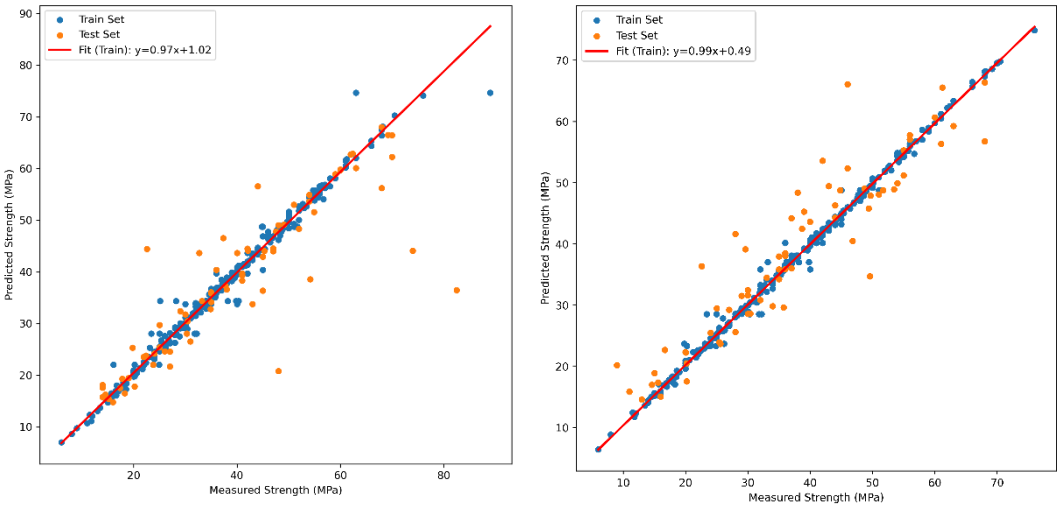
5.3.4 XGBoost model

XGBoost is an ensemble regression algorithm based on gradient-boosted decision trees, widely adopted in material property prediction due to its efficiency and strong

generalization capability. To systematically evaluate the impact of different feature engineering methods on the predictive performance of the XGBoost model, this thesis presents a visual comparison between the conventional XGBoost model and the enhanced XGBoost model optimized using CVAE and Optuna, based on their prediction scatter plots and point-to-point trend curves.

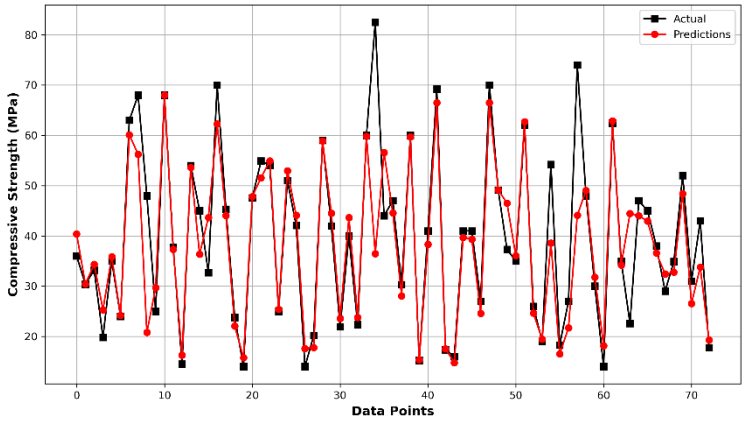
As shown in Figure 5.9 and Figure 5.10, the comparison results indicate that although the prediction points of the traditional XGBoost model are generally distributed along the diagonal in the scatter plot, there are some deviations in specific regions and slight underestimation in extreme values. The fitted regression line has a slope slightly less than 1 ($y=0.97x+1.02$), with some test samples deviating on both sides of the ideal fitting line, and local errors still present. In the point-to-point trend curve, the traditional XGBoost model shows a generally consistent trend with the actual values, but the fitting accuracy and detail in extreme cases could still be improved.

In contrast, the CVAE+Optuna-enhanced XGBoost model demonstrates a noticeable improvement in predictive performance. The scatter plot shows that predicted points are more densely concentrated along the diagonal line, and the slope of the fitted line is closer to 1 ($y=0.99x+0.49$). The distribution of points is tighter, indicating better consistency between the training and test sets. In the trend curve, the enhanced predictions align closely with the actual strength curve, not only capturing the overall trend accurately but also effectively reflecting extreme values and local fluctuations. The robustness and generalization capability of the model are improved.

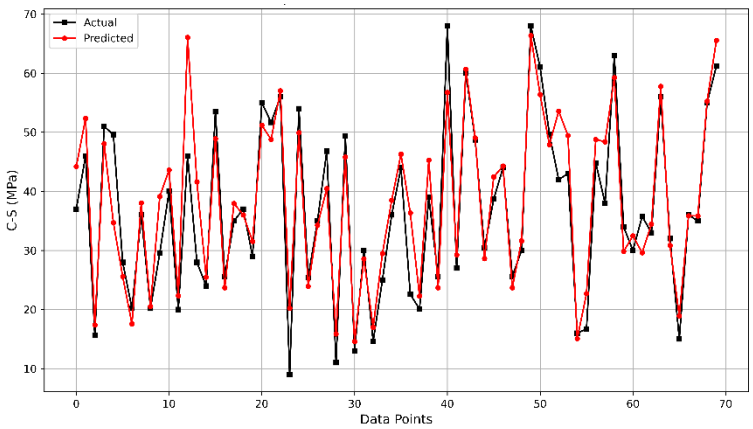


a. XGB b. CVAE+XGB

Figure 5.9 Scatter Plot of the XGB Model



a. XGBoost



b. CVAE+XGBoost

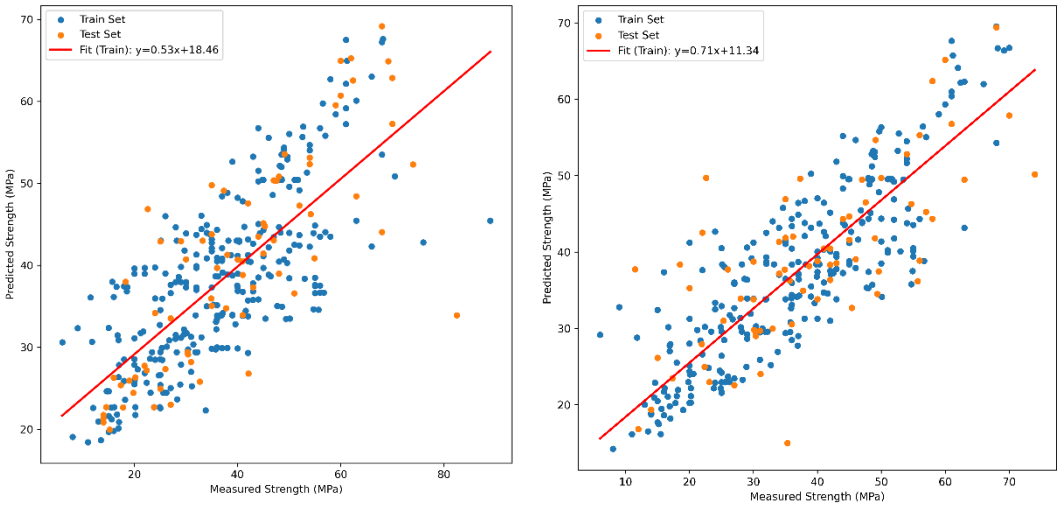
Figure 5.10 Comparison of Prediction Curves for XGB Models

5.3.5 MLP model

MLP is a typical feedforward neural network regression algorithm with strong nonlinear modeling and feature representation capabilities. To systematically evaluate the impact of different feature engineering strategies on the prediction performance of the MLP model, this thesis compares the traditional MLP approach with a CVAE- and Optuna-enhanced MLP model using both scatter plots and point-to-point trend graphs, as shown in Figure 5.11 and Figure 5.12.

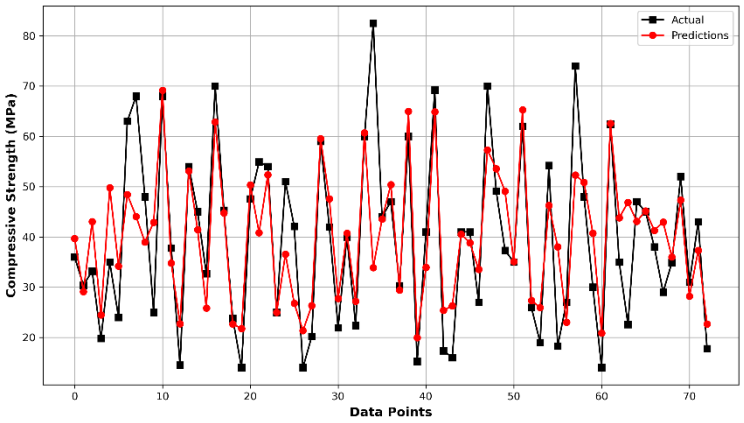
According to the results, the scatter plot of the traditional MLP model shows that the predicted points generally follow a diagonal distribution, yet they are relatively dispersed. A notable underestimation is observed in the high-strength range, and the slope of the fitted line ($y = 0.71x + 11.34$) is markedly lower than the ideal value, suggesting limited capacity to fit outliers and extreme values. In the point-to-point trend graph, while the model captures the general variation trend of the actual compressive strength, considerable discrepancies exist between the predicted and measured curves. These deviations reflect insufficient capability in restoring extreme values and local fluctuations, indicating a lack of robustness.

In comparison, the MLP model augmented with CVAE exhibits improved predictive performance. Its scatter plot reveals a denser clustering of predicted points around the diagonal line, indicating enhanced accuracy and consistency. In the point-to-point trend graph, the enhanced model more closely tracks the actual strength variation, particularly in extreme value regions and local details, with reduced prediction errors. These improvements reflect a stronger generalization ability and better applicability to practical scenarios. While MLP possesses inherent nonlinear representation capabilities suited for material strength prediction, conventional direct modeling approaches are often prone to overfitting and limited generalizability. By incorporating CVAE-based latent feature enhancement and automated optimization strategies, the overall robustness and predictive accuracy of the model are substantially improved.

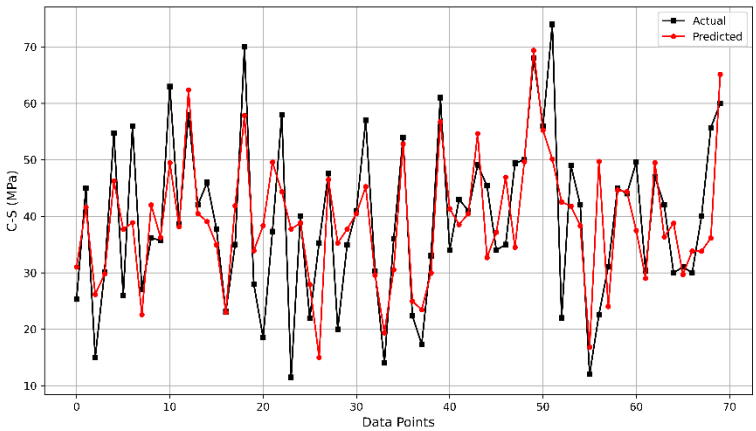


a. MLP b. CVAE+MLP

Figure 5.11 Scatter Plot of the MLP Model



a. MLP



b. CVAE+MLP

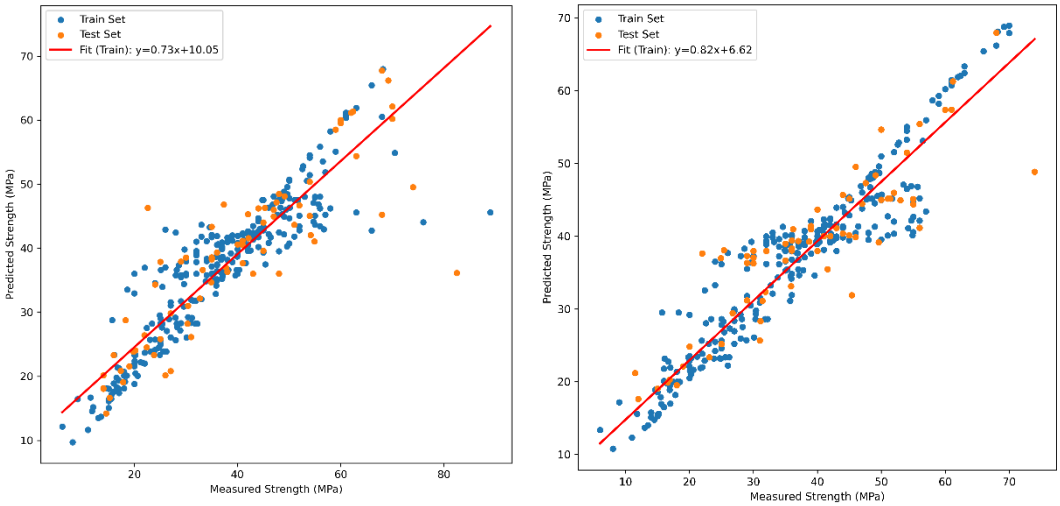
Figure 5.12 Comparison of Prediction Curves for MLP Models

5.3.6 ET model

ET Regressor, as a nonlinear ensemble regression algorithm composed of multiple extremely randomized trees, demonstrates strong capabilities in modeling complex nonlinear relationships and resisting noise. To comprehensively evaluate the impact of different feature engineering strategies on the prediction performance of the ET model, this thesis compares the traditional ET model with the enhanced ET model incorporating CVAE-based feature augmentation and Optuna optimization, as illustrated in Figure 5.13 and Figure 5.14.

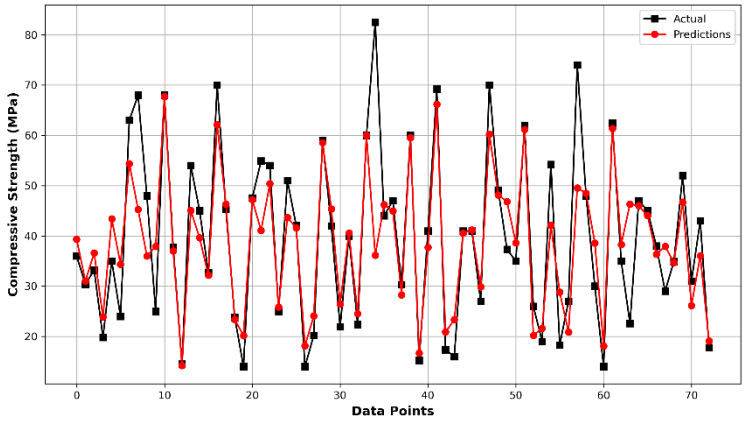
The results show that in the traditional ET model, the scatter plot reveals that most predicted points lie close to the diagonal, but noticeable prediction errors occur in the high-strength range, where some test samples deviate from the expected distribution. The data points are slightly dispersed, with a fitted regression line of y=0.73x+10.05, indicating suboptimal fitting performance. In the point-to-point trend plot, while the traditional ET model generally follows the actual strength trends, its predictions of extreme values and local fluctuations still exhibit certain deviations, suggesting room for improvement in local fitting precision.

In contrast, the CVAE-enhanced ET model demonstrates significantly better performance. The scatter plot shows predicted points more tightly clustered around the diagonal line, with improved point cloud convergence. The regression line improves to y=0.82x+6.62, indicating enhanced consistency and fitting accuracy. In the point-to-point trend plot, the enhanced model tracks actual strength variations more accurately, especially in extreme value and high-variability regions, showing notable improvements in robustness and generalization capability. In summary, although the ET model already exhibits solid nonlinear fitting ability and resilience due to its ensemble structure, it still suffers from prediction errors in direct modeling. With the introduction of CVAE-based deep feature augmentation and automated optimization via Optuna, the model's prediction accuracy, consistency, and generalization are all improved.

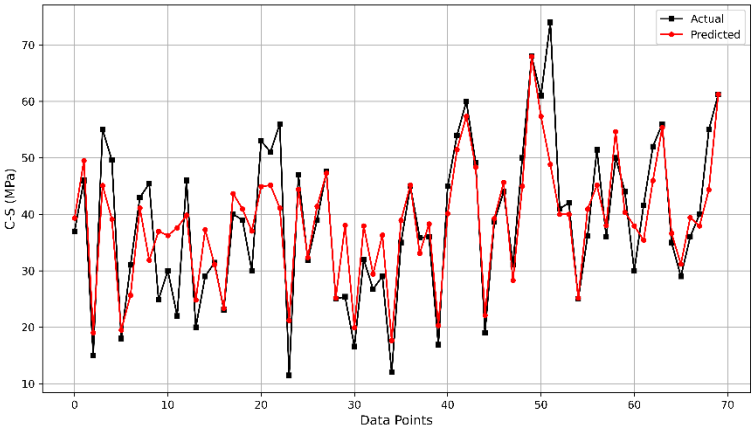


a. ET b. CVAE+ET

Figure 5.13 Scatter Plot of the ET Model



a. ET



b. CVAE+ET

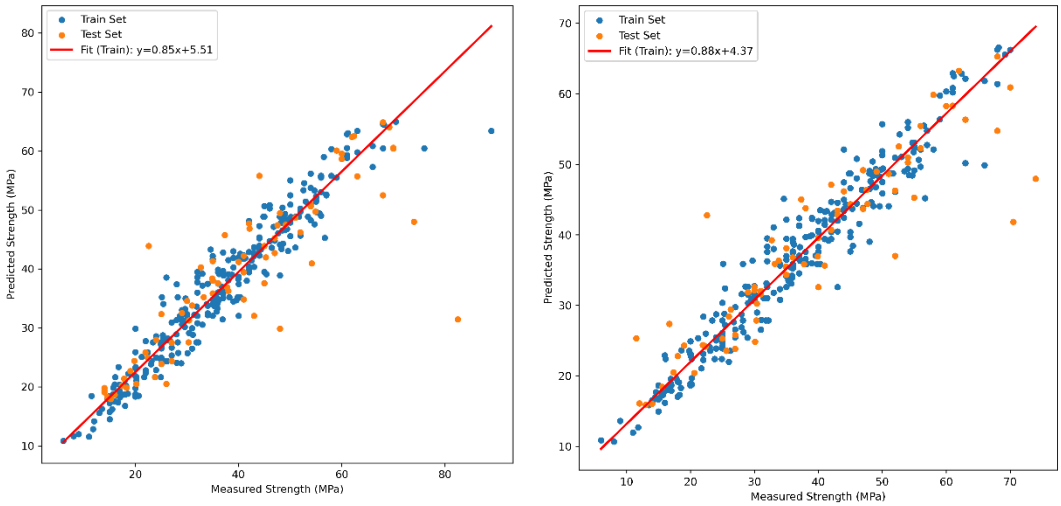
Figure 5.14 Comparison of Prediction Curves for ET Models

5.3.7 CatBoost model

CatBoost, a high-performance gradient boosting tree algorithm optimized for categorical features, is known for its strong nonlinear modeling capability and robustness. To systematically evaluate the effect of different feature engineering strategies on the predictive performance of CatBoost, this section compares the traditional CatBoost model with the CVAE-enhanced CatBoost model using scatter plots and point-to-point trend plots, as shown in Figure 5.15 and Figure 5.16.

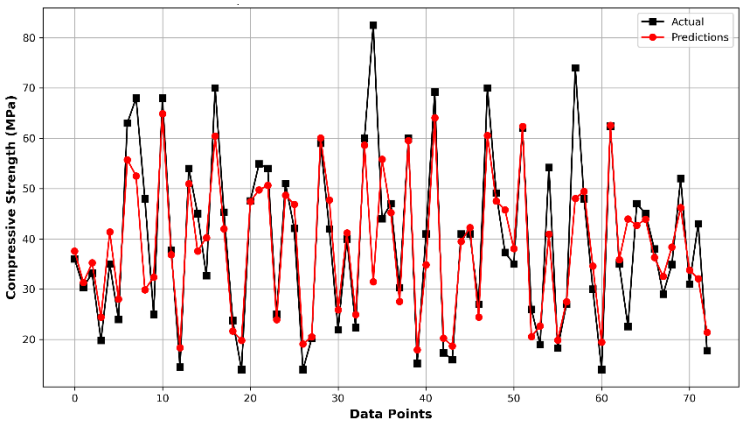
From the scatter plot analysis, the traditional CatBoost model shows predicted values generally distributed near the ideal diagonal, with a fitted line of y=0.85x+5.51, indicating a good representation of actual strength. However, noticeable prediction deviations still exist in the high-strength range, and several test samples exhibit large fluctuations. The point-to-point trend plot further illustrates that while the traditional CatBoost model follows the overall strength trend reasonably well, it struggles to capture extreme values and local variations with sufficient accuracy.

In contrast, the CVAE-enhanced CatBoost model shows clear performance improvements. The scatter plot reveals that predicted points are more tightly clustered along the diagonal, with an improved regression line of y=0.88x+4.37, indicating enhanced fitting ability and prediction consistency. In the trend plot, the optimized model more accurately captures the actual variation in compressive strength, with substantial improvements in fitting extreme values and local fluctuations. The generalization and robustness of the model are enhanced. Overall, the CatBoost model demonstrates excellent nonlinear fitting and noise resistance in predicting the compressive strength of GPC. When combined with CVAE-based deep feature augmentation and Optuna-based optimization, the model achieves further improvements in prediction accuracy, consistency, and practical applicability in engineering scenarios.

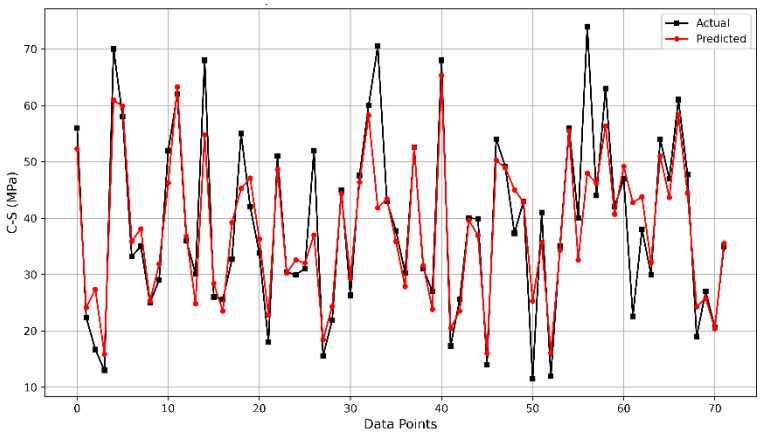


a. CatBoost b. CVAE+ CatBoost

Figure 5.15 Scatter Plot of the CatBoost Model



a. CatBoost



b. CVAE+ CatBoost

Figure 5.16 Comparison of Prediction Curves for CatBoost Models

5.4 Comparative Analysis of Strength Prediction Model Results

To evaluate the effectiveness of different machine learning models and feature optimization strategies in improving compressive strength prediction for GPC, this thesis conducts a systematic comparison between traditional regression models and models enhanced by CVAE. The performance metrics are presented in Tables 5.3 and 5.4.

Across the seven models, CVAE-based feature augmentation improved test-set performance for most models (KNN, SVR, RF, ET, CatBoost, XGBoost); the MLP showed a slight deterioration in R^2 and RMSE, although Max AE improved marginally.

The CVAE-XGBoost model demonstrated the highest predictive accuracy, achieving an R^2 of 0.8572, along with significant reductions in error metrics (RMSE = 5.83, MAE = 3.06). Compared to the baseline XGBoost model (R^2 = 0.7387, RMSE = 8.70, MAE = 4.43), this represents an absolute improvement of 16.0% in R^2, alongside 32.9% and 30.9% reductions in RMSE and MAE, respectively. Notably, RMSE exhibited the most substantial decrease, highlighting the model's enhanced precision in error minimization. Across all seven models, the integration of CVAE-based feature augmentation led to consistent improvements in key evaluation metrics. On average, R^2 increased by 9.1%, while RMSE and MAE decreased by 20.4% and 14.0%, respectively. Although traditional neural networks (MLP) and kernel-based methods (SVR) showed only moderate responsiveness to deep feature representations, they still exhibited improvements in secondary metrics such as maximum absolute error (Max AE), suggesting potential robustness enhancements.

The CVAE-XGBoost model also excelled on the training set, achieving an R^2 of 0.9892, with RMSE and MAE reduced to 1.47 and 0.89, respectively. These results further validate the efficacy of deep feature augmentation combined with Bayesian optimization in enhancing model fitting capacity. Other ensemble models, including XGBoost, RF, ET, and CatBoost, exhibited notable improvements in training performance. Specifically, XGBoost and CatBoost achieved R^2 values exceeding

0.92, accompanied by low RMSE and MAE values. In contrast, simpler models such as KNN and MLP exhibited limited gains from feature augmentation, with some showing mild overfitting, whereas SVR remained comparatively stable without clear overfitting.

The CVAE-based deep feature augmentation method significantly improves both the accuracy and robustness of machine learning models in predicting GPC compressive strength. This approach is particularly effective for ensemble learning models (XGBoost, RF, CatBoost), underscoring the practical utility of deep generative features and automated hyperparameter optimization in intelligent engineering modeling. These findings contribute to advancing data-driven methodologies in construction materials science, offering a pathway for more reliable predictive modeling in material performance assessment.

Table 5.3 Comparative Analysis of Machine Learning Model Performance Based on Optimization Algorithms (Test Set)

Model	R^2	MAE	RMSE	MSE	MAPE	Median AE	Max AE
KNN	0.6192	6.99	10.51	110.41	21.07%	4.80	54.10
SVR	0.5561	7.73	11.34	128.70	21.77%	5.58	51.97
RF	0.7026	5.84	9.29	86.21	16.41%	3.80	48.63
XGBoost	0.7387	4.43	8.70	75.75	11.32%	1.98	46.04
MLP	0.5356	7.34	9.64	92.88	26.75%	5.56	43.55
ET	0.7211	5.45	8.99	80.87	15.49%	3.34	46.34
CatBoost	0.7233	5.34	8.96	80.21	14.49%	3.36	51.04
CVAE-KNN	0.6539	6.81	9.00	81.06	26.42%	5.10	27.70
CVAE-SVR	0.627	7.32	9.49	90.14	34.01%	5.91	25.82
CVAE-RF	0.7866	4.33	6.42	41.17	12.08%	3.42	31.20
CVAE-XGB	0.8572	3.06	5.83	33.94	8.24%	1.59	29.46
CVAE-MLP	0.5274	7.25	9.88	97.60	24.17%	5.50	27.12
CVAE-ET	0.7718	4.73	6.50	42.29	14.46%	3.47	25.16
CVAE-CAT	0.8082	4.33	6.92	47.88	13.33%	2.68	28.67

Table 5.4 Comparative Analysis of Machine Learning Model Performance Based on Optimization Algorithms (Training Set)

Model	R^2	MAE	RMSE	MSE	MAPE	Median AE	Max AE
KNN	0.7301	5.38	7.35	53.98	18.45%	4.03	37.80
SVR	0.5377	6.79	9.61	92.44	24.49%	4.81	49.30
RF	0.8922	3.27	4.64	21.55	10.68%	2.41	28.70
XGBoost	0.9834	0.94	1.82	3.32	2.86%	0.49	14.30
MLP	0.6268	7.12	10.40	108.20	22.12%	4.87	48.59
ET	0.8136	3.83	6.10	37.27	11.85%	2.29	43.40
CatBoost	0.9214	2.91	3.97	15.73	9.29%	2.19	25.60
CVAE-KNN	0.7878	5.08	6.54	42.78	17.10%	4.08	20.60
CVAE-SVR	0.6386	5.91	7.97	63.45	20.65%	4.72	24.69
CVAE-RF	0.9255	2.86	3.83	14.67	9.58%	2.11	13.87
CVAE-XGB	0.9892	0.89	1.47	2.15	2.845	0.43	7.76
CVAE-MLP	0.7199	5.63	7.31	53.46	20.50%	4.48	23.79
CVAE-ET	0.8877	3.23	4.70	22.05	10.80%	2.08	14.87
CVAE-CAT	0.936	2.62	3.51	12.32	8.34%	1.87	16.12

5.5 Model Interpretability

All SHAP results are computed for the same train/test split as Table 5.3 and for the final CVAE-XGBoost model. As illustrated in Figure 5.17, the SHAP summary plot offers an intuitive representation of the distribution and directional impact of each feature on the model's predictions. The analysis revealed that GGBFS, MK, water content, and coarse aggregates were the most influential features governing the prediction results. Notably, GGBFS exhibits the highest absolute SHAP values, indicating its dominant role in determining the model's output. Both Water content and MK also demonstrate substantial contributions to strength prediction, with their varying input values exerting either positive or negative effects on the predicted compressive strength.

Furthermore, secondary features such as Age, Na_2SiO_3, NaOH, and Fly Ash exhibit marginal contributions, though their overall influence remains comparatively limited relative to the primary predictors. This hierarchical distinction in feature importance provides valuable insights for optimizing mix design and material selection in GPC applications. By leveraging SHAP-based interpretability, this thesis enhances the transparency of machine learning-driven predictions and supports mix-design understanding under the investigated sulfate-exposure conditions. The CVAE latent representation is discussed in Section 4.3.2; this section focuses on SHAP-based attribution of the original mix-design variables and age.

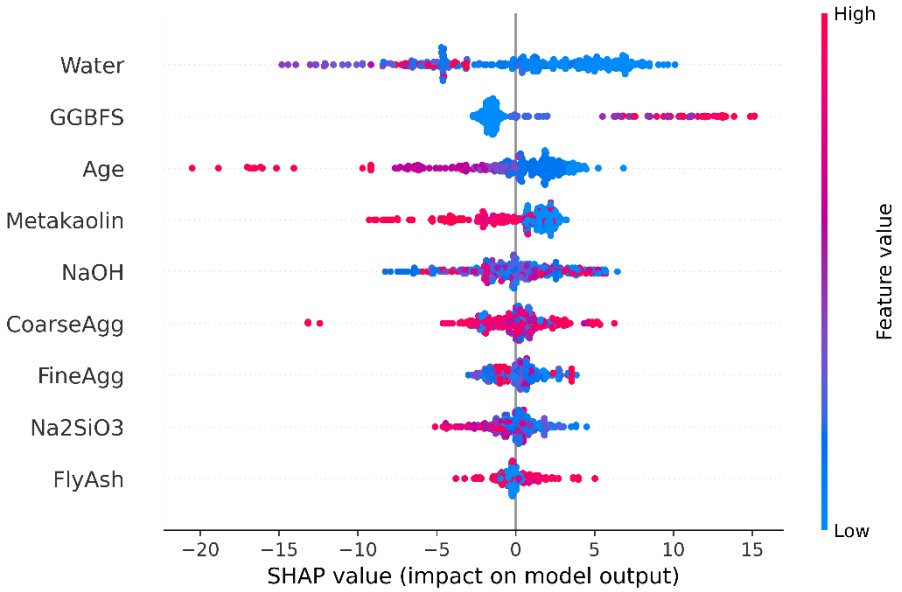


Figure 5.17 SHAP Analysis of the Strength Prediction Model

Further analysis of the feature importance rankings, as illustrated in Figure 5.18, demonstrates that GGBFS exhibits the most substantial contribution at 38%, surpassing all other variables. MK, Water, and Coarse Aggregates follow with contributions of 23.5%, 10.3%, and 7.6%, respectively, while the remaining features each account for less than 10% of the total influence. These findings are consistent with established engineering understanding of geopolymer binder systems, thereby supporting the physical plausibility of the prediction trends under the investigated data scope.

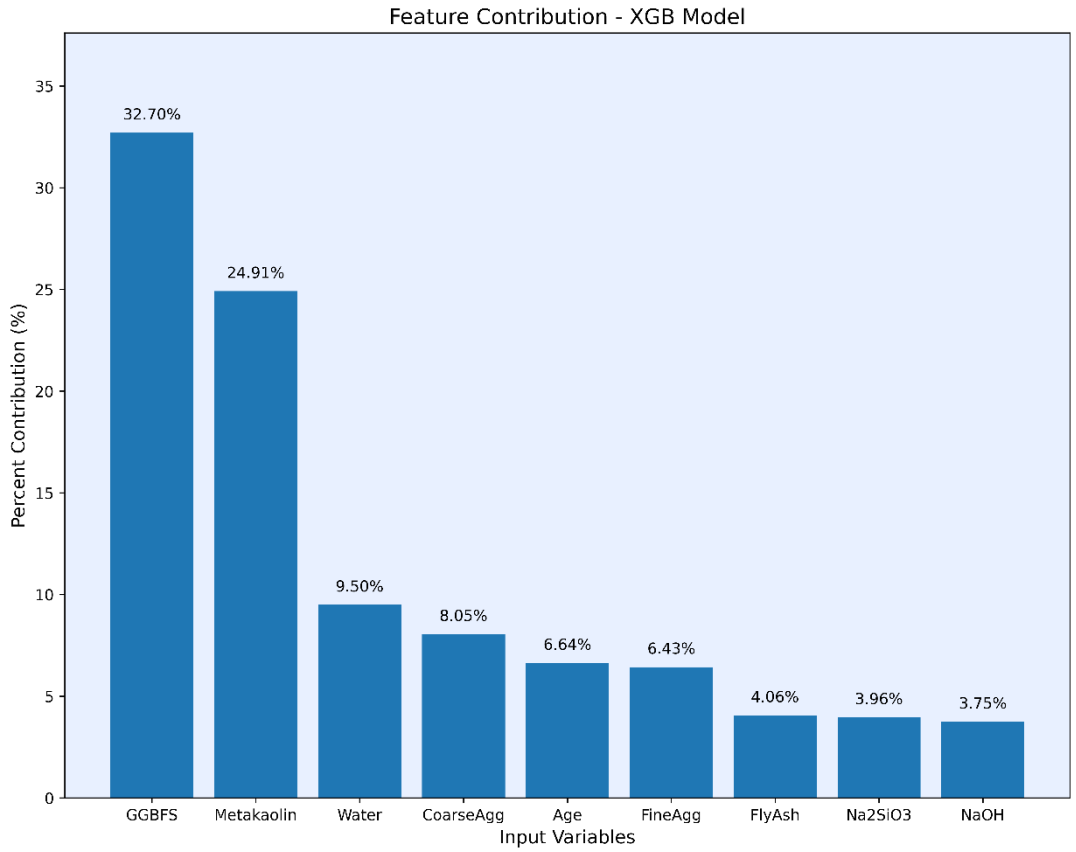


Figure 5.18 Feature Importance Ranking Based on SHAP Values

To further examine the practical interpretability of the proposed model, this thesis investigated the temporal evolution of predicted compressive strength under fixed mix proportions using the SHAP-informed CVAE-XGBoost framework. By holding all major input features constant at their mean values while systematically varying the age parameter, a clear strength degradation trend was obtained, as illustrated in Figure 5.19.

The results demonstrate a nonlinear deterioration pattern, with predicted compressive strength declining from 36.3 MPa at day 1 to 20.3 MPa by day 241. Notably, the degradation rate was more pronounced in the initial phase (1-81 days), followed by a gradual stabilization in the later stage (beyond 161 days). This observed behavior aligns with established material science principles governing time-dependent strength deterioration under sulfate exposure.

The time-dependent strength prediction not only demonstrates the temporal robustness of the CVAE-XGBoost model but also highlights its utility for analysing strength evolution over exposure time under the investigated sulfate conditions. When integrated with SHAP analysis, the predicted degradation trends provide quantitative insights into the influence of key variables, particularly exposure duration, on the predicted strength response. These findings provide a useful basis for understanding time-dependent strength trends within the compiled dataset and may support subsequent performance-oriented modelling when complemented by additional environmental descriptors and validation.

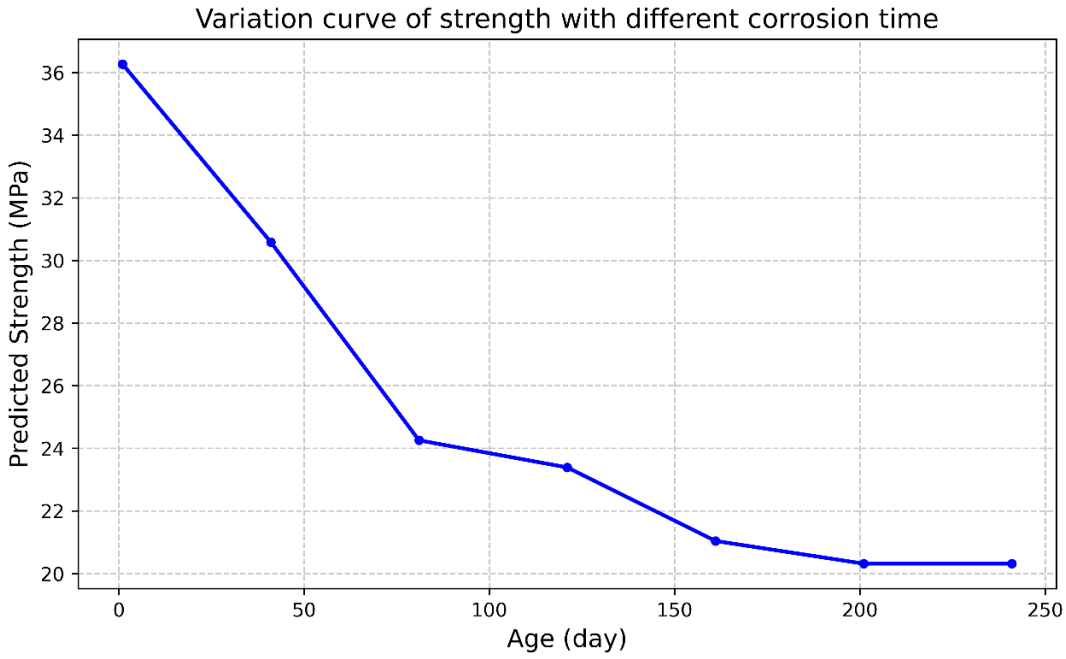


Figure 5.19 Predicted strength degradation over time under fixed mix proportion

5.6 Influence of the Number of CVAE Latent Features on Model Performance

This study systematically examines how the dimensionality of the CVAE latent space affects prediction performance through controlled hyperparameter tuning. To isolate the effect of latent dimensionality, the architecture and parameters of the downstream regressor are held constant while the number of latent features is varied. Performance is evaluated on the test set using the coefficient of determination R^2 and the root mean square error RMSE. The results are shown in Figures 5.20 and 5.21.

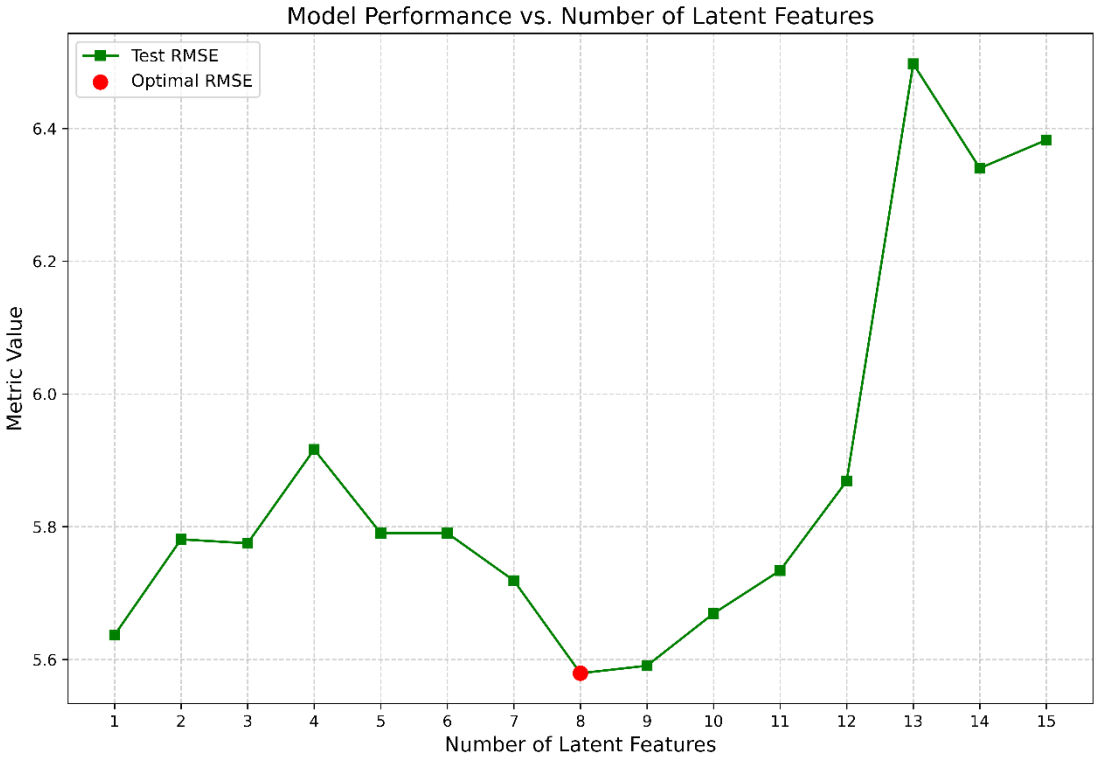


Figure 5.20 Impact of Different Latent Feature Dimensions on Model RMSE

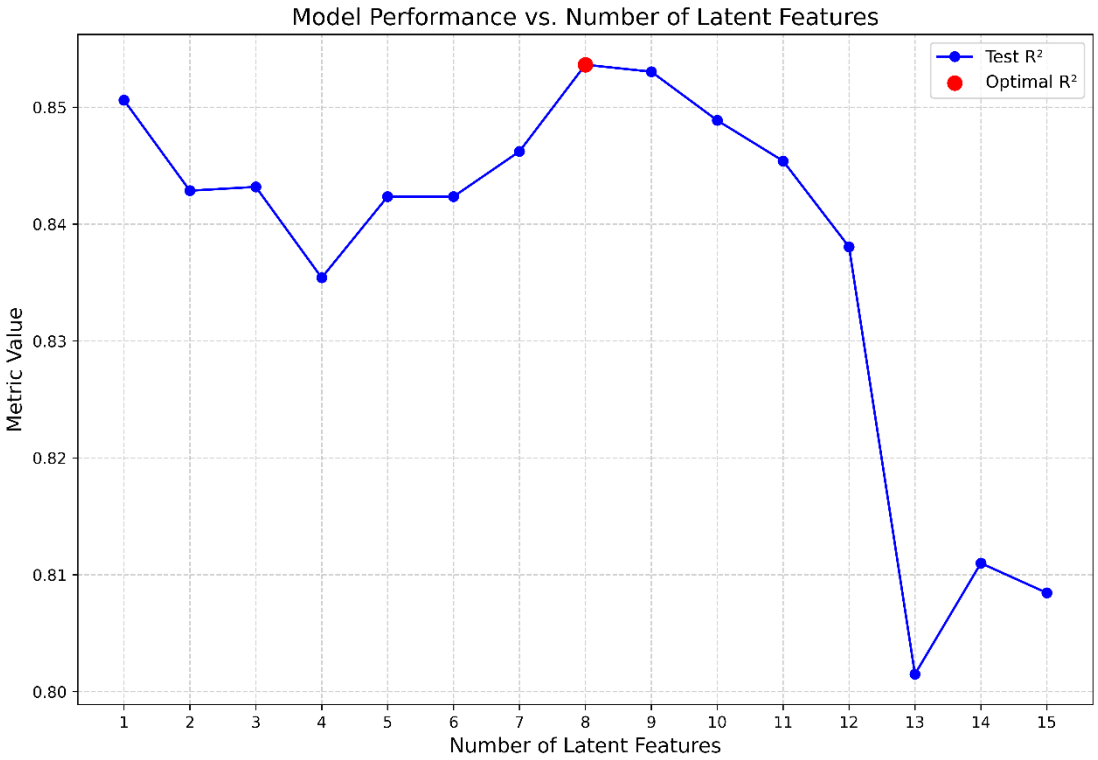


Figure 5.21 Impact of Different Latent Feature Dimensions on Model R^2

The experimental results show that increasing the latent dimensionality initially improves predictive accuracy, with the optimum at eight latent features. At this setting, the model achieves $R^2 = 0.8572$ and RMSE = 5.83 on the test set, indicating strong generalization and aligning with the best CVAE-XGBoost performance reported in Section 5.4. Beyond this threshold, further increases in latent features lead to a decline in test performance: R^2 decreases and RMSE increases, which is a characteristic signature of overfitting. This behavior suggests that a moderate latent dimension is sufficient to encode the essential nonlinear interactions embedded in the mix-design variables and age, whereas overly large latent spaces may capture dataset-specific noise rather than transferable material-response patterns.

These findings indicate that an appropriate latent dimension effectively captures the underlying patterns and complex relationships within the mix design data, thereby improving prediction accuracy. In contrast, excessive latent features introduce noise and redundancy that compromise generalization. Accordingly, eight latent features are adopted as the default CVAE configuration to balance accuracy and robustness while avoiding unnecessary model complexity. This setting offers adequate capacity to encode the key nonlinear interactions present in the dataset, and the latent representation is interpreted qualitatively as a compact descriptor of coupled material-response effects, rather than as a set of directly measurable microstructural or chemical variables. Synthetic samples are generated within the learned conditional latent distribution and constrained to remain within the observed feature ranges, so the augmented data remain anchored to the dataset scope.

5.7 Summary

This chapter systematically investigates machine learning-based prediction of compressive strength for GPC, focusing on model performance enhancement and engineering applicability. Key findings are summarized as follows:

1. Seven regressors were developed and evaluated: KNN, SVR, RF, XGBoost, MLP, ET, and CatBoost. Ensemble methods (XGBoost, RF, CatBoost) outperformed the basic models. Before augmentation they showed limitations on the strongly nonlinear

geopolymer dataset; these were substantially mitigated after CVAE-based enhancement.

2. Integrating CVAE deep feature augmentation with Bayesian hyperparameter optimization (Optuna) produced consistent gains across models. The CVAE-XGBoost model achieved the best test performance with $R^2 = 0.8572$, RMSE = 5.83, and MAE = 3.06, improving over baseline XGBoost by 16.0% in R^2, 32.9% in RMSE, and 30.9% in MAE.

3. SHAP analysis identified GGBFS as the most influential feature (38%), followed by metakaolin (23.5%), water (10.3%), and coarse aggregates (7.6%); age and activator variables contributed secondarily. Under fixed mix proportions in immersion, the model predicts a nonlinear deterioration: strength decreases from 36.3 MPa on day 1 to 20.3 MPa by day 241, with a faster decline during days 1 to 81 and a slower rate after about day 161.

4. A sensitivity study on the CVAE latent space showed that eight latent features provide the best balance between accuracy and robustness; larger latent dimensions reduced generalization due to overfitting.

These results establish a data-driven, interpretable framework for predicting GPC strength deterioration and support durable, sustainable mix design.

Chapter 6. SUMMARY AND CONCLUSIONS

6.1 Conclusions

This thesis presents a systematic investigation into the durability-related performance of GPC under simulated marine sulfate exposure. In addition, a data-driven framework is developed to predict compressive strength degradation under sulfate immersion, where strength retention is treated as one durability-related performance indicator. Through a comprehensive series of experimental studies and machine learning analyses, the research examines deterioration behaviour under the investigated exposure regimes and evaluates the efficacy of advanced ML approaches for time-dependent strength prediction within the scope of the compiled dataset.

6.1.1 Experimental Evaluation of Durability Performance

This thesis presents an experimental investigation into the durability performance of BFRGC under simulated marine erosion conditions. The methodology encompasses specimen preparation, exposure conditions, and testing procedures, including mechanical property evaluation and pore structure characterization. The key findings are summarized as follows:

1. The incorporation of 1% basalt fiber improved the durability-related performance of GPC under the investigated simulated marine sulfate exposure regimes. Compared to plain GPC, BFRGC demonstrates superior strength retention and reduced stiffness degradation under both W-D cycling and continuous immersion conditions.
2. Wet-dry cycling induced more pronounced deterioration than continuous immersion, leading to accelerated reductions in compressive strength and elastic modulus. This is attributed to the exacerbation of microcracking and microstructural damage under cyclic exposure.
3. LF-NMR analysis indicates that GPC incorporating basalt fibers is associated with slower microstructural deterioration under W-D cycles, consistent with improved crack control and delayed crack coalescence within the investigated material system. Specifically, the enlargement of pore size is less significant in BFRGC (1% fiber content) than in fiber-free GPC. Notably, after 96 days of erosion testing, fiber-free specimens develop a distinct third peak in pore size distribution, which is absent in

BFRGC, indicating improved pore-structure stability under the investigated conditions.

4. GPC exhibits good resistance to prolonged sulfate exposure under immersion conditions. Both fiber-reinforced and unreinforced specimens retain structural integrity over the 192-day testing period, with only marginal increases in porosity observed.

6.1.2 Performance of the CVAE-Optimized Strength Prediction Model

A dataset comprising 369 GPC mix designs was compiled from 36 peer-reviewed publications. All data corresponded to specimens exposed to 5% sulfate solutions and tested after standard 28-day curing, ensuring consistency and comparability across sources. On this basis, the ML models were developed to predict compressive strength degradation under sulfate immersion within the scope of the standardized dataset compiled in this thesis, where strength retention is treated as a durability-related performance indicator.

1. The CVAE-XGBoost model demonstrated superior predictive performance, achieving an R^2 of 0.8493, RMSE of 5.66 MPa, and MAE of 4.07 MPa, significantly outperforming baseline models and validating the efficacy of the hybrid approach.

2. SHAP analysis revealed that GGBFS, MK, water content, and coarse aggregate were the most influential factors in compressive strength predictions, consistent with established engineering principles.

3. Sensitivity analysis on the CVAE's latent variables indicated that eight features provided the optimal balance between accuracy and generalization, with additional dimensions leading to overfitting and reduced test-set robustness.

6.2 Recommendation for Future Studies

While this thesis advances the understanding of durability-related behaviour of GPC under sulfate-rich exposure conditions and demonstrates a data-driven framework for strength-degradation prediction, several limitations remain. To enhance applicability and scientific rigor, the following directions are recommended for future work:

1. This thesis exclusively examined sulfate-induced degradation using a 5% sodium sulfate solution to simulate marine exposure. However, chloride-induced corrosion—

particularly from seawater or de-icing salts—also significantly compromises concrete durability in coastal and cold regions. Future work should incorporate chloride attack or combined sulfate-chloride exposure to better replicate real-world marine conditions.

2. The current predictive model is developed based on compressive strength data under continuous immersion conditions and a standardized sulfate-exposure protocol. Future studies may improve generalizability by expanding the dataset and feature set when consistently available, and by extending the framework to account for additional exposure descriptors and coupled actions.

3. All experimental and ML data in this thesis were obtained under controlled laboratory conditions. To validate these findings, long-term field exposure tests or natural aging studies should be conducted, ensuring that laboratory-based conclusions accurately reflect GPC performance in real-world service environments.

4. The study focused on compressive strength degradation as a primary durability-related performance metric. Future research should adopt a multi-parameter approach, incorporating mass loss, sulfate penetration depth, porosity changes, and crack propagation kinetics to establish a more comprehensive durability assessment framework.

5. Building on the current prediction-oriented framework, future work may extend the methodology toward inverse mix design by exploring generative AI models that recommend candidate geopolymer formulations to meet a required compressive strength at specified service-life milestones aligned with design specifications. Such an approach would enable performance-driven mix optimization under prescribed durability and exposure constraints, provided that sufficiently comprehensive datasets and experimental validation are available.

6. Due to the use of fiber-reinforced, steel-free specimens, reinforcement corrosion was not considered. To facilitate GPC adoption in structural applications, subsequent studies should investigate the corrosion behavior of steel reinforcement embedded in geopolymer matrices, particularly under chloride-rich conditions.

REFERENCES

Darmawan, M.S. et al. (2024). "Case study of performance of a jetty structure after 25 years of exposure in a marine environment considering earthquake load." *Engineering Failure Analysis*, 156, 107831.

Poloju, K.K., Srinivasu, K. and Rao, M. (2020). "Study on mechanical characterization of geopolymer cement mortar with single solution and combined solution." *Journal of Xi'an University of Architecture & Technology*, 12, pp. 481-487.

Işıkdağ, B. and Yalghuz, M.R. (2023). "Strength development and durability of metakaolin geopolymer mortars containing pozzolans under different curing conditions." *Minerals*, 13(7), 857.

Chakkor, O. (2025). "Durability and mechanical analysis of basalt fiber reinforced metakaolin-red mud-based geopolymer composites." *Buildings*, 15(12), 2010.

Taffese, W.Z., Hilloulin, B., Zaccardi, Y.V. et al. (2025). "Machine learning in concrete durability: challenges and pathways identified by RILEM TC 315-DCS towards enhanced predictive models." *Materials and Structures*, 58, 145.

Meng, Q., Wu, C., Hao, H., Li, J., Wu, P., Yang, Y. and Wang, Z. (2020). "Steel fiber reinforced alkali-activated geopolymer concrete slabs subjected to natural gas explosion in buried utility tunnel." *Construction and Building Materials*, 246, 118447.

Rahman, S.K. and Al-Ameri, R. (2021). "A newly developed self-compacting geopolymer concrete under ambient condition." *Construction and Building Materials*, 267, 121822.

Zaidi, F.H.A. et al. (2019). "Performance of geopolymer concrete when exposed to marine environment." *IOP Conference Series: Materials Science and Engineering*, 551(1), 012001.

Chen, K., Wu, D., Xia, L., Cai, Q. and Zhang, Z. (2021). "Geopolymer concrete durability subjected to aggressive environments—A review of influence factors and comparison with ordinary Portland cement." *Construction and Building Materials*, 279, 122496.

Aldawsari, S. et al. (2022). "Setting time, microstructure, and durability properties of low calcium fly ash/slag geopolymer: a review." *Materials*, 15(3), 876.

Deb, P., Sarker, P.K. and Nath, P. (2013). “Sulphate resistance of slag blended fly ash based geopolymer concrete.” *Proceedings of the 26th Biennial National Conference of the Concrete Institute Australia: Understanding Concrete*, Gold Coast, QLD, Australia.

Rahman, S.K. and Al-Ameri, R. (2023). “Long-term performance of basalt fiber-reinforced marine geopolymer concrete in harsh environment.” *Magazine of Concrete Research*, 75(22), pp. 1165-1187.

Sutrisno, W. et al. (2017). “Cracking process of reinforced concrete induced by non-uniform reinforcement corrosion.” *Jurnal Teknologi (Sciences & Engineering)*, 79(3).

Ganesan, N., Abraham, R. and Raj, S.D. (2015). “Durability characteristics of steel fiber reinforced geopolymer concrete.” *Construction and Building Materials*, 93, pp. 471-476.

Amorim Jr, N.S., Andrade Neto, J.S. and Santana, H.A. (2021). “Durability and service life analysis of metakaolin-based geopolymer concretes with respect to chloride penetration using chloride migration test and corrosion potential.” *Construction and Building Materials*, 287, 122970.

Mohseni, E., Kazemi, M.J., Koushkbaghi, M., Zehtab, B. and Behforouz, B. (2019). “Evaluation of mechanical and durability properties of fiber-reinforced lightweight geopolymer composites based on rice husk ash and nanoalumina.” *Construction and Building Materials*, 209, pp. 532-540.

Hodhod, O.A. and Salama, G. (2013). “Developing an ANN model to simulate ASTM C1012-95 test considering different cement types and different pozzolanic additives.” *HBRC Journal*, 9(1), pp. 1-14.

Bassuoni, M.T. and Nehdi, M.L. (2009). “Durability of self-consolidating concrete to different exposure regimes of sodium sulfate attack.” *Materials and Structures*, 42, pp. 1039-1057.

Wang, K., Guo, J., Wu, H. and Yang, L. (2020). “Influence of dry-wet ratio on properties and microstructure of concrete under sulfate attack.” *Construction and Building Materials*, 263, 120635.

Meng, Z., Liu, Q.-F., Ukrainczyk, N., Mu, S., Zhang, Y. and De Schutter, G. (2024). “Numerical study on the chemical and electrochemical coupling mechanisms for

concrete under combined chloride-sulfate attack." *Cement and Concrete Research*, 175, 107368.

Song, P.S., Huang, S. and Shen, B.C. (2005). "Strength properties of nylon and polypropylene fiber-reinforced concretes." *Cement and Concrete Research*, 35(8), pp. 1346-1550.

Clarke, J., Peaston, C. and Swannell, N. (2007). *TR 65 Guidance on the Use of Macro-Synthetic-Fiber Reinforced Concrete*. Concrete Society, Blackwater, Camberley.

Bo-wei, Y. et al. (2015). "Study of durability of reactive magnesia-activated ground granulated blast-furnace slag stabilized soil attacked by sulfate sodium solution." *Rock and Soil Mechanics*, 36, pp. 64-72.

Ariffin, M.A.M., Bhutta, M.A.R., Hussin, M.W., Tahir, M.M. and Aziah, N. (2013). "Sulfuric acid resistance of blended ash geopolymer concrete." *Construction and Building Materials*, 43, pp. 80-86.

Wang, L., Gao, M. and Zhang, J. (2022). "Effect of continuous loading coupled with wet-dry cycles on strength deterioration of concrete." *Sustainability*, 14(20), p. 13407.

Horpibulsuk, S., Phojan, W., Suddeepong, A., Chinkulkijniwat, A. and Liu, M.D. (2016). "Durability against wetting-drying cycles of water treatment sludge-fly ash geopolymer and water treatment sludge-cement and silty clay-cement systems." *Journal of Materials in Civil Engineering*, 28(1), p. 04015078.

Guo, X. and Xiong, G. (2021). "Resistance of fiber-reinforced fly ash-steel slag based geopolymer mortar to sulfate attack and drying-wetting cycles." *Construction and Building Materials*, 269, p. 121326.

Nagaraju, T.V., Bahrami, A., Azab, M. and Naskar, S. (2023). "Development of sustainable high performance geopolymer concrete and mortar using agricultural biomass—A strength performance and sustainability analysis." *Frontiers in Materials*, 10, p. 1128095.

Liu, X.X., Wang, S.X., Han, F., Qin, J.Y., Lu, L.M., Xue, Q. and Ji, Y.S. (2024). "Mechanical relationship between compressive strength and sulfate erosion depth of basalt fiber reinforced concrete." *Construction and Building Materials*, 411, p. 134412.

Furtos, G., Prodan, D., Sarosi, C., Moldovan, M., Korniejenko, K., Miller, L., Fiala, L. and Nováková, I. (2024). "Mechanical properties of MiniBars™ basalt fiber-reinforced geopolymer composites." *Materials*, 17(5), p. 1707.

Jiang, Z., Zhu, Z. and Accornero, F. (2024). "Tensile-to-shear crack transition in the compression failure of steel-fiber-reinforced concrete: insights from acoustic emission monitoring." *Buildings*, 14(7), p. 2039.

Şahin, F., Uysal, M., Canpolat, O., Aygörmez, Y., Cosgun, T. and Dehghanpour, H. (2021). "Effect of basalt fiber on metakaolin-based geopolymer mortars containing Rilem, basalt and recycled waste concrete aggregates." *Construction and Building Materials*, 301, p. 124113.

Yavana Rani, S., Saleh Nusari, M., bin Non, J., Poddar, S. and Bhaumik, A. (2022). "Durability of geopolymer concrete with addition of polypropylene fiber." *Materials Today: Proceedings*, 56, pp. 2846-2851.

Kantarcı, F. (2022). "Influence of fiber characteristics on sulfate resistance of ambient‐cured geopolymer concrete." *Structural Concrete*, 23(2), pp. 775-790.

Mirgozar Langaroudi, M.A., Khademi, F., Rahmani, I., Ebrahimi, G. and Afshari, M. (2025). "Investigation of the mechanical properties and durability of fiber-reinforced geopolymer mortars containing metakaolin and glass powder." *Infrastructures*, 10(2), p. 25.

Abrams, D.A. (1918). *Design of Concrete Mixtures*. Bulletin No. 1, Lewis Institute, Chicago, Illinois.

Li, S., Yang, J. and Zhang, P. (2020). "Water‐cement‐density ratio law for the 28‐day compressive strength prediction of cement‐based materials." *Advances in Materials Science and Engineering*, 2020, p. 7302173.

Yeh, I-C. (1998). "Modeling of strength of high-performance concrete using artificial neural networks." *Cement and Concrete Research*, 28(12), pp. 1797-1808.

Rathnayaka, M., Karunasinghe, D., Gunasekara, C., Wijesundara, K., Lokuge, W. and Law, D.W. (2024). "Machine learning approaches to predict compressive strength of fly ash-based geopolymer concrete: a comprehensive review." *Construction and Building Materials*, 419, p. 135519.

Wang, Y., Iqtidar, A., Amin, M.N., Nazar, S., Hassan, A.M. and Ali, M. (2024). "Predictive modelling of compressive strength of fly ash and ground granulated blast

furnace slag based geopolymer concrete using machine learning techniques." *Case Studies in Construction Materials*, 20, p. e03130.

Van Dao, D., Trinh, S.H., Ly, H.B. and Pham, B.T. (2019). "Prediction of compressive strength of geopolymer concrete using entirely steel slag aggregates: novel hybrid artificial intelligence approaches." *Applied Sciences*, 9(6), p. 1113.

Alaneme, G.U., Olonade, K.A. and Esenogho, E. (2023). "Critical review on the application of artificial intelligence techniques in the production of geopolymer-concrete." *SN Applied Sciences*, 5(5), p. 217.

Siva Krishna, A. and Ranga Rao, V. (2019). "Strength prediction of geopolymer concrete using FUZZY." *International Journal of Recent Technology and Engineering*, 7, pp. 661-667.

Jin, L., et al. (2025). "Compressive strength prediction of concrete under sulfate attack using coupled machine learning methods." *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 49(2), pp. 1577-1590.

Hilloulin, B., et al. (2023). "Interpretable ensemble machine learning for the prediction of the expansion of cementitious materials under external sulfate attack." *Journal of Building Engineering*, 80, p. 107951.

Zhou, F.Y., et al. (2023). "Prediction of chloride diffusion coefficient in concrete based on machine learning and virtual sample algorithm." *Sustainability*, 15(24), p. 16896.

Kechroud, F., Benzaamia, A. and Ghrici, M. (2025). "Optimized regression-based machine learning models for predicting chloride diffusion in concrete." *Asian Journal of Civil Engineering*, pp. 1-14.

Ehsani, M., et al. (2024). "Machine learning for predicting concrete carbonation depth: a comparative analysis and a novel feature selection." *Construction and Building Materials*, 417, p. 135331.

Mansouri, S., et al. (2025). "Predicting concrete carbonation depth and investigating the influencing factors through machine learning approaches and optimization." *Amirkabir Journal of Civil Engineering*, 56(12), pp. 1583-1604.

Wong, L.S. (2022). "Durability performance of geopolymer concrete: a review." *Polymers*, 14(5), p. 868.

Ge, X., et al. (2022). “Accelerated design and deployment of low-carbon concrete for data centers.” *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*.

Shmuel, A., Glickman, O. and Lazebnik, T. (2025). “Data augmentation for deep learning regression tasks by machine learning models.” *arXiv preprint arXiv:2501.03654*.

ASTM (2017). *Standard specification for coal fly ash and raw or calcined natural pozzolan for use in concrete*. ASTM C 618, ASTM International, West Conshohocken, PA.

ASTM International. (2022). “C597-22: Standard Test Method for Ultrasonic Pulse Velocity Through Concrete.” ASTM International, West Conshohocken, PA.

GB/T 50081 (2019). Standard for test method of mechanical properties on ordinary concrete. Ministry of Transport of the People’s Republic of China, Beijing.

Scherer, G.W. (2004). “Stress from crystallization of salt.” *Cement and Concrete Research*, 34(9), pp. 1613-1624.

Bassuoni, M.T. and Nehdi, M.L. (2009). “Durability of self-consolidating concrete to sulfate attack under combined cyclic environments and flexural loading.” *Cement and Concrete Research*, 39(3), pp. 206-226.

Omikrine Metalssi, O., Quiertant, M., Jabbour, M. and Baroghel-Bouny, V. (2024). “Effect of exposure conditions on mortar subjected to an external sulfate attack.” *Materials*, 17(13), p. 3198.

Wang, Y., Hu, S. and Sun, X. (2022). “Experimental investigation on the elastic modulus and fracture properties of basalt fiber-reinforced fly-ash geopolymer concrete.” *Construction and Building Materials*, 338, p. 127570.

Flatt, R.J. (2002). “Salt damage in porous materials: how high supersaturations are generated.” *Journal of Crystal Growth*, 242(3-4), pp. 435-454.

Steiger, M. (2005). “Crystal growth in porous materials—I: the crystallization pressure of large crystals.” *Journal of Crystal Growth*, 282(3-4), pp. 455-469.

Guo, J.J., Guo, J., et al. (2019). “Effect of dry-wet ratio on properties of concrete under sulfate attack.” *Applied Sciences*, 9(18), p. 3778.

Liu, F., Zhang, Y., et al. (2021). “Effects of sodium sulfate attack on concrete incorporated with drying-wetting cycles.” *Advances in Materials Science and Engineering*, 2021, p. 5393504.

Yang, H., Zhao, G., et al. (2025). “Study on mass and performance deterioration of concrete under multiple corrosive environments.” *Materials*, 18(9), p. 1931.

ASTM International. (2021). “ASTM C642-21: Standard test method for density, absorption, and voids in hardened concrete.” *ASTM International*, West Conshohocken, PA, p. C642-21.

He, Z., Li, Q., et al. (2024). “Study on the Effect of Basalt Fiber Content and Length on Mechanical Properties and Durability of Coal Gangue Concrete.” *Sustainability*, 16(21), p. 9310.

Hu, Y., Zhao, G., et al. (2024). “The Improving Role of Basalt Fiber on the Sulfate-Chloride Corrosion Resistance of Concrete in the Splash Zone.” *Materials*, 17(18), p. 4454.

Rodríguez-Navarro, C., Doehne, E., & Sebastián, E. (2000). “How does sodium sulfate crystallize? Implications for the decay and testing of building materials.” *Cement and Concrete Research*, 30(10), pp. 1527-1534.

Zhang, Y.H., Zhong, W.L., & Fan, L.F. (2024). “Long-term durability investigation of basalt fiber-reinforced geopolymer concrete in marine environment.” *Journal of Materials Research and Technology*, 31, pp. 593-605.

Wang, K., Guo, J. and Yang, L. (2021). “Effect of dry-wet ratio on sulfate transport-reaction mechanism in concrete.” *Construction and Building Materials*, 302, p. 124418.

Guo, J.J., Liu, P.Q., Wu, C.L. and Wang, K. (2021). “Effect of dry-wet cycle periods on properties of concrete under sulfate attack.” *Applied Sciences*, 11(2), p. 888.

Hong, S., Yoon, S., Kim, J., Lee, C., Kim, S. and Lee, Y. (2020). “Evaluation of condition of concrete structures using ultrasonic pulse velocity method.” *Applied Sciences*, 10(2), p. 706.

Qin, S., Chen, C. and Zhang, M. (2024). “Modeling of concrete deterioration under external sulfate attack and drying-wetting cycles: a review.” *Materials*, 17(13), p. 3334.

Fan, H., Wang, C., Hu, Y. and Zhao, G. (2024). “Durability performance of basalt fiber-reinforced concrete subjected to sulfate-magnesium combined attack.” *Materials*, 17(5), p. 1128.

Brownstein, K. and Tarr, C.E. (1977). “Spin-lattice relaxation in a system governed by diffusion.” *Journal of Magnetic Resonance*, **26**(1), pp. 17-24.

Liu, F., Zhang, T., Luo, T., Zhou, M., Zhang, K. and Ma, W. (2020). “Study on the deterioration of concrete under dry-wet cycle and sulfate attack.” *Materials*, **13**(18), p. 4095.

Zeng, H., Zhang, J., Li, Y., Su, X., Gu, C. and Zhang, K. (2022). “Mechanical properties and microstructure of basalt fiber reinforced concrete under the single-side salt-freezing-drying-wetting cycles.” *International Journal of Concrete Structures and Materials*, **16**(1), p. 44.

Espinosa, A.B., Revilla-Cuesta, V., Skaf, M., Faleschini, F. and Ortega-López, V. (2023). “Utility of ultrasonic pulse velocity for estimating the overall mechanical behavior of recycled aggregate self-compacting concrete.” *Applied Sciences*, **13**(2), p. 874.

ACTE Technologies (2021). *Support Vector Machine (SVM) Algorithm*. Available at: https://www.acte.in/support-vector-machine-algorithm-machine-learning-article

Breiman, L. (2001). “Random forests.” *Machine Learning*, 45(1), pp. 5-32.

Anasbrital98 (2021). *Random Forest*. Available at: https://anasbrital98.github.io/blog/2021/Random-Forest/

Sohn, K., Lee, H. and Yan, X. (2015). “Learning structured output representation using deep conditional generative models.” *Advances in Neural Information Processing Systems*, 28, pp. 3483-3491.

APPENDIX A

Summary of GPC Datasets from Literature

Title	Authors	Year	Publication Source
Mechanical and durability assessment of cement-based and alkali-activated coating mortars in an aggressive marine environment	Lashkari, S.; Yazdipanah, F.; Shahri, M.; Sarker, P.	2021	SN Applied Sciences, 3: 618
Long-term durability investigation of basalt fiber-reinforced geopolymer concrete in marine environment	Zhang, Y.H.; Zhong, W.L.; Fan, L.F.	2024	Journal of Materials Research and Technology, 31: 593-605
Effect of sodium hydroxide concentration on chloride penetration and steel corrosion of fly ash-based geopolymer concrete under marine site	Chindaprasirt, P.; Chalee, W.	2014	Construction and Building Materials, 63: 303-310
偏高岭土基地聚物混凝土力学性能及耐久性能研究 (Mechanical and durability study of metakaolin-based geopolymer concrete)	Ren Yi	2024	MSc Thesis, Xi'an University of Technology
Compressive strength and workability characteristics of low-calcium fly ash-based self-compacting geopolymer concrete	Ahmed, M.F.; Nuruddin, M.F.; Shafiq, N.	2011	International J. of Civil & Environmental Engineering, 3(2)
Durability of fly ash-based geopolymer concrete in the presence of silica fume	Okoye, F.N.; Prakash, S.; Singh, N.B.	2017	Journal of Cleaner Production, 149: 1062-1067
Development and properties of low-calcium fly ash-based geopolymer concrete	Rajan, M.S.; AnuPriya, A.	2024	International Journal of Innovative Research in Advanced Engineering, 11: 465-469.
Fly Ash Based Geopolymer Concrete with	Galvin, B.; Lloyd, N.	2011	Proc. Concrete 2011 Conf., Perth,

Title	Authors	Year	Publication Source
Recycled Concrete Aggregate			Australia (Concrete Institute)
Fly ash-based geopolymer concrete	Rangan, B.V.	2008	Technical Report, Curtin Univ. of Technology (Perth, Australia)
Effect of elevated temperatures on geopolymer paste, mortar and concrete	Kong, D.L.Y.; Sanjayan, J.G.	2010	Cement and Concrete Research, 40(2): 334-339
Influence of aggregate content on the behavior of fly ash based geopolymer concrete	Joseph, B.; Mathew, G.	2012	Scientia Iranica, 19(5): 1188-1194
The effect of microwave incinerated rice husk ash on the compressive and bond strength of fly ash based geopolymer concrete	Kusbiantoro, A.; Nuruddin, M.F.; Shafiq, N.; Qazi, S.A.	2012	Construction and Building Materials, 36: 695-703
Geopolymer concrete with fly ash	Lloyd, N.A.; Rangan, B.V.	2010	Proc. 2nd Int. Conf. on Sustainable Construction Materials & Technologies, Vol.3, pp.1493-1504
Utilisation of waste material in geopolymeric concrete	Nuruddin, M.F.; Qazi, S.; Kusbiantoro, A.; Shafiq, N.	2011	Proc. ICE - Construction Materials, 164(6): 315-324
Use of OPC to improve setting and early strength properties of low calcium fly ash geopolymer concrete cured at room temperature	Nath, P.; Sarker, P.K.	2015	Cement and concrete composites 55: 205-214.
Properties of fly ash geopolymer concrete designed by Taguchi method	Olivia, M.; Nikraz, H.	2012	Materials & Design, 36: 191-198
Geopolymer concrete columns under combined axial load and biaxial bending	Rahman, M.M.; Sarker, P.K.	2011	Proc. Concrete 2011 Conf., Perth, Australia (Concrete Institute)

Title	Authors	Year	Publication Source
Low-Calcium fly ash-based geopolymer concrete: Long-term properties	Wallah, S.E.; Rangan, B.V.	2006	Research Report GC2, Curtin Univ. of Technology (Perth, Australia)
Fracture behaviour of heat cured fly ash based geopolymer concrete	Sarker, P.K.; Haque, R.; Ramgolam, K.V.	2013	Materials & Design, 44: 580-586
Mechanical properties and microstructure analysis of fly ash geopolymeric recycled concrete	Shi, X.S.; Collins, F.G.; Zhao, X.L.; Wang, Q.Y.	2012	Journal of Hazardous Materials, 237-238: 20-29
Engineering properties of inorganic polymer concretes (IPCs)	Sofi, M.; van Deventer, J.S.J.; Mendis, P.A.; Lukey, G.C.	2007	Cement and Concrete Research, 37(2): 251-257
Strength assessment of heat cured geopolymer concrete slender column	Sujatha, T.; Kannapiran, K.; Nagan, S.	2012	Asian Journal of Civil Engineering, 13(5): 635-646
Low-Calcium fly ash-based geopolymer concrete: Reinforced beams and columns	Sumajouw, M.D.J.; Rangan, B.V.	2006	Research Report GC3, Curtin Univ. of Technology (Perth, Australia)
Fly ash-based geopolymer concrete: study of slender reinforced columns	Sumajouw, M.D.J.; Hardjito, D.	2007	Journal of Materials Science, 42(9): 3124-3130.
Parametric studies on compressive strength of geopolymer concrete	Vora, P.R.; Dave, U.V.	2013	Procedia Engineering, 51: 210-219
The mechanical properties of fly ash-based geopolymer concrete with alkaline activators	Ryu, G.S.; Lee, Y.B.; Koh, K.T.; Kim, S.H.	2013	Construction and Building Materials, 47: 409-418
Long-Term Mechanical Properties of Different Fly Ash Geopolymers	Gunasekara, C.; Setunge, S.; Law, D.W.	2017	ACI Structural Journal, 114(3): 743-752
The Durability of Fly Ash Geopolymer and Alkali-Activated Slag Concretes	Wardhono, A.	2024	Ph.D. Thesis, RMIT University (Australia)
Effect of carbonation curing on the physical, mechanical, and microstructural properties	Beltrame, N.A.M.; Dias, R.L.; Witzke, F.B.; Medeiros-Junior, R.A.	2023	Construction and Building Materials, 406: 133403

Title	Authors	Year	Publication Source
of metakaolin-based geopolymer concrete			
Behavior of Metakaolin-Based geopolymer concrete at ambient and elevated temperatures	Albidah, A.M.; Alqarni, A.S.; Abbas, H.; Almusallam, T.; Al-Salloum, Y.	2022	Construction and Building Materials, 317: 125910
Characteristics of metakaolin-based geopolymer concrete for different mix design parameters	Alghannam, M.; Albidah, A.; Abbas, H.; Al-Salloum, Y.	2021	Journal of Materials Research and Technology, 10: 84-98
Influence of Critical Parameters of Mix Proportions on Properties of MK-Based Geopolymer Concrete	Alghannam, M.; Albidah, A.; Abbas, H.; Al-Salloum, Y.	2021	Arabian Journal for Science and Engineering, 46(5): 4399-4408
Physical & mechanical properties of fiber reinforced metakaolin-based geopolymer concrete	Moradikhou, A.B.; Esparham, A.	2020	Construction and Building Materials, 251: 118965.
Influence of reclaimed asphalt pavement aggregate on the performance of metakaolin-based geopolymer concrete at ambient and elevated temperatures	Albidah, A.S.	2023	Construction and Building Materials, 402: 132945.
Development of metakaolin based high strength recycled aggregate geopolymer concrete	Gopalakrishna, B.; Pasla, D.	2023	Construction and Building Materials, 391: 131810.
Marine Geopolymer Concrete—A Hybrid Curable Self-Compacting Sustainable Concrete for Marine Applications	Rahman, S.K.; Al-Ameri, R.	2022	Applied Sciences, 12(6): 3116.