AI-Driven Utilisation of Alkali-Activated Aluminosilicate Wastes for Sustainable Soil Stabilisation

HAMED FAIZI

Doctor of Philosophy

Supervisor: Professor Abbas El-Zein
Associate Supervisor: Professor David Airey

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

School of Civil Engineering
Faculty of Engineering
The University of Sydney
Australia

9 December 2025

Statement of Originality

I declare that this thesis is an original work and has not been submitted, in whole or in part, for the award of any other qualification. All intellectual content presented in this thesis is the outcome of my independent research, except where explicit reference has been made to the contributions of others. All sources of information and forms of assistance have been acknowledged.

Hamed Faizi
PhD Candidate
September 2025

Author Attribution Statement

This thesis presents research findings that have been, or are in the process of being, published as detailed below:

Results of predictive modelling presented in Chapter 5 has been submitted to Neural Computing and Applications under the title *'Tree-based ensembles and genetic programming for predicting strength of aluminosilicate waste-stabilised soils using an expanded dataset'* (Faizi and El-Zein, 2025d), which is under its second-round reviews at the time of writing this. Additional data that were not used in this thesis or the above paper, were presented at the 6th International Conference on Environmental Geotechnology, Recycled Waste Materials, and Sustainable Engineering (EGRWSE-2025), Vigo, Spain, 11–14 June 2025, under the title *'Predicting Dry Density of Subgrades Stabilised with Agro-Industrial Byproducts: Comparison of Tree-based Ensembles and Genetic Programming'* (Faizi and El-Zein, 2025b). These papers were authored by the candidate, Hamed Faizi, and co-authored by the primary thesis supervisor, Professor Abbas El-Zein. The candidate was responsible for literature review, data collection, data preprocessing, model development, and performance evaluation. Professor El-Zein contributed to conceptual design of the study, and reviewed, refined, and provided feedback on the manuscript.

Experimental findings presented in Chapter 6 have been published by Transportation Geotechnics under the title *'Stabilising Expansive Subgrades with Agro-industrial Waste: An Experimental Investigation'* (Faizi et al., 2025b). Parts of the experimental results were presented at the 5th European Conference on Unsaturated Soils as *'Evaluating swelling behaviour and durability of geopolymerised expansive soils under cyclic moisture conditions'* (Faizi and El-Zein, 2025a), and other findings were accepted for presentation at the 2025 Australian Geomechanics Society (Sydney) Symposium under the title *'Balancing swelling control and strength in clayey subgrades using waste-based geopolymers'* (Faizi et al., 2025a), scheduled for publication in November 2025. These experimental papers were authored by the candidate and co-authored by Professors Abbas El-Zein and David Airey. The candidate was responsible for the design and execution of the experimental program, data analysis and interpretation, and preparation and refinement of the manuscripts. Professor El-Zein contributed to detailed review of the laboratory program, experimental results, and manuscript.

Professor Airey reviewed the manuscript and provided feedback to strengthen the experimental findings.

Key findings reported in Chapter 7 were presented at the 3rd International Conference on Geomechanics and Geoenvironmental Engineering (iCGMGE-2025) under the title *'Predicting soil swelling stabilisation using agro-industrial by-products: Comparing Deep Learning with Genetic Programming'* (Faizi and El-Zein, 2025c). This paper was authored by Hamed Faizi and co-authored by Professor Abbas El-Zein. The candidate developed the methodology, and performed data preparation and preprocessing, model development and evaluation, and preparing and finalising the manuscript. Both authors jointly contributed to the conceptual design of the study, interpretation of findings, and manuscript refinement.

The knowledge and expertise in data modelling (relational and non-relational databases, and schema design) and artifical intelligence (image classification, regression, and natural language processing) gained through this doctoral research has enabled the candidate, to contribute to industrial practice, beyond the scope of this thesis, by proposing ways of using AI tools to enhance data management and knowledge exchange in geotechnical firms. The candidate adapted and customised an AI systems-level architecture called Retrieval Augmented Generation (RAG) to automate the data retrieval process to enable generation of customised and verifiable responses to geotechnical enquiries. The abstract, authored by the candidate and reviewed by Professor David Airey, was submitted to be presented at the seminar Practical Applications of Generative AI in Geotechnical Engineering, organised by the Australian Geomechanics Society (Sydney Chapter) on 10 September 2025. Following acceptance of the abstract, the candidate presented the work titled *Unlocking the Value of Archived Geotechnical Data through the Development and Deployment of a Retrieval-Augmented Generation*, showcasing how underutilised geotechnical archives can be leveraged to improve problem-solving, decision-making, and workflow efficiency in engineering practice (AGS, 2025).

Hamed Faizi
PhD Candidate
September 2025

As the primary research supervisor of the candidate, Hamed Faizi, I confirm that the authorship attribution statements provided above are accurate.

Professor Abbas El-Zein
Primary Supervisor
September 2025

Use of Generative AI

Generative AI (GenAI) has been used to assist in debugging Python scripts, automating the conversion of tables and equations from Microsoft Word to LaTeX format. GenAI was also utilised as a proofreading aid to identify and correct grammatical and typographical errors in selected sections of this thesis. All outputs were critically reviewed and edited by the author to ensure accuracy and originality.

Acknowledgements

I am deeply grateful to my primary research supervisor, Professor Abbas El-Zein, for his continuous support, guidance, and constructive feedback throughout this research. His ability to creatively challenge my perspectives has been crucial in shaping both the quality of this research and my growth as an independent researcher.

I am also thankful to my secondary supervisor, Professor David Airey, for generously sharing his deep knowledge in soil mechanics and for his thoughtful feedback on the adequacy and feasibility of my experimental program.

I would like to acknowledge the former Senior Technical Officer at the Soil Mechanics Laboratory, Mr Ross Barker, for his assistance and support throughout the experimental phase of this research. I am equally grateful to my colleagues and friends in HDR Office 101, as well as members of the SciGEM group and the geo-environmental team at the University of Sydney, for the helpful conversations, ideas, and support that greatly enriched this work.

I would also like to thank the reviewers of my papers for their constructive feedback and valuable suggestions, which helped me improve the quality and clarity of my work.

Finally, I would like to express my deepest gratitude to my parents, my brothers, and my extended family for their continuous support and encouragement.

I am especially indebted to my father for his constant prayers, unwavering dedication to my academic growth and personal development, and for the encouragement that sustained me throughout this research and will continue to inspire me beyond it. I am equally grateful to my mother, whose kindness, resilience, and prayers have been a source of strength and comfort throughout my life.

I am also deeply grateful to my wife, whose patience, dedication, and constant support made it possible for me to devote myself fully to this research.

Australian Government Support

This research was undertaken with the support of a postgraduate research scholarship, the Australian Government Research Training Program, provided by the Australian Government.

Abstract

Background and Objectives. The problematic behaviour of low-strength and expansive soils poses major challenges to lightweight structures such as roads, railways, bridges, and retaining walls. Reusing industrial by-products with high aluminosilicate content, such as fly ash and slag, is an active area of research in soil stabilisation, driven by three key benefits: improving soil properties, minimising waste, and reducing dependence on CO_2-intensive stabilisers such as cement. The conventional approach to chemical soil stabilisation relies heavily on experimentally determined mix designs, which are labour-intensive, time-consuming, and costly. Also, the variability in soil properties, mixture compositions, and curing conditions often limits the generalisability of findings. To address these inefficiencies, there is a need for versatile and reliable predictive models capable of accurately estimating key soil properties, such as unconfined compressive strength (UCS) and swelling behaviour. Such models can reduce reliance on extensive laboratory testing by narrowing experimental work to a targeted set of validation tests. Empirical research has primarily focused on UCS, providing a large body of data to support the development of predictive models. However, while the experimental literature is rich, existing UCS models have not made full use of this literature and are currently based on a limited set of tests which rely on cement and lime, while considering a relatively narrow range of soil types. On the other hand, stabilisation of expansive soils by aluminosilicates remains experimentally under-researched, with knowledge gaps found on stabilisation mechanisms, key factors and interactions, durability and environmental impact under cyclic wetting and drying; in other word, both experimental data and numerical models are lacking in this respect. All three gaps in the field of stabilisation by aluminosilicate waste are addressed in this thesis resulting in the development of wide-scope UCS prediction models, experimental insights and data on swelling stabilisation, and new models predicting swelling of stabilised soils.

Methods. The thesis methodology is comprised of three components corresponding to the three gaps identified above. In the first component (UCS models), two new models predicting UCS of stabilised soils are developed and tested. The models are based exclusively on waste-based aluminosilicates data and developed using an expanded dataset comprising 1,310 unique test records collected from 23 experimental studies, which is the largest dataset of its kind to date. The compiled data encompasses broad range of soil types, curing conditions, mixture compositions, and sample preparation methods, enabling the models to capture complex relationships and interactions among predictors. Genetic programming (GP) and tree-based ensembles (TbE) are used as modelling methods. In the second component (swelling tests), an experimental program of one-dimensional free swelling tests is performed under both monotonic and cyclic wet-dry conditions, to assess the effects of different stabilisation variables on the swelling of stabilised soils. Additional experiments, including X-ray fluorescence, X-ray diffraction, mercury intrusion porosimetry, scanning electron microscopy, unconfined compressive strength testing, and pH measurements, are also conducted to develop a deeper understanding of the underlying stabilisation mechanism, durability, and environmental impacts. Finally, in the third component, 260 data rows generated from the second component are used as the basis for developing and testing two models to predict swelling of stabilised soil, using Extra Trees (ET) and GP methods.

Results. The new UCS models achieved significant improvements in accuracy relative to the previous models. While the ET regressor, a TbE method, achieved slightly higher accuracy than GP, the latter showed comparable performance with better transparency. UCS was found to be strongly correlated to total CaO concentration, while combined silicon, aluminium, and iron oxides had a non-linear effect on UCS, peaking at 25% content. The total water content of the mixture had an exponential decay relationship with UCS, hence highlighting the importance of moisture control. Alkali-liquid content emerged as the most influential predictor in both models, showing a decaying quasi-linear relationship with swelling, while ash content exhibited a positive correlation, suggesting an increased swelling tendency under low alkali-liquid dosages. A data gap in long-term curing was also identified, indicating potential biases in existing models. Laboratory investigations showed that swelling stabilisation requires lower activator concentrations and shorter curing periods than strength development, reducing swelling from 56% to under 3% within 24 hours using 1 M NaOH. Activator content was the dominant experimental factor, helping to achieve high stabilisation even in soils with greater than 50% initial swelling. Microstructural analysis showed denser matrices and reduced number of pores, while pH levels and leachability results highlighted the need for additional design measures to mitigate the impact of high pH. The ET model once again showed marginally better performance compared to the GP model, however, the difference between their performances were not statistically significant.

Conclusions. This thesis has made a significant contribution of soil stabilisation by aluminosilicate waste, especially in developing more advanced models of UCS prediction and showing that aluminosilicate waste is highly promising as a cost-effective stabilisation approach for shallow soils. Further research can explore more energy efficient activators and investigate the effect of varying dry density on swelling. Also, to move research from proof-of-concept to practical applications, site testing of stabilisation is required. Given that predictive models are based on laboratory-cured samples, site testing will allow the development of calibration methods to align model predictions with in-situ curing, because the latter may yield different stabilisation outcomes.

Contents

Nomenclature

$clay\%$	%	Clay fraction of the host soil
$silt\%$	%	Silt fraction of the host soil
$sand\%$	%	Sand fraction of the host soil
$organic\%$	%	Organic content of the soil
LL	%	Liquid limit
PL	%	Plastic limit
PI	–	Plasticity index
LS	%	Linear shrinkage
MC_0	%	Initial moisture content of the host soil
DUW_soil	kN/m^3	Dry unit weight of host soil
OMC_0	%	Optimum moisture content of host soil
MDD_0	g/cm^3	Maximum dry density of host soil
UCS_0	kPa	Unconfined compressive strength of host soil
FSI_0	%	Free swell index of host soil
PL_0	%	Plastic limit of host soil
LL_0	%	Liquid limit of host soil
PI_0	%	Plasticity index of host soil
$ash\%$	%$^{(1)}$	Fly ash content in the mixture
$ash\_C\%$	%$^{(1)}$	Class C fly ash content in the mixture
$ash\_F\%$	%$^{(1)}$	Class F fly ash content in the mixture
$slag\%$	%$^{(1)}$	Slag content in the mixture
$lime\%$	%$^{(1)}$	Lime content in the mixture
$cement\%$	%$^{(1)}$	Cement content in the mixture
$water\%$	%$^{(1)}$	Water content in the mixture
$NaOH\%$	%	Sodium hydroxide content in the mixture
$SS\%$	%$^{(1)}$	Sodium silicate content (Na_2SiO_3)
M	M	Molarity of sodium hydroxide
$soil\%$	%	Host soil percentage in the mixture
DUW_ash	kN/m^3	Dry unit weight of ash
DUW_lime	kN/m^3	Dry unit weight of lime
DUW_mix	kN/m^3	Dry unit weight of mixture
$ALC\%$	%	Alkaline liquid percentage

SS/SH	–	Ratio of sodium silicate to sodium hydroxide
A/ash	–	Ratio of activator to ash
A/B	–	Ratio of activator to binder
$ash/soil$	–	Ratio of fly ash to untreated soil
$CaO\_ash\%$	%	Calcium oxide content in fly ash
$CaO/SiO2\_ash$	–	CaO to SiO_2 ratio in ash
$LOI\_ash\%$	%	Loss on ignition of ash
Na/Al_mix	–	Sodium-to-aluminium ratio in mix
Si/Al_mix	–	Silicon-to-aluminium ratio in mix
$total\_CaO\%$	%	Combined CaO content in mix
$total\_oxides\%$	%	Combined Al, Si, Fe oxides in mix
t_curing	days	Curing time
T_curing	°C	Curing temperature
$sample\_type$	–	Sample preparation method
$initDD$	g/cm^3	Initial dry density before stabilisation
XRD	–	X-ray Diffraction
XRF	–	X-ray Fluorescence
SEM	–	Scanning Electron Microscopy
MIP	–	Mercury Intrusion Porosimetry
LDI	mL/g	Log differential intrusion
CI	mL/g	Cumulative intrusion
1DFS	%	One-dimensional free swelling
PV	mL	Pore volume
WAC	%	Water absorption capacity
H_i	mm	Initial sample thickness
H_f	mm	Final (maximum) sample thickness
ML	–	Machine Learning
ANN	–	Artificial Neural Network
GP	–	Genetic Programming
GEP	–	Gene Expression Programming
RM	–	Regression Model
ET	–	Extra Trees
RF	–	Random Forest
AB	–	AdaBoost
GB	–	Gradient Boosting
BR	–	Bagging Regressor
TbE	–	Tree-based Ensemble
ST	–	Sub-tree in GP expression tree

R^2	–	Coefficient of determination
R^2_{train}	–	R^2 on training set
R^2_{test}	–	R^2 on testing set
r_s	–	Pearson correlation coefficient
MSE	–	Mean squared error
RMSE	–	Root mean squared error
MAE	–	Mean absolute error
VIF	–	Variance inflation factor
SD	–	Standard deviation
CV	–	Cross-validation
OOB	–	Out-of-bag error
SHAP	–	SHapley Additive exPlanations
ReLU	–	Rectified Linear Unit

Notes:

(1) *Percentage by dry soil mass.*

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Motivation

Expansive and low-strength soils present persistent challenges in geotechnical engineering, contributing to ground movement, infrastructure deterioration, and costly remediation. Sustainable stabilisation using recycled aluminosilicate waste such as fly ash and slag has gained momentum. However, the performance of such systems depends on complex interactions between soil mineralogy, byproduct chemistry, alkaline activator dosage, curing conditions, and environmental exposure. The existing literature contains numerous UCS models but lacks a rigorous evaluation of their consistency, mechanisms, and applicability to waste-based stabilisation. Published data on swelling stabilisation, particularly under low-molarity alkali activation, remain scarce, inconsistent, and insufficient for reliable design. Addressing these gaps is essential for moving from experimental trial-and-error toward more efficient data driven stabilisation design.

Machine-learning approaches provide an opportunity to capture nonlinear behaviour and reveal key mix-design relationships, but the lack of structured datasets, inconsistent methodologies, and limited interpretability have prevented their practical adoption in geotechnical engineering. This thesis is motivated by the need to compile a comprehensive UCS database from published studies, generate new high-quality swelling data through laboratory testing, and establish a unified, reproducible machine-learning framework capable of predicting stabilisation performance in a physically meaningful and interpretable manner. By integrating model evaluation, experimental investigation, and comparative machine learning analysis, the research aims to deliver predictive tools and engineering insights that support more sustainable and efficient stabilisation design using industrial aluminosilicate waste.

1.2 Background

Problematic soils can be broadly classified into low-strength and expansive soils. Low-strength soils have weak inter-particle bonds, which result in poor load-bearing capacity. These soils are prone to failure under load. On the other hand, expansive soils are characterised by their high potential for volumetric change due to moisture variations. These soils can exhibit significant swelling or shrinkage, which can lead to structural failure if not properly stabilised. In expansive clayey soils, swelling and strength are closely interrelated as swelling increases the void ratio, thereby weakening the soil structure and reducing its strength.

As shown in Figure 1.1, expansive soils are typically encountered in arid and semiarid regions around the world with approximately 20% of Australia covered by such soils (Vorwerk et al., 2015). These soils can pose significant challenges to the built environment, particularly lightweight structures such as roads and railways, making them an ongoing challenge for geotechnical researchers and engineers (Abdila et al., 2022). Bypassing project areas that are affected by problematic soils is becoming more difficult due to land scarcity (Vijayan and Parthiban, 2020; Tiwari and Satyam, 2021), particularly in major coastal regions where problematic soils are encountered more frequently.

Repairing damages caused by problematic soils is costly. For example, it has been estimated that the US alone spends more than 10 billion dollars annually on repairing infrastructure effected by problematic soils behaviour (Ahmad et al., 2024). To mitigate this risk, problematic soils are either removed and replaced with suitable engineered fill material or chemically stabilised to improve their engineering properties (Ayub and Khan, 2023).

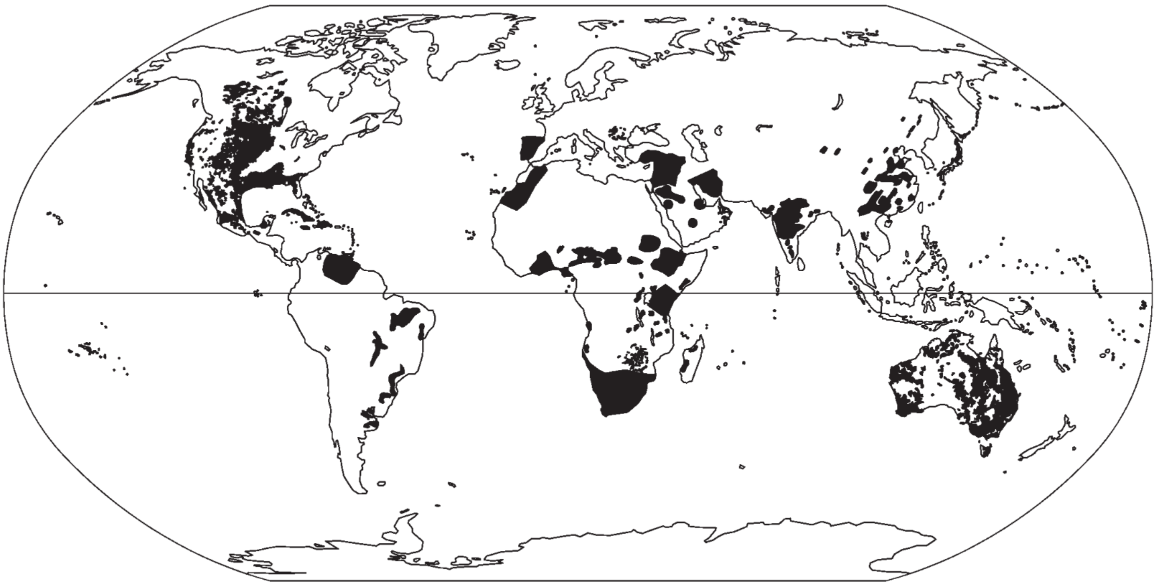


Figure 1.1. Reported locations of expansive soils (Reproduced from Nelson et al. (2015))

Cement and lime have traditionally been the preferred materials for chemical soil stabilisation, with cement favoured due to its widespread availability and ease of use (Ahmad et al., 2024; Lakkimsetti and Nayak, 2022; Parthiban et al., 2022). However, nearly one tonne of carbon dioxide (CO_2) is emitted for every tonne of cement produced (Adedokun and Oluremi, 2019). Global cement consumption currently exceeds 10 billion tonnes annually and is projected to double by 2050 (Liu et al., 2022; Odeh and Al-Rkaby, 2022). In response to these environmental challenges, there is growing interest in developing sustainable alternatives to the current cement-dominated, 'one-size-fits-all' approach to soil stabilisation (Abdullah et al., 2020; Adedokun and Oluremi, 2019; Mypati and Saride, 2022; Parthiban et al., 2022; Provis and Bernal, 2014).

A life cycle assessment has shown that alkali-activated aluminosilicate systems, particularly those combining fly ash and ground granulated blast furnace slag (slag), offer significant environmental advantages over conventional cement (Habert et al., 2011). While fly ash and slag have been the most investigated wastes, there are also others, rich in cementitious oxides, such as paper mill ash, sawdust ash, coconut husk ash, millet husk ash, corn cob ash, rice husk ash, bagasse ash, and locust bean pod ash, that have shown effectiveness in various soil stabilisation applications due to their inherent cementitious properties (Adedokun and Oluremi, 2019; Blesson and Rao, 2023; de Oliveira et al., 2022; Hebib and Farrell, 2003; Khadka et al., 2020; Mypati and Saride, 2022; Nordmark et al., 2022; Parthiban et al., 2022; Yang et al., 2014). To activate their cementitious properties, solid aluminosilicates must be mixed in an alkaline solution; however, the production of alkaline activators such as sodium hydroxide (NaOH) is energy-intensive and associated with a high carbon footprint. Its procurement cost and corrosive nature can also pose practical and environmental challenges in geotechnical applications, if used in high concentrations. Alternative, activators with low carbon footprints, such as glass waste, silica fume, and calcium and aluminium-rich residues, are gaining some attention (Parthiban et al., 2022), but NaOH remains the primary choice. Therefore, seeking ways to reduce its concentration is essential for improving the economic and environmental viability of aluminosilicate wastes, while still achieving acceptable improvements in swelling behaviour and strength. While NaOH concentrations of above 10M are commonly used for strength enhancement (Odeh and Al-Rkaby, 2022; Cristelo et al., 2011), requirements for effective swelling stabilisation remain under-researched.

Repurposing such by-products in soil stabilisation is intended to complement, rather than replace, the use of cement. While cement remains the most appropriate choice for certain projects, particularly those requiring rapid strength gain or subject to strict time constraints, the development of such binders aims to reduce overall reliance on cement while reducing waste. By promoting the use of sustainable materials where feasible, cement can be reserved for applications where its specific properties are critical.

The physicochemical processes involved in soil stabilisation through the activation of aluminosilicates are governed by complex interactions between key variables related to mix design, curing conditions, and the properties of the host soil. From a practitioner's perspective, a critical stage in the planning and design process is the determination of a mix design capable of meeting specific engineering objectives, such as increasing soil strength and dry density, or reducing swelling potential. However, determining resource-efficient and performance-optimised mix designs remains a significant challenge (Consoli et al., 2016). The conventional approach involves experimentally evaluating the effects of individual variables, e.g., binder content, curing time, curing temperature, on a target soil property for the specific site to be stabilised. While this trial-and-error approach can be effective for site-specific applications, it is time-consuming and costly due to the extensive sampling and laboratory testing required. Also, its strictly empirical nature limits the generalisability of its findings across different soils, as soil characteristics vary widely across space and time and no unifying theory currently exists to predict behaviour across soil types. This limitation has hindered the broader adoption of aluminosilicate wastes in soil stabilisation practice.

To overcome the above obstacle, further research is needed to develop reliable and versatile predictive models trained on rich experimental datasets. Experimental studies have mainly focused on the confined compressive strength (UCS) with key variables adequately investigated and sufficient empirical data available to support predictive modelling efforts.

In contrast, less attention has been given to swelling stabilisation, despite early evidence indicating effectiveness in reducing linear shrinkage, liquid limit, and plasticity index (Mypati and Saride, 2022; Mandal and Singh, 2007; Phanikumar, 2022), which suggest reduced volumetric reactivity due to changes in moisture content. However, further experimental investigations are required to address key knowledge gaps related to swelling stabilisation, which will pave the way for the subsequent development of predictive models. Further investigation is required to identify key factors that govern the effectiveness of aluminosilicate-based stabilisation and their interactions. Also, while alkaline activators such as NaOH is required for the activation of aluminosilicates, its use in high concentrations may act as a barrier to the use of aluminosilicate by-products due to its relatively higher cost and potentially negative environmental impacts as highlighted by Habert et al. (2011). Therefore, further investigation is needed to determine the optimum molarity of NaOH for swelling stabilisation. Also, the effectiveness of stabilisation across a range of expansive soils must be explored, and the durability of the stabilisation treatment under cyclic wetting and drying conditions needs thorough assessment. Once these gaps are addressed, experimental programs can be expanded to generate the necessary training data for predictive modelling.

Most of the published research have focused on modelling UCS due to availability of data. However, progress has been limited due to inadequate utilisation of existing data leading to the narrow experimental scopes, which are either constrained by factors such as limited datasets often considering one or very few soil types (Leong et al., 2018; Mozumder and Laskar, 2015; Zeini et al., 2023) or are reliant on other stabilisers such as cement and lime (Pham et al., 2021, 2022).

Conventional regression methods (RM) and machine learning (ML) algorithms have been explored in the development of predictive models in soil stabilisation. Key ML algorithms used in this domain include evolutionary methods, artificial neural networks (ANNs), and random forests, a type of tree-based ensemble method (TbE) (Sari Ahmed et al., 2020; Zeini et al., 2023; Leong et al., 2018). Overall, ML algorithms have demonstrated better performance in modelling soil stabilisation compared to RM (Mozumder and Laskar, 2015). While ANNs have consistently performed well, evolutionary models offer a unique advantage. Despite showing slightly lower performance than ANNs in certain scenarios (Leong et al., 2018; Hanandeh et al., 2020), evolutionary methods, particularly genetic programming (GP) stand out due to their inherent transparency. The ability to express GP models in explicit mathematical expressions quantifies explicitly the relationships between predictors and response, which is a significant advantage in chemical soil stabilisation that requires interpretability.

However, further comparative analyses are needed as transparency alone is not enough to justify the adoption of evolutionary methods in modelling soil stabilisation problems. Existing comparisons between evolutionary methods and ANNs are often based on small datasets with fewer than 130 test records, typically covering only one or very few soil types. These studies primarily focus on performance metrics without evaluating the overall predictor-response behaviours or how well these models align with empirical findings. This gap limits the generalisability of the results, particularly when considering real-world applications with diverse soil conditions, mix designs, and curing conditions.

In addition to ANN and GP algorithms, TbE also require further investigation. These methods have proven to be highly effective in modelling complex regression datasets that exhibit irregular and non-uniform distributions (Zhuang et al., 2020), which are common features of engineering datasets. Despite their known capability, TbEs have only been explored once in soil stabilisation modelling, using a small dataset of just 283 test records (Zeini et al., 2023).

To address the above-mentioned gaps, further research is needed to provide a comprehensive performance comparison between GP and: i) TbE, and ii) ANN. This comparison should be conducted using a richer dataset that encompasses various soil types, mixture designs, and curing conditions, to evaluate not only predictive performance but also the interpretability of models and their consistency with empirical findings.

1.3 Objectives

The overall goal of this thesis is to improve the effectiveness, efficiency, and sustainability of the use of recycles alkali-activated aluminosilicates to address problems of low strength and high swelling in problematic soils. This goal is achieved through three core research objectives:

i. To generate and analyse a high-quality experimental dataset that characterises the swelling behaviour, durability, and environmental impact of soils stabilised with aluminosilicate waste and low-molarity alkaline activators, thereby improving understanding of stabilisation for swelling and providing data for modelling swelling stabilisation.
ii. To establish and verify a unified and reproducible machine learning framework for both strength and swelling prediction, incorporating consistent data preparation, predictor selection, model training, evaluation, and interpretability analyses.
iii. To develop and evaluate benchmarkable machine learning models for predicting unconfined compressive strength and swelling potential, capable of systematic predictor ranking and interaction analysis to identify the most influential mix-design parameters.

1.4 Research Questions

Below are a set of research questions addressed in this thesis.

i. How can insights from the capabilities and limitations of existing strength prediction models be used improve the design of reliable waste-based soil stabilisation systems?
ii. What are the key physicochemical factors governing the swelling behaviour of alkali-activated aluminosilicate-stabilised soils, and how can new laboratory data address gaps in current experimental knowledge?
iii. How can a unified and reproducible machine-learning framework be developed to predict the strength and swelling behaviour of soils stabilised with recycled aluminosilicates?
iv. How reliably can machine-learning models, trained on both literature-derived strength data and newly generated swelling data, predict stabilisation performance and identify the most influential mix-design parameters?
v. How do most adopted methods such as Artificial Neural Networks and Genetic Programming compare with Tree-based Ensembles in their ability to model soil stabilisation using alkali-activated aluminosilicates?

1.5 Thesis Layout

This thesis is organised into eight chapters, structured around two core themes: strength gain and swelling control in soils stabilised with alkali-activated aluminosilicate byproducts. Chapter 1 introduces the background, motivation, and research objectives. Chapter 2 reviews the state of knowledge on waste-derived aluminosilicates and their role in soil stabilisation. Chapter 3 outlines the methodological framework used to address the research questions.

As illustrated in Figure 1.2, Chapters 4 to 7 present the major analytical components of the research. Chapter 4 reports a meta-analysis of the previously developed strength prediction models to understand their capabilities and limitations. Building on these insights, Chapter 5 reports on the compiles published datasets relevant to the strength of stabilised soils, and develops new machine learning models to predict unconfined compressive strength. Chapter 6 addresses identified gaps in the literature on swelling stabilisation through targeted experimental investigations. Finally, Chapter 7 develops new models to predict swelling behaviour using the experimental dataset generated in Chapter 6. Chapter 8 concludes the thesis, summarises key findings, and provides recommendations for future research.

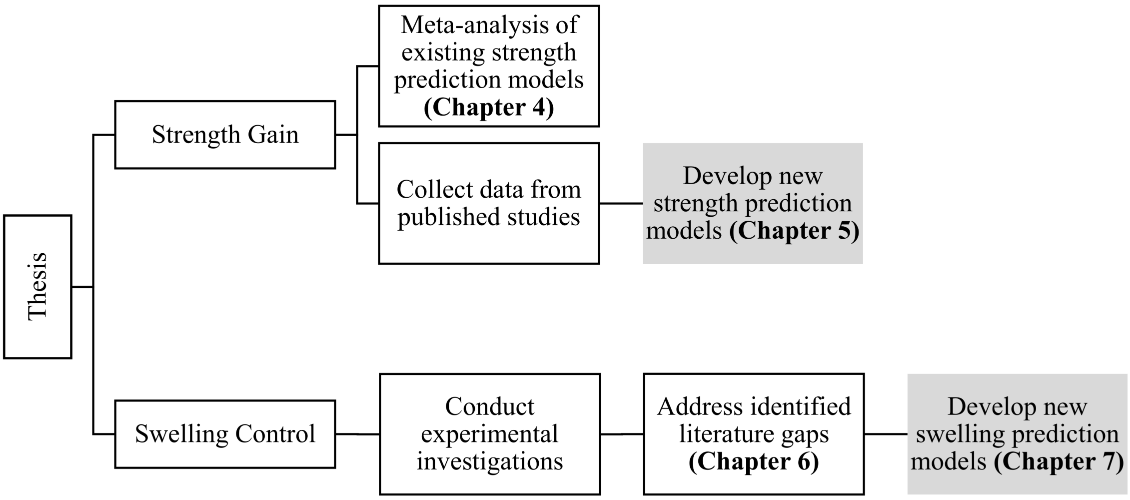


Figure 1.2. Overview of the thesis structure highlighting the key analytical components across strength gain (Chapters 4 & 5) and swelling control (Chapters 6 & 7)

CHAPTER 2

STATE OF THE ART

This chapter provides an overview of the current state of knowledge and practice in soil stabilisation and associated predictive modelling. It begins by discussing the influence of soil molecular structure and mineralogy on the engineering behaviour of soils. Then a review of laboratory analyses of problematic soils follows, highlighting key findings. Next, common stabilisation methods used to address problematic soils are briefly evaluated. The chapter then explores the synthesis of geopolymer-stabilised soils, focusing on the effects of key influencing factors on soil strength and swelling stabilisation. Recent advancements in predictive modelling techniques for soil stabilisation are also presented. Finally, a comparison of the most effective modelling approaches is provided, along with a summary of key literature gaps in both experimental research and predictive modelling.

2.1 Soil Molecular Structure and Mineralogy

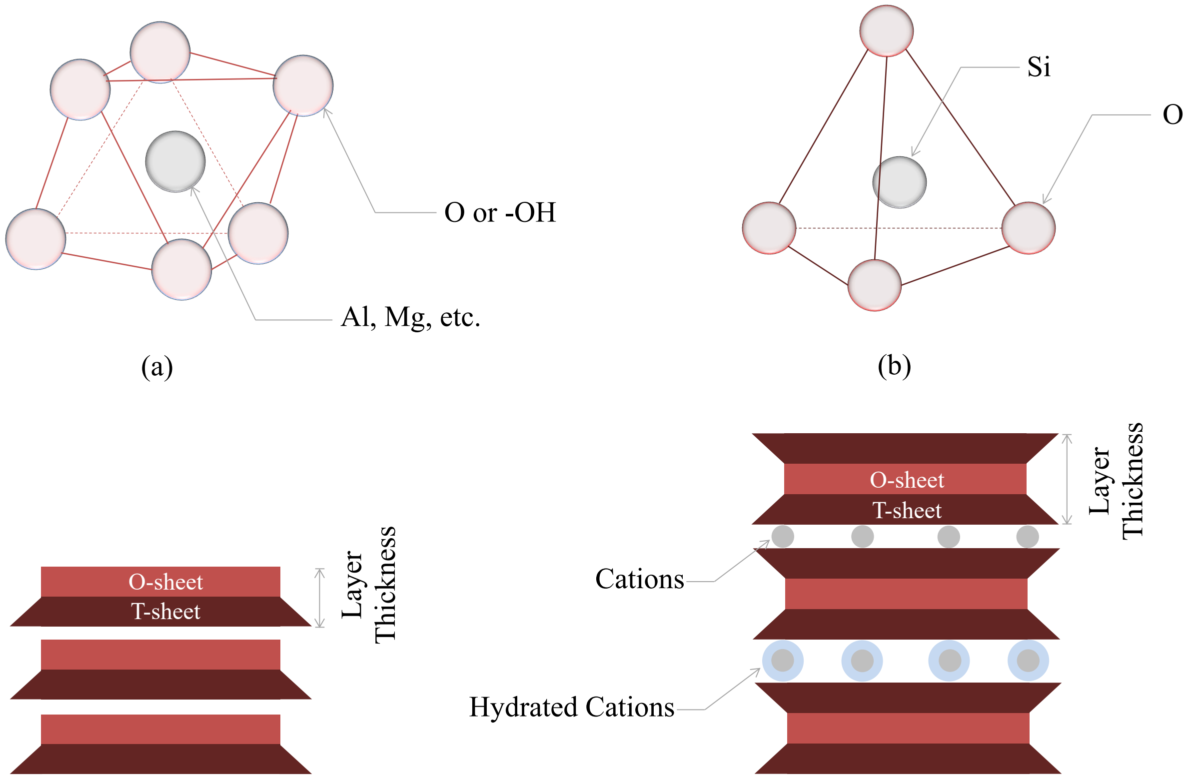
Understanding the problematic behaviour of soils requires an appreciation of their structural complexity at the molecular scale. Sands and silts, characterised by particle sizes greater than 2 μm, generally exhibit a granular physical structure with relatively simple molecular compositions, consisting predominantly of quartz. This structural simplicity results in correspondingly simpler behavioural characteristics making them non-reactive to changes in moisture conditions. Hence, they are not susceptible to swelling and shrinkage due to moisture changes.

In contrast, clay has a complex molecular structure. As shown in Figure 2.1, clayey soils are aluminosilicates consisting of flat layers or platelets of elements arranged in octahedral and tetrahedral configurations (Sposito, 2016). This structural complexity in clays gives rise to complex behavioural characteristics, e.g., swelling when exposed to water. As illustrated in Figure 2.1, in the octahedral arrangement (a), the central atom, typically aluminium or magnesium, is surrounded by six oxygen or hydroxide ions. In the tetrahedral arrangement (b), a silicon atom occupies the central position, surrounded by four oxygen atoms. These fundamental units are stacked in either 1:1 (octahedral–tetrahedral, O–T) (c) or 2:1 (tetrahedral–octahedral–tetrahedral, T–O–T) (d) layer configurations, to make up distinct types of clay minerals.

Table 2.1 presents the main clay mineral groups and their corresponding layer structures. Kaolinites are the only group of minerals with a 1:1 layer structure, whereas all other groups have a 2:1 layer configuration. Clay sheets carry a net negative surface charge (referred to as 'layer charge' in the table), which facilitates the adsorption of water molecules and the attraction of positively charged cations within the interlayer spaces. Higher layer charge values indicate a greater number of negatively charged sites, which enhances the clay's capacity for water absorption.

The negatively charged sites on clay surfaces attract and hold cations from water stored in pores. These exchangeable cations can vary in valency and play an important role in the clay's swelling behaviour. During cation exchange processes, monovalent cations, e.g., Na^{+}, are generally more weakly bound to clay surfaces and are easily hydrated, allowing water molecules to infiltrate between clay layers, causing swelling. In contrast, divalent cations, e.g., Ca_{2+} or Mg_{2+}, due to their higher charge and stronger electrostatic attraction to the clay surfaces, reduce the thickness of the diffuse double layer and limit water permeation into clay layers, thereby reducing the swelling potential. While the initial type of interlayer cation is determined during the geological formation of clay minerals, subsequent changes in the chemistry of water in clay pores can change the dominant exchangeable cations through cation exchange process.

The volumetric expansion and contraction of clayey soils due to changes in moisture conditions is primarily driven by the amount of smectite minerals such as montmorillonite (Patel, 2019; Chen, 1975). Montmorillonites, with their small particle size, ranging from 0.05 to 10^{-6} m in diameter, offer a large specific surface area (between 50 and 840 m^2/g), which allows for greater adsorption of water molecules and promotes prolonged water retention (Vorwerk et al., 2015). Hence, they are the most reactive clay minerals. On the other hand, kaolinite has almost no ability to absorb water due to its low surface charge and layered structure. Other clay minerals fall in between and are less reactive to water when compared with the smectites. Therefore, expansivity depends on the content of reactive clay minerals present in the soil.



(a) (b) (c) (d)

Figure 2.1. Molecular structure of clay particles: (a) a single octahedron of aluminium or magnesium surrounded by oxygen atoms or hydroxide ions, (b) a single tetrahedron of a silicon surrounded by oxygen atoms, (c) O-T structure of clay particles, and (d) T-O-T structure of clay particles

Table 2.1. Clay minerals, layer structure, and charge, and the typical cations in their inter-layer spaces (adapted from Sposito (2016))

Mineral Groups	**Layer Structure**	**Layer Charge**	**Typical Cations**
Kaolinite	1:1	< 0.01	None or minimal
Illite	2:1	1.2 to 1.7	Monovalent: K^+
Vermiculite	2:1	1.2 to 1.8	Divalent: Ca^{2+}, Mg^{2+}
Smectite	2:1	0.4 to 1.2	Monovalent/Divalent: Na^+, K^+, Ca^{2+}, Mg^{2+}
Chlorite	2:1	Variable	None or minimal

2.2 Laboratory Analysis of Problematic Soils

The unconfined compressive strength (UCS) test is a widely used method for evaluating soil strength (ASTM International, 2024). In contrast, there is currently no consensus in the literature about a standard testing method to assess swelling in soils (Vorwerk et al., 2015). National standards use different laboratory tests to analyse soil swelling. For example, in Australia, expansive soils are commonly characterised using a combination of Atterberg limits and shrink-swell tests (Vorwerk et al., 2015), whereas the Bureau of Indian Standards relies on Atterberg limits alone for classification (Patel, 2019) (see Table 2.2). These methods differ in practicality, particularly with respect to sample condition, material requirement, and testing complexity. Atterberg limit tests are simple, relatively fast to conduct, and require about 500 g of soil per test. In contrast, shrink–swell tests are more resource-intensive, typically requiring undisturbed soil samples to preserve natural structure. However, remoulded specimens can also be evaluated, generally requiring approximately 200 to 400 g of compacted material per test. Although shrink-swell tests provide more direct and meaningful quantification of expansive behaviour of clayey soils, their complexity, cost, and material requirements can limit widespread application.

Table 2.2. Classifying soil expansivity based on Atterberg limits (Patel, 2019)

Atterberg limits	**Low**	**Medium**	**High**	**Very High**
Liquid limit	20 to 35	35 to 50	50 to 70	70 to 90
Plasticity index	<12	12 to 23	23 to 32	>32

Another widely recognised approach to analyse soil expansivity is through the use of the activity parameter, originally proposed by Skempton (1984). Soil activity is defined as the ratio of the plasticity index to the percentage of clay-sized particles (less than 2 µm) in the soil. This method integrates both Atterberg limits and particle size distribution, offering a more realistic assessment of swell potential compared to classification systems that are solely based on plasticity index. While this parameter is commonly used in research settings (Ishibashi and Hazarika, 2015), it has not been universally adopted in geotechnical practice. The classification chart shown in Figure 2.2, has been used to categorise expansive soils based on their activity values. Soils with high activity exhibit a greater tendency for volumetric changes and are thus considered highly expansive, whereas soils with low activity values tend to be more stable (Pratama et al., 2021).

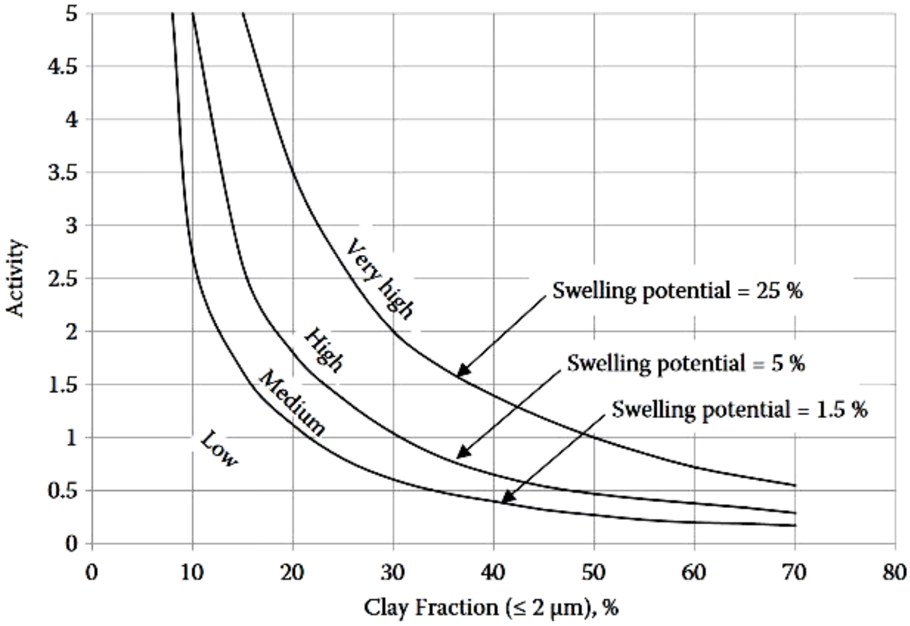


Figure 2.2. Swelling potential classification chart (Reproduced from Seed and Lundgren (1962))

Linear shrinkage test is another method used to analyse soil swelling. It requires about 300 g of disturbed soil enough to fill a standard shrinkage mould. The linear shrinkage parameter is calculated using Equation 2.1 (Pratama et al., 2021).

$$LS = \left(1 - \frac{L_{\mathrm{D}}}{L_0}\right) \times 100, \tag{2.1}$$

In Equation 2.1, LS denotes the linear shrinkage value; L_0 is the initial length of a soil sample moulded at a moisture content equal to, or greater than, the liquid limit; and L_{D} is the length after oven-drying. As shown in Table 2.3, soil expansivity can be categorised based on the LS value (Altmeyer, 1955; Chen, 1975).

Table 2.3. Clay expansivity based on linear shrinkage (Altmeyer, 1955; Chen, 1975)

LS	**Expansivity**
> 8 %	Critical degree of expansion
5 to 8 %	Marginal degree of expansion
0 to 5 %	Non-critical degree of expansion

The ASTM Standard D4546-21, provides another practical and widely accepted procedure for estimating the one-dimensional free swelling (1DFS) potential of expansive clays due to wetting under confined conditions (ASTM International, Committee D18, 2021). The test can either be carried out under a constant load or load-free. Unlike shrink–swell or swell pressure tests, which often require undisturbed soil specimens for best results, large sample volumes, and long testing durations, 1DFS test can be performed on remoulded specimens using significantly smaller quantities of soil, typically between 30 to 50 grams per test. Given the minimal sample requirements, this method is particularly useful in research projects in which a large number of tests need to be conducted under variable experimental conditions or, more generally, those constrained by time and material availability.

2.3 Soil Stabilisation Methods

Soil stabilisation methods can be broadly categorised into two groups: mechanical and chemical techniques (Jiao et al., 2024; Khazaei and Moayedi, 2019). These methods can be utilised individually or combined to achieve enhanced performance (Toksöz Hozatlıoğlu and Yılmaz, 2021). There are also other techniques such as biotechnical and thermal approaches, and consolidation with or without drains and consolidation with vacuum (Zhu et al., 2024; Schaefer et al., 2012), although their applications are generally site-specific and less widespread. The selection of a suitable stabilisation approach depends on the specific requirements of the project, including soil type, site conditions, performance expectations, as well as time and budget constraints.

One conventional approach to improving problematic soils is the ’cut-and-fill’ method, which involves the excavation of the unsuitable material and replacement with crushed rock (Abdila et al., 2022; Patel, 2019) or other engineered fill materials. Standards such as AS 3798–2007 (Standards Australia Limited, 2007) provide detailed procedures for achieving stable engineered fills, typically involving, removal of problematic soil, subgrade preparation and the placement of compacted fill in 200 mm thick layers. However, depending on the extent of excavation and filling required, this approach can lead to substantial project costs.

The above approach relies heavily on mechanical compaction to improve the physical properties of soils or backfill materials. Techniques such as static and dynamic compaction are commonly employed to achieve the required material density (Abdila et al., 2022). Other ground improvement methods are also adopted, depending on site-specific conditions, including vibro-compaction, dynamic compaction, blasting compaction, compaction grouting, and surface compaction (Schaefer et al., 2012).

An alternative approach to improving problematic soils is to backfill with compacted and chemically stabilised site-won materials. This method typically involves the addition of cementitious binders, such as cement, lime, and aluminosilicates, to enhance soil properties through specific chemical reaction mechanisms (Abdila et al., 2022).

Cement stabilisation rely on the hydration reactions, where water reacts with cement particles to form cementitious compounds, primarily calcium–silicate–hydrate (C–S–H) gel. The formation of C–S–H gel progressively binds soil particles together, leading to improvements in strength, stiffness, and durability. The process is relatively rapid, typically requiring about 30 days to achieve significant strength development. The efficiency and rate of stabilisation depend on factors such as the binder content, soil characteristics, moisture content, and curing conditions.

Lime stabilisation operates through two primary mechanisms: cation exchange and pozzolanic reactions. Initially, cation exchange occurs, where calcium ions (Ca^{2+}) from the lime replace weaker exchangeable cations in the interlayer space of clay particles, leading to a reduction in swell potential, densification, and the development of long-term strength (Sargent, 2015). This process reduces the thickness of the diffuse double layer surrounding the clay particles, thereby decreasing soil plasticity and promoting a more granular structure with enhanced friction angle and shear strength (Rogers and Glendinning, 1996). Lime stabilisation also relies on pozzolanic reactions, which require a high-pH environment and sufficient water availability. In these reactions, dissolved calcium ions (Ca^{2+}) interact with silicon dioxide (SiO_2) and aluminium oxide (Al_2O_3) present either within the soil matrix or introduced through binders, resulting in the formation of cementitious compounds such as C-S-H and calcium-aluminate-hydrates (C–A–H) (Sithole and Mashifana, 2020; Sargent, 2015; Liu et al., 2022). The pozzolanic process is gradual, with significant strength gains occurring over extended periods. Complete stabilisation can take up to one year as the cementitious products continue to develop with time (Sargent, 2015).

Aluminosilicate stabilisation primarily relies on the geopolymerisation process involving the dissolution of alumina (Al_2O_3) and silica (SiO_2) from precursors, including clay itself, leading to the formation of three-dimensional networks of geopolymers. A more detailed review of the geopolymerisation process is provided next, as it forms the core focus of this study.

2.4 Geopolymerisation

2.4.1 Background

Alkali-activated materials were first introduced by Glukhovsky in the 1950s. His work involved the use of volcanic ash activated with sodium hydroxide (Parthiban et al., 2022). In the early 1970s, the term ’geopolymer’ and the initial description of its chemical processes were proposed by Davidovits (Davidovits, 2020). Knowledge on the topic of geopolymerisation went through a gradual phase of development and maturation until the mid-1990s, in which Davidovits played a prominent role (Parthiban et al., 2022).

In the literature, the terms ’alkali-activated materials’ and ’geopolymer’ are used interchangeably (de Oliveira et al., 2022) for aluminosilicate material. However, for clarity and consistency, it is important to differentiate between the two terms (Provis and Bernal, 2014), with the latter defined as a special case of the former (de Oliveira et al., 2022). 'Alkali-activated materials' refers to activated aluminosilicate materials with a range of calcium content (de Oliveira et al., 2022) from none to high. High calcium content is defined as a ratio of calcium to combined aluminium and silicon (Ca : [Si + Al]) greater than 1 (de Oliveira et al., 2022). 'Geopolymer', on the other hand, refers to alkali-activated aluminosilicate materials with low calcium and high aluminosilicate content (de Oliveira et al., 2022; Abdullah et al., 2019) such as fly ash and metakaolin.

2.4.2 Synthesis of Geopolymerised Soils

Geopolymerisation is a multistage chemical process as illustrated in Figure 2.3. The process begins with (i) *dissolution*, during which hydroxide ions (OH^-) in the alkaline solution attack the aluminosilicate material, breaking it down into reactive species. Alumina and silica components react with hydroxide to form aluminate ($Al_2O_3 + 6OH^- \rightarrow 2AlO_4^{5-} + 3H_2O$) and silicate ($SiO_2 + 4OH^- \rightarrow SiO_4^{4-} + 2H_2O$) ions, respectively (Zhu et al., 2024). This is followed by (ii) *diffusion*, where the dissolved aluminate and silicate species migrate and disperse throughout the mixture. Subsequently, (iii) *polycondensation* occurs, during which reactive species polymerise to form an amorphous aluminosilicate gel network composed of Si–O–Si and Si–O–Al bonds. During this stage, aluminium (Al^{3+}) begins to substitute silicon (Si^{4+}) within the tetrahedral framework, introducing a net negative charge that is balanced by alkali cations such as Na^+. Finally, (iv) *hardening* takes place, where the aluminosilicate gel progressively reorganises into a stable geopolymer matrix. These reactions may run concurrently, rather than sequentially, depending on the mixing method, material characteristics, and curing conditions.

A general empirical formula has been provided for the geopolymerisation of aluminosilicates (Sata and Chindaprasirt, 2020):

$$Mn[-(SiO_2)_z - AlO_2]n.wH_2O \tag{2.2}$$

In Equation 2.2, *M* represents the alkaline cation (either K^+ or Na^+), *n* denotes the degree of consolidation (also referred to as the degree of polymerisation), *w* is the number of moles of water, and *z* is the Si:Al molar ratio. Si:Al molar ratios greater than 3 lead to the formation of a two-dimensional network of rubbery gels, whereas Si:Al molar ratios less than 3 result in a three-dimensional network of brittle and cementitious gels characterised by Si–O–Al bonds, which are particularly suitable for soil stabilisation applications. The geopolymer gel structure consists of tetrahedral SiO_4 and AlO_4 units interconnected by O_2 bridges (Abdullah et al., 2020).

Depending on the calcium content of the system, the final stabilisation product may consist of a sodium–aluminosilicate–hydrate (N–A–S–H) gel for low-calcium systems, or a calcium–aluminosilicate–hydrate (C–A–S–H) gel for high-calcium systems. In mixed systems containing both low and high-calcium precursors, a hybrid

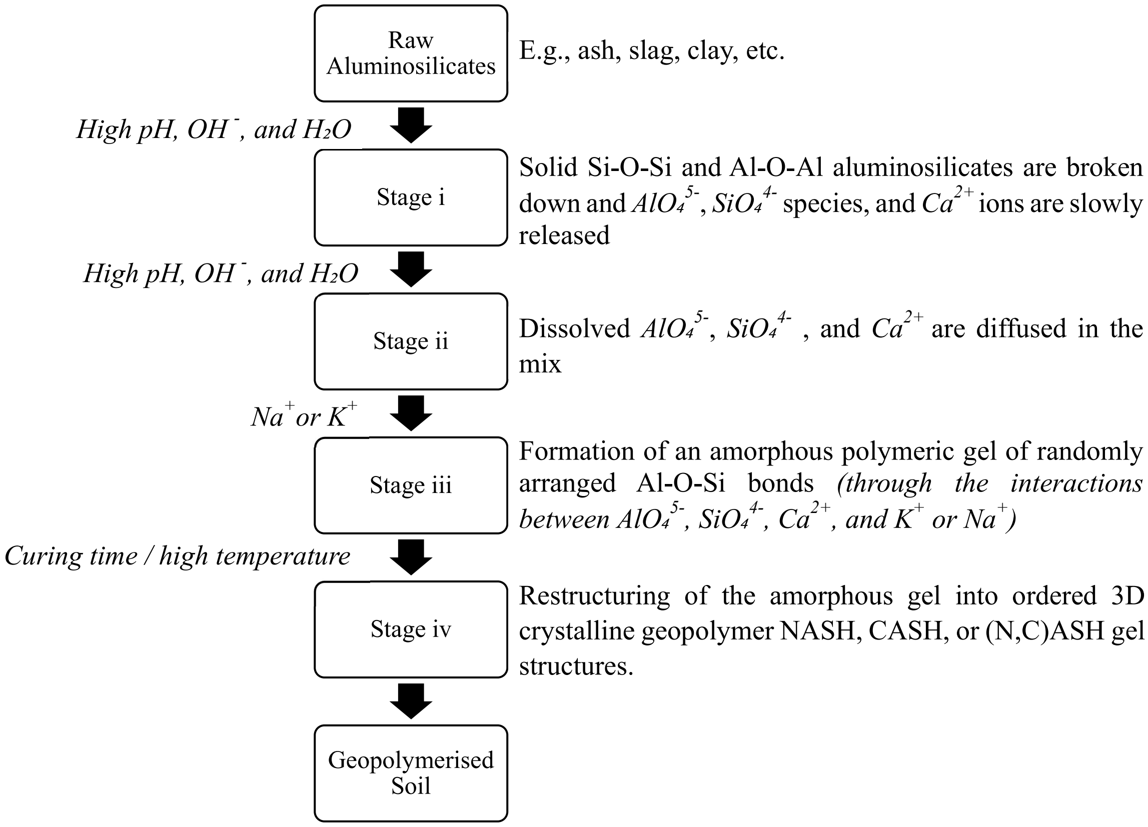


Figure 2.3. The process of soil geopolymerisation (Ahmed Hisseini et al., 2021; Marvila et al., 2021)

(N,C)–A–S–H gel structure is typically formed. Unlike N–A–S–H and (N,C)–A–S–H gels, the high-calcium C–A–S–H gel is not classified as a true geopolymer (Abdullah et al., 2020). This distinction arises because C–A–S–H gels are formed through reaction mechanisms similar to the hydration processes observed in cement stabilisation (Abdullah et al., 2020). The primary reaction mechanisms governing the formation of these gels can be represented as follows:

$$\mathrm{Na^{+} + AlO_4^{5-} + SiO_4^{4-} + H_2O \rightarrow N\text{-}A\text{-}S\text{-}H\ gel} \tag{2.3}$$

$$\mathrm{Ca^{2+} + AlO_4^{5-} + SiO_4^{4-} + H_2O \rightarrow C\text{-}A\text{-}S\text{-}H\ gel} \tag{2.4}$$

$$\mathrm{Na^{+} + Ca^{2+} + AlO_4^{5-} + SiO_4^{4-} + H_2O \rightarrow (N,C)\text{-}A\text{-}S\text{-}H\ gel} \tag{2.5}$$

2.4.3 Geopolymer Precursors

Geopolymer precursors, i.e., raw materials with high alumina (Al_2O_3) and silica (SiO_2) content, can be either naturally occurring or artificially produced (Ram and Mohanty, 2022). Natural aluminosilicates are found in nature and include materials such as pyroclastic rocks, tuffs, metakaolin, and calcined clays (Ram and Mohanty, 2022). Artificial aluminosilicates, on the other hand, are derived from industrial processes and include ash, slag, silica fume, burned clay, and shale (Ram and Mohanty, 2022). Among these, fly ash and slag are particularly important due to their wide availability as agro-industrial geopolymer precursors.

Fly ash is a by-product generated from coal-burning power plants (Puppala et al., 2015). It can be described as a rounded (Yaghoubi et al., 2018), transparent (glassy) (Ram and Mohanty, 2022), non-plastic fine silt material (Phanikumar, 2022) with particle sizes typically less than 20 µm (Yao et al., 2015). The spherical particles of fly ash can be either hollow or solid (Bhatt et al., 2019). The pH of fly ash varies widely, ranging from 1.2 to 12.5, however, it generally tends to be alkaline, depending on the calcium-to-sulphur ratio (Yao et al., 2015).

Fly ash has a complex chemical composition, comprising over 300 different mineral phases (Yao et al., 2015). The chemical composition plays a critical role in determining its effectiveness as a cementitious binder in soil stabilisation. Among the many constituents, calcium oxide (CaO), silicon dioxide (SiO_2), aluminium oxide (Al_2O_3), iron oxide (Fe_2O_3), magnesium oxide (MgO), and sulphur trioxide (SO_3) are particularly important for soil stabilisation applications.

Certain properties of fly ash are influenced by the type of coal burned (Bhatt et al., 2019). Coals are generally categorised into four groups: sub-bituminous, lignite, anthracite, and bituminous (Bhatt et al., 2019). Sub-bituminous and lignite coals, also known as brown coals, produce high-calcium fly ash (Puppala et al., 2015), whereas anthracite and bituminous coals, which are black, yield low-calcium fly ash (Puppala et al., 2015). In Australia, most coal-fired power stations burn black coal, resulting in the production of low-calcium but high aluminosilicate content fly ash (CRC for Low Carbon Living, 2015). Victoria is the only state that generates some quantities of high-calcium fly ash (CRC for Low Carbon Living, 2015).

The physicochemical characteristics of fly ash are also influenced by combustion temperature and cooling rate (Ram and Mohanty, 2022). Rapid cooling tends to produce spherical, non-crystalline (glassy) fly ash particles, while slow cooling results in the formation of crystalline structures (Ram and Mohanty, 2022).

The American Society for Testing and Materials (ASTM) standard C618 classifies fly ash into two broad categories based on its chemical composition (ASTM C618, 2019). Fly ash with high calcium content is designated as Class C, whereas fly ash with low calcium content is classified as Class F. Calcium in fly ash typically exists in the form of calcium sulphate and calcium hydroxide, with contents ranging from 1–12% in Class F and 30–40% in Class C fly ash (Puppala et al., 2015). Although originally developed for cement manufacturing, the ASTM classification system has been widely adopted in geotechnical engineering studies. Class C fly ash is characterised by a calcium oxide (CaO) content greater than 10%, a combined silicon dioxide (SiO_2), aluminium oxide (Al_2O_3), and iron oxide (Fe_2O_3) content of greater than 50%, and a sulphur trioxide (SO_3) content of less than 5%. In contrast, Class F fly ash is defined by a CaO content of less than 10%, a combined SiO_2, Al_2O_3, and Fe_2O_3 content of greater than 70%, and an SO_3 content of less than 5%.

Fly ash can also be classified as acidic, mildly alkaline, and strongly alkaline based on its pH. Acidic ashes have pH values in the range of 1.2 to 1.7. Mildly alkaline ashes have pH values of 8 to 9. Strongly alkaline ashes have pH values in the range of 11 to 13 (Bhatt et al., 2019).

Slag is a by-product generated by the iron manufacturing industry (de Oliveira et al., 2022). It can exist either as fine particles or as a granular, gravel-like material (Parthiban et al., 2022). The ground granulated blast furnace slag (slag) form is the finest and most commonly preferred for soil stabilisation applications (Puppala et al., 2015). Chemically, slag is comparable to cement, consisting primarily of calcium oxide (CaO), silicon dioxide (SiO_2), aluminium oxide (Al_2O_3), and magnesium oxide (MgO) (Abdila et al., 2022). The particles are typically irregular and exhibit angular to sub-angular shapes (Yaghoubi et al., 2018). Slag possesses high alumina, silica, and calcium contents, making it a suitable material for enhancing soil properties (Abdila et al., 2022).

Previous studies have demonstrated that granulated blast furnace slag can significantly improve the strength and durability of soils while reducing their permeability (Abdila et al., 2022). Although slag is slightly more expensive than fly ash, its use can be optimised by blending it with fly ash not only to enhance cost-efficiency but also to improve the stabilisation outcome (Provis and Bernal, 2014).

2.4.4 Alkaline Activators

Until recently, potassium hydroxide was among the most commonly used alkali activators in soil stabilisation. However, sodium hydroxide has emerged as a more effective and cost-efficient alternative (Odeh and Al-Rkaby, 2022). Compared to potassium hydroxide, sodium hydroxide has a greater ability to free aluminate and silicate monomers, thereby enhancing geopolymer gel formation (Odeh and Al-Rkaby, 2022). It is also more widely available. Other hydroxide-based activators, such as lithium hydroxide, rubidium hydroxide, and cesium hydroxide, have been explored less frequently and primarily in the context of large-scale or specialised applications (Provis and Bernal, 2014).

The addition of silicates, such as sodium silicate, to an alkali-activator mix enhances its ability to reduce soil permeability, thereby improving the overall effectiveness of soil stabilisation (Provis and Bernal, 2014). Furthermore, the inclusion of silicates can lower the risk of efflorescence formation (Provis and Bernal, 2014), where salts migrate and crystallise on the surface of polymerised materials.

Hydroxide-based activators offer enhanced resistance to high temperatures, tolerating conditions up to 1000°C, and improve workability due to the low viscosity of the resulting mix (Provis and Bernal, 2014). Empirical studies recommend a minimum alkaline molar concentration of 3 M for effective soil stabilisation, corresponding to a pH of approximately 12.36 (Eades and GRTh'I, 1966; Mypati and Saride, 2022). However, certain applications, particularly those involving low-calcium binders, may require higher molar concentrations (Provis and Bernal, 2014). While higher molarity improves chemical reactivity, it also increases the hazard level, as highly concentrated alkaline solutions can be corrosive and dangerous to handle (Provis and Bernal, 2014). To address these concerns, solid-state alkaline activators have been proposed as a safer alternative for soil stabilisation projects, offering simpler handling and transportation compared to liquid activators (Shah et al., 2020; Oderji et al., 2019).

2.5 Published Reviews of the Use of Aluminosilicates

Soil stabilisation through the use of alkaline activated aluminosilicates has emerged as a significant area of research within the broader field of chemical soil stabilisation (Sharma and Kumar, 2020). Since the year 2000, the number of publications focused on this approach has grown exponentially (Sharma and Kumar, 2020).

Since the early 1980s, numerous review papers have examined the engineering and environmental benefits of aluminosilicates, with a growing body of work focusing specifically on their application in soil stabilisation. Several studies have highlighted the potential of agricultural-based ash wastes in improving the engineering properties of problematic soils, particularly expansive and organic soils (Adedokun and Oluremi, 2019; Roy et al., 1981; Khanday et al., 2021). Others have explored the use of aluminosilicates as engineered fill and pavement materials, offering viable alternatives to conventional binders (Andavan and Pagadala, 2020; Roy et al., 1981; Ahmaruzzaman, 2010).

While less directly related to soil stabilisation, other research has addressed the broader use of aluminosilicates in sustainable absorbents (Ahmaruzzaman, 2010; Navin et al., 2018) and low-permeability barriers (Roy et al., 1981; Devarangadi and M, 2020). Other studies have investigated the environmental implications of aluminosilicate use (Habert et al., 2011; Yao et al., 2015; Munawar et al., 2021), strategies for enhancing durability in aggressive environments (Millán Ramírez et al., 2021; de Oliveira et al., 2022), and the chemical and microstructural mechanisms underlying geopolymerisation (Provis and Bernal, 2014; Bhatt et al., 2019; Abdullah et al., 2020; Sharma and Kumar, 2020; Ram and Mohanty, 2022).

The following sections summarizes relevant findings from an extensive number of experimental studies related to the application of aluminosilicates in soil stabilisation.

2.6 Empirical Findings on the Use of Aluminosilicates

The majority of research on soil stabilisation by aluminosilicates has focused on the use of fly ash, either as a stand-alone binder or in combination with slag (Sharma and Kumar, 2020). High-silica wastes have also been evaluated for their stabilisation potential, although to a lesser extent than fly ash. Pourakbar et al. (2015) examined the use of palm oil fuel ash in stabilising clayey soils. Sharma and Sharma (2021) investigated the combined effects of rice husk ash, demolition waste, and lime in improving the properties of low-strength clayey subgrade materials. Blesson and Rao (2023) reviewed the application of various agro-industrial aluminosilicate wastes in the stabilisation of problematic soils via alkali activation. Similarly, Adedokun and Oluremi (2019) provided a comprehensive review of agricultural ash waste applications in the treatment of lateritic soils. A key finding across these studies is that many agricultural-based aluminosilicate wastes exhibit chemical compositions comparable to low-calcium fly ash, making them viable alternatives in alkali-activated soil stabilisation systems (Blesson and Rao, 2023).

The empirical findings discussed in the following sections contribute to identifying and evaluating the key parameters influencing stabilisation performance through the alkali-activation of aluminosilicates. A summary of the relevant empirical findings related to increasing strength and reducing swelling of problematic soils is provided below highlighting the diversity of soil types, binder and activator combinations, and curing conditions studied to date. Key findings relevant to soil stabilisation have also been synthesised.

2.6.1 Strength Development

Geopolymerisation is influenced by a range of factors, including soil characteristics, chemical composition, activator molarity, curing temperature and time, and water content. Empirical studies have shown that high clay content generally reduces strength development (Toufigh et al., 2022), whereas the presence of sand and low organic content enhances strength by optimising the activator-to-ash ratio and improving reaction efficiency (Hebib and Farrell, 2003; Leong et al., 2018; Odeh and Al-Rkaby, 2022; Tastan et al., 2011). Binders characterised by low calcium but high alumina and silica contents promote long-term strength gains (Abdila et al., 2022; Abdullah et al., 2020; Cristelo et al., 2012; Wardhono, 2018), while high calcium content favours early strength development (Cristelo et al., 2012; Wardhono, 2018).

Blending low-calcium ash with slag has been shown to significantly improve strength, reduce the quantity of activator required, and mitigate environmental impacts and costs (Abdila et al., 2022; Abdullah et al., 2020, 2021; Lakkimsetti and Nayak, 2022; Pham et al., 2022; Shah et al., 2020; Toufigh et al., 2022; Wardhono, 2018). Magnesium oxide (MgO) content is also beneficial to strength development (Provis and Bernal, 2014), whereas elevated sulphate content (SO_3) and high Loss on Ignition (LOI) levels in binders have negative effects (Abdullah et al., 2021; Chen et al., 2009; Detphan et al., 2022; Shah et al., 2020).

Curing conditions, particularly temperature and time, play a critical role in the stabilisation process. Elevated curing temperatures accelerate early strength development (Cristelo et al., 2012; Hoy et al., 2016; Leong et al., 2018; Phummiphan et al., 2018; Ram and Mohanty, 2022; Sukprasert et al., 2021; Suksiripattanapong et al., 2021); however, maintaining adequate moisture during curing is essential to ensure long-term strength development (Phummiphan et al., 2018).

Factors related to the mixture design, including binder, water, and activator contents, are equally important for strength optimisation. Increasing binder and activator contents generally improves strength (ASTM C618, 2019; Odeh and Al-Rkaby, 2022); however, exceeding optimal ratios can lead to strength reductions (Odeh and Al-Rkaby, 2022). The molar concentration of the alkaline activator significantly influences the rate and extent of the stabilisation process (Abdullah et al., 2020; Chen et al., 2022).

2.6.2 Swelling Stabilisation

Empirical findings reveal that aluminosilicate binders have considerable potential to reduce swelling. For instance, alkali-activated ash has been shown to reduce linear shrinkage by up to 80%, however, a slight increase in swelling has been observed when ash and activator contents are not optimally proportioned (Mypati and Saride, 2022; Nalbantoğlu, 2004). Mypati and Saride (2022) reported that activator-to-ash ratios of 0.75 or lower may lead to a marginal increase in soil swelling relative to the swelling limit achieved by the optimum activator content, potentially compromising both swelling reduction and strength development. These findings highlight the importance of optimising the proportions of ash and alkaline activator not only for controlling swelling but also for achieving adequate strength performance (Mypati and Saride, 2022; Sukprasert et al., 2021).

In a recent study, Saride and Mypati (2024) reported reductions in swelling and shrinkage of up to 72%. Even in the absence of an alkaline activator, the combination of ash with slag has demonstrated the potential to reduce the liquid limit and plasticity index of expansive soils (Mandal and Singh, 2007; Phanikumar, 2022). Slag plays a critical role in stabilisation, particularly in enhancing early strength development and accelerating reaction rates due to its high calcium oxide (CaO) content (Cristelo et al., 2012; Wardhono, 2018). However, the specific role of an activator and the influence of its molarity in ash–slag systems remain under-explored in the context of swelling behaviour.

Several studies have investigated the use of aluminosilicates in combination with other additives such as lime, gypsum, and metakaolin to enhance soil stabilisation performance (Khadka et al., 2020; Zha et al., 2024). Zha et al. (2024) examined the stabilisation of expansive soils using fly ash blended with phospho-gypsum and lime. Their results indicated that phospho-gypsum, used alone or in combination with ash, led to a modest reduction in the plasticity index (approximately 5%).

However, the addition of 2% lime significantly improved performance, reducing the plasticity index by approximately 23%. In another study, Khadka et al. (2020) assessed the performance of alkali-activated ash combined with metakaolin, lime, and gypsum. They found that the effectiveness of ash–metakaolin blends could be optimised based on their chemical compositions, particularly the alumina, silica, and sodium contents. Khadka et al. (2020) also showed that swelling potential was minimised when the concentration of additives such as lime and gypsum ranged between 6% and 9.5% by weight of the binder. Gypsum, in particular, was found to be more effective than lime in reducing swelling.

2.6.3 Durability and Environmental Impact

The leachability of cementitious elements is a critical factor influencing the long-term durability of geopolymer-stabilised soils. This is particularly important during the early stages of curing, when the progression of chemical reactions relies on the continued availability of reactive species such as calcium, silica, and alumina. In addition, studying the changes in pH levels in the vicinity of the stabilised zone is essential for assessing the potential environmental impact of stabilisation, particularly on sensitive natural ecosystems, groundwater quality, and adjacent infrastructure.

Avirneni et al. (2016) investigated leachate characteristics from alkali-activated ash-stabilised recycled asphalt pavement, described as ’sandy gravel’, with added calcium carbonate and sodium. Leachate parameters, including pH, calcium, and sodium concentrations, were monitored across multiple pore volumes. Elevated levels of calcium and sodium were observed during the first four pore volumes, after which concentrations declined, while pH gradually decreased over 14 cycles. Durability was assessed by measuring weight loss under wetting–drying cycles, which revealed a gradual mass loss over the first ten cycles before stabilising. However, the study did not evaluate how wetting–drying cycles influence leachability.

In a related study, Abdullah et al. (2019) examined pH variation in leachate from stabilised clay over nine pore volumes. They reported a relatively large decline in pH from 13.53 to 12.85 over the first three pore volumes, followed by a slower reduction until the sixth pore volume, with minimal change thereafter. To the best of the authors’ knowledge, no study to date has directly investigated the effect of wetting–drying cycles on the alkalinity or leaching behaviour of stabilised soils. This remains a critical gap, as high pH has the potential to negatively impact sensitive natural ecosystems.

An additional gap is the leaching of cementitious elements such as calcium, silicon, and aluminium, as it may adversely affect the long-term performance of stabilised soils, an area that also requires further investigation.

2.7 Modelling Soil Stabilisation

The conventional approach to chemical soil stabilisation, as illustrated in Figure 2.4, involves experimentally assessing the effects of individual variables, such as binder content, curing time, activator content and molarity, and curing temperature and time, on a target soil property e.g., UCS or swelling potential. While this approach can be effective for site-specific applications, it is time consuming, labour-intensive, and costly due to the extensive sampling and laboratory testing required. Furthermore, its purely empirical nature limits the ability to generalise findings to other problematic sites with different soils and environmental conditions due to the absence of a unified theory to predict behaviour across diverse soil types.

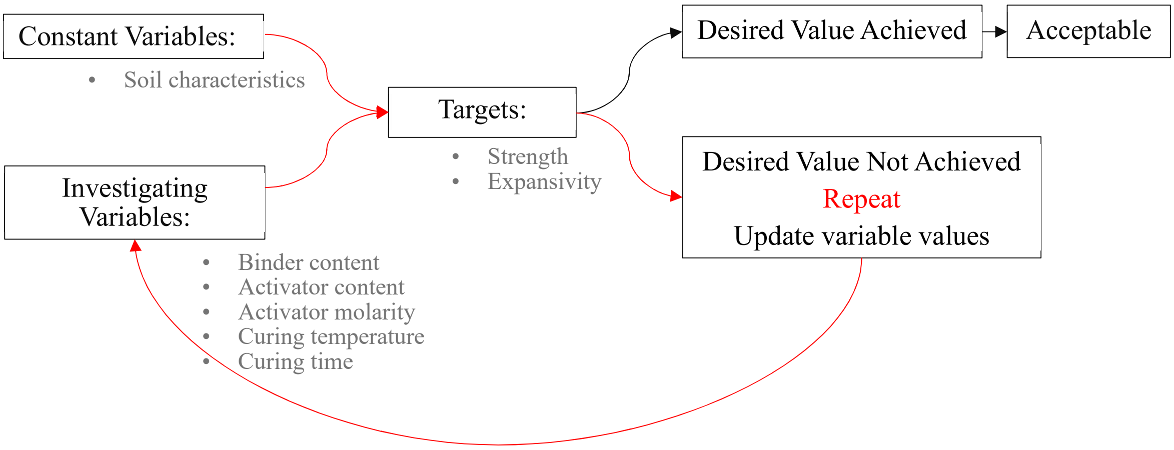


Figure 2.4. The conventional approach to soil stabilisation relies on cyclic laboratory testing of experimental variables until the desired outcome is achieved

To overcome these limitations, predictive models targeting a range of soil properties have been developed based on existing experimental studies. Reliable and versatile predictive models have the potential to reduce the amount of trial and error laboratory investigations by limiting them to a narrow range of validation tests.

2.7.1 Modelling Approaches

Broadly, three different approaches have been pursued in developing predictive models in the broader engineering context. The first and most commonly used approach for predicting the engineering behaviour of a material is the development of phenomenological models, based on empirically determined constitutive relationships and mass and momentum conservation equations. The second approach starts from fundamental micro-mechanical laws of physics and upscale equations to tackle specific engineering contexts and problems. The first approach is usually considered the most pragmatic and accessible. However, in the case of geopolymerisation of problematic soils, it suffers from several limitations, the most important of which is the complexity of processes to be captured.

The second approach, while in principle the most rigorous from a theoretical point of view, remains computationally expensive, often prohibitively so, while requiring input data that are difficult to determine in practice. These two obstacles generally rule it out as a tool to be used by engineers in designing materials, systems, and processes, except in very specific research contexts. Combinations of the first and second approaches have been developed too, but remain, by and large, confined to highly specialised research contexts and their added complexity, if anything, tends to make them even less suitable for engineering design and practice than either of the two individual approaches.

The third approach is to apply statistical or machine learning methods, that do not require prior theories of behaviour and are instead driven by data. The approach is hence practical and relatively easy to implement while producing ready-to-use and computationally inexpensive predictive models. Consequently, it has become the most widely adopted method for the development of predictive models in soil stabilisation.

2.7.2 Overview of Advances in Predictive Modelling for Soil Stabilisation

Given the rapidly growing body of literature on predictive modelling in soil stabilisation, it is not feasible to review every model proposed to date. This review therefore focuses on a representative subset of recent developments that are methodologically diverse and particularly relevant to waste-based aluminosilicate soil stabilisation. While the primary emphasis is on models that utilise agro-industrial aluminosilicate by-products such as fly ash and slag, a limited number of models incorporating cement and lime have also been included in this review to extract transferable insights into modelling strategies, methodologies, and performance evaluation.

Since 2015, there has been a notable increase in the number of published studies addressing predictive modelling in soil stabilisation (Mozumder and Laskar, 2015; Consoli et al., 2016; Leong et al., 2018; Hanandeh et al., 2020; Sari Ahmed et al., 2020; Nissanka et al., 2021; Pham et al., 2021, 2022; Zeini et al., 2023). Table 2.4 provides a summary of the predictive models reviewed in this study, including their target output parameters and the modelling techniques used. Broadly, two approaches have been followed including conventional regression methods (RM) and machine learning (ML) techniques. ML techniques explored are evolutionary algorithms, including genetic programming (GP) and gene expression programming (GEP), artificial neural networks (ANN), and one type of tree-based ensemble method called random forests (RF).

While UCS remains the most commonly modelled parameter, several studies have also developed models for other geotechnical parameters, such as Resilient Modulus (M_r), Liquid Limit (LL), Plasticity Index (PI), Free Swell Index (FSI), Maximum Dry Density (MDD), and Optimum Moisture Content (OMC). These models reflect the growing interest in using data-driven approaches to predict a broader range of engineering soil properties.

Table 2.4. Existing predictive models, their adopted approaches, and target parameters

Year	**Source**	**Target**	**Method**	**Model ID**
2015	Mozumder and Laskar (2015)	UCS	ANN	ANN-Moz
			RM	RM-Moz
2016	Consoli et al. (2016)	UCS	RM	RM-Con
2018	Leong et al. (2018)	UCS	ANN	ANN-Leo
			GP	GP-Leo
2020	Hanandeh et al. (2020)	M_r	ANN	ANN-Han
			GEP	GEP-Han
2020	Sari Ahmed et al. (2020)	LL	RM	RM-Sar-1
		PI	RM	RM-Sar-2
		UCS	RM	RM-Sar-3
		FSI	RM	RM-Sar-4
		MDD	RM	RM-Sar-5
		OMC	RM	RM-Sar-6
2021	Nissanka et al. (2021)	OMC	ANN	ANN-Nis-1
		MDD	ANN	ANN-Nis-2
		UCS	ANN	ANN-Nis-3
		UCS	ANN	ANN-Nis-4
2021	Pham et al. (2021)	UCS	GEP	GEP-Pha-1
2022	Pham et al. (2022)	UCS	GEP	GEP-Pha-2
2023	Zeini et al. (2023)	UCS	RF	RF-Zei

Extensive research has been undertaken to identify the critical factors influencing soil stabilisation and to assess their effects under varying conditions (Abdullah et al., 2020; Sharma and Kumar, 2020; Provis and Bernal, 2014; Bhatt et al., 2019; Ram and Mohanty, 2022). These studies form the basis for selecting appropriate predictors for a given modelling objective. As discussed in subsequent sections, researchers have explored a range of predictor combinations, offering valuable insights into the relative importance of different predictors across various modelling approaches. A consolidated list of predictors and modelling parameters used in the development and training of the reviewed models is provided in Table 2.5. To ensure consistency and facilitate meaningful comparison, predictor acronyms and symbols have been standardised in this study, with modifications made where necessary to align with the nomenclature adopted across all evaluated models.

Table 2.5. List of predictors and predicted outputs used in the development of existing models

Predictors and Parameters	Acronym
Properties of the host soil:	
Clay content	*clay%*
Silt content	*silt%*
Sand content	*sand%*
Soil organic content	*organic%*
Plastic Limit of the untreated soil	*PL_0*
Liquid Limit of the untreated soil	*LL_0*
Plasticity Index of the untreated soil	*PI_0*
Moisture content of the untreated soil	*MC_0*
UCS of the untreated soil	*UCS_0*
FSI of the untreated soil	*FSI_0*
MDD of the untreated soil	*MDD_0*
Dry unit weight of the untreated soil	*DUW_soil*
OMC of the untreated soil	*OMC_0*
Properties of the mixture:	
Slag content	*slag%*
Host soil content	*soil%*
Mix water content	*water%*
Cement content	*cement%*
Fly ash content	*ash%*
Class C ash content	*ash_C%*
Class F ash content	*ash_F%*
Lime content	*lime%*
Volumetric lime content	*vol_lime*
Dry unit weight of the lime	*DUW_lime*
Dry unit weight of the ash	*DUW_ash*
Dry unit weight of the mix	*DUW_mix*
NaOH content	*SH%*
NaOH Molar concentration	*M*
Na_2SiO_3 content	*SS%*
Na_2SiO_3:NaOH ratio	*SS_to_SH*
Activator to ash ratio	*A_to_ash*
Activator to binder ratio	*A_to_B*
Fly ash to untreated soil ratio	*ash_to_soil*
CaO content in ash	*CaO_ash%*
CaO:SiO_2 ratio in fly ash	*Ca_to_Si_ash*
Loss of ignition of ash	*LOI_ash%*
Na to Al ratio of the mix	*Na_to_Al_mix*
Si to Al ratio of the mix	*Si_to_Al_mix*
Curing conditions:	
Curing time (days)	*t_curing*
Curing temperature (°C)	*T_curing*

2.7.3 Conventional Regression Methods

Two studies have developed eight regression-based models to predict various geotechnical properties, including UCS, LL, PI, FSI, MDD, and OMC (Consoli et al., 2016; Sari Ahmed et al., 2020). These models use different combinations of predictors, as listed in Table 2.6, and are primarily based on published experimental datasets. Mathematical formulas of the models are presented in Appendix A (Equation A.1 to Equation A.9). Conventional regression approaches offer interpretability and simplicity, making them a practical starting point for modelling efforts despite their limitations in capturing complex non-linear interactions that exist in chemical soil stabilisation.

Table 2.6. Target parameters and predictors of the regression models

Model ID	**Target & Predictors**
RM-Moz	Target: UCS Predictors: LL_0, PL_0, $soil\%$, $ash\%$, M, A_to_B, $Na\_to\_Al\_mix$, $Si\_to\_Al\_mix$.
RM-Con	Target: UCS Predictors: DUW_mix, $lime\%$, and $ash\%$.
RM-Sar-1	Target: LL Predictors: LL_0, $ash\%$.
RM-Sar-2	Target: PI Predictors: PI_0, $ash\%$.
RM-Sar-3	Target: UCS Predictors: UCS_0, $ash\%$.
RM-Sar-4	Target: FSI Predictors: FSI_0, $ash\%$.
RM-Sar-5	Target: MDD Predictors: MDD_0, $ash\%$.
RM-Sar-6	Target: OMC Predictors: OMC_0, $ash\%$.

A multivariate regression model, RM-Moz, was developed by Mozumder and Laskar (2015) to predict the 28-day UCS of geopolymerised soils. The model estimates UCS in MPa based on a set of geotechnical and chemical predictors, formulated in a closed-form equation. The formulation is presented in Equation A.1.

Another UCS model, RM-Con, was proposed by Consoli et al. (2016). This model expresses UCS as a function of ash and lime content, the dry unit weight of the soil–lime mixture, and the dry unit weight of lime. The general formulation of the model is presented in Equation A.2. Beside including lime, the model has been recommended only for scenarios that meet the following conditions: (i) $ash\%$ should be within 0 to 25%, (ii) $lime\%$ should be within 3 to 9%, and (iii) DUW_mix should be within 14 to 17 kN/m^3.

Six simple regression models were proposed by Sari Ahmed et al. (2020), linking fly ash content to LL, PI, UCS, FSI, MDD, and OMC of stabilised soils. These relationships are represented by linear equations and are presented in Equation A.4 to Equation A.9, respectively.

2.7.4 Evolutionary Methods

Evolutionary methods are inspired by Darwin's theory of natural selection (Pham et al., 2022). Among the various subsets of these methods, GP and GEP have been applied to develop predictive models for soil stabilisation. A key advantage of GP and GEP is their ability to produce interpretable mathematical expressions derived from expression trees, allowing the final model to be represented as a single explicit formula. Five evolutionary models have been developed targeting (UCS) and resilient modulus (M_r) of stabilised soils. Mathematical expressions of these models are presented in Appendix A (Equation A.10 to Equation A.13). A summary of the predictors used in these models is provided in Table 2.7.

Table 2.7. Target parameters and predictors of models based on evolutionary methods

Model ID	Target & Predictors
GEP-Pha-1	Target: UCS Predictors: *PL_0*, *LL_0*, *clay%*, *silt%*, *sand%*, *organic%*, *water%*, *t_curing*, *cement%*, *ash%*, *CaO_ash%*, *Ca_to_Si_ash*, *LOI_ash*.
GEP-Pha-2	Target: UCS Predictors: *PI_0*, *clay%*, *silt%*, *sand%*, *water%*, *t_curing*, *lime%*, *cement%*, *ash_C%*, *ash_F%*, *slag%*.
GP-Leo	Target: UCS Predictors: *soil%*, *ash%*, *water%*, *SS%*, *SH%*, *SS_to_SH*, *A_to_ash*.
GEP-Han	Target: $\mathrm{M_r}$ Predictors: *PI_0*, *MC_0*, *clay%*, *silt%*, *OMC_0*, *cement%*, *lime%*, *ash%*.

The GP-Leo model, developed by Leong et al. (2018), was trained and tested using a relatively small dataset comprising 64 test records. Its structure consisted of four interconnected sub-expression trees, mathematically represented in Equation A.10. Although the model includes several input parameters, it omits key soil characteristics such as Atterberg limits, particle size distribution, and the chemical composition of both the host soil and the ash. Furthermore, all specimens were cured at a constant temperature of 100°C, and as a result, curing temperature was not included as a predictor. The use of such an elevated curing temperature limits the model's applicability in practical in-situ soil stabilisation scenarios, where ambient or moderately elevated temperatures are more appropriate.

The GEP-Han model, developed by Hanandeh et al. (2020), was designed to predict the resilient modulus (M_r) of stabilised soils. A dataset comprising 125 test records sourced from published literature was used for training and testing. As summarised in Table 2.7, the model incorporates eight predictors. Its structure comprises three expression trees, which are mathematically represented in Equation A.11. The model estimates M_r in kilo-pounds per square inch (Ksi), and employs floor functions, denoted by the symbols '$\lfloor$' and '$\rfloor$', within its formulation. However, it does not account for several important factors, including sand content, the chemical composition of the clay and binders, or the presence and concentration of alkali activators.

Another *UCS* model, GEP-Pha-1, developed by Pham et al. (2021), was trained and validated using a dataset comprising 489 test records of data compiled from previously published studies. This represents a significantly larger dataset than those used in earlier models. The mathematical formulation of the model GEP-Pha-1 is presented in Equation A.12. While it is one of the few models in the literature that partially incorporates chemical composition as an input, it relies on cement-based stabilisation.

Following the development of GEP-Pha-1, Pham et al. (2022) proposed an improved predictive model, referred to as GEP-Pha-2. This model was developed using a larger dataset comprising 1,183 test records compiled from published experimental studies. The set of predictors used in GEP-Pha-2 is summarised in Table 2.7. The main differences between GEP-Pha-2 and its predecessor lie in the selection of predictors: $CaO\_ash\%$ and $Ca\_to\_Si\_ash$ ratio, used in GEP-Pha-1, were replaced by two new parameters including $ash\_C\%$ and $ash\_F\%$. Additionally, $slag\%$ was introduced as a new predictor. The resulting relationship between UCS and its predictors is mathematically represented in Equation A.13.

Model GEP-Pha-2 relies on a combination of binders, including lime, cement, fly ash, and slag, for stabilising low-strength soils. However, there is no clear consensus in the literature on the adequacy of classifying ash wastes into the two broad cement-based categories, Class C and Class F, for use in soil stabilisation. This binary classification, although widely adopted, may be inadequate for chemical stabilisation contexts, where even slight variations in the concentrations of key oxide compounds (e.g., CaO, SiO_2, Al_2O_3, Fe_2O_3) may significantly affect reaction mechanisms and the overall effectiveness of stabilisation.

2.7.5 Artificial Neural Networks

ANN's have gained increasing popularity in geotechnical engineering due to their ability to capture complex, non-linear relationships within data (Zhang et al., 2021). Eight ANN-based models have been developed targeting key geotechnical parameters such as UCS, M_r, OMC, and MDD. The predictors used across these models are summarised in Table 2.8.

Mozumder and Laskar (2015) conducted a series of laboratory experiments to develop a dataset comprising 283 test records, with 198 allocated for training and 85 for testing their predictive model. The predictors used in this model are listed in Table 2.8. The model does not consider the effects of soil fractions or the chemical composition of the materials used. A shallow artificial neural network (ANN) architecture was adopted, consisting of a single hidden layer with nine nodes, selected through a trial-and-error process. The hyperbolic tangent function was used as the activation function for the hidden layer, while a pure linear function was used for

Table 2.8. Target parameters and predictors of ANN-based models

Model ID	Target & Predictors
ANN-Moz	Target: UCS Predictors: LL_0, PL_0, $soil\%$, $ash\%$, M, A_to_B, $Na\_to\_Al\_mix$, $Si\_to\_Al\_mix$.
ANN-Leo	Target: UCS Predictors: $soil\%$, $ash\%$, $water\%$, $SS\%$, $SH\%$, SS_to_SH, A_to_ash.
ANN-Han	Target: M_r Predictors: PI_0, MC_0, $clay\%$, $silt\%$, OMC_0, $cement\%$, $lime\%$, $ash\%$.
ANN-Nis-1	Target: OMC Predictors: ash_to_soil, OMC_0.
ANN-Nis-2	Target: MDD Predictors: ash_to_soil, MDD_0.
ANN-Nis-3	Target: UCS Predictors: ash_to_soil, t_curing.
ANN-Nis-4	Target: UCS Predictors: ash_to_soil, M, SS_to_SH, t_curing.

the output layer. Model training was performed using the Bayesian regularisation algorithm to improve generalisation and reduce over-fitting.

Nissanka et al. (2021) developed four shallow artificial neural network (ANN) models to predict various geotechnical parameters using published experimental data. The models include ANN-Nis-1 for predicting OMC, ANN-Nis-2 for MDD, ANN-Nis-3 for UCS of ash-stabilised soil, and ANN-Nis-4 for the UCS of soils stabilised with alkali-activated fly ash. The datasets used for training and testing these models comprised between 112 and 169 data test records each. A detailed list of predictors used in the models is provided in Table 2.8. In all four models, the number of hidden nodes was determined through a trial-and-error process, while the number of hidden layers was arbitrarily fixed at one. The hyperbolic tangent function was used as the activation function in the hidden layer, and a pure linear function was applied in the output layer.

As shown in Table 2.9, existing studies have predominantly employed shallow ANN architectures, typically consisting of a single hidden layer. These models have demonstrated reasonable predictive capabilities, however, they may be limited in their capacity to capture more complex, non-linear relationships in datasets involving chemical reactions. The performance of deeper ANN architectures, i.e., those incorporating multiple hidden layers, has not been explored in previous studies. Deeper networks have the potential to model higher order interactions between predictors and response. The absence of such exploration presents an opportunity to evaluate whether deeper ANN structures could enhance predictive accuracy in modelling soil stabilisation.

Table 2.9. ANN models' reported information on their architecture and hyperparameters selection and tuning

Model ID	Hyperparameters		Tuned
ANN-Moz	No. of input nodes:	8	
	No. of hidden layers:	1	
	No. of hidden nodes:	9	✓
	Activation functions:	Hyperbolic tangent and pure linear	
	Optimisation function:	Bayesian	
ANN-Leo	No. of input nodes:	7	
	No. of hidden layers:	1	
	No. of hidden nodes:	10	
	Activation functions:	Logistic	
	Optimisation function:	Not specified	
ANN-Han	No. of input nodes:	8	
	No. of hidden layers:	1	
	No. of hidden nodes:	9	
	Activation functions:	Log sigmoid	
	Optimisation function:	Levenberg-Marquardt	
ANN-Nis-1 & 2	No. of input nodes:	2	
	No. of hidden layers:	1	
	No. of hidden nodes:	4	✓
	Activation functions:	Hyperbolic tangent and pure linear	
	Optimisation function:	Not specified	
ANN-Nis-3	No. of input nodes:	2	
	No. of hidden layers:	1	
	No. of hidden nodes:	6	✓
	Activation functions:	Hyperbolic tangent and pure linear	
	Optimisation function:	Not specified	
ANN-Nis-4	No. of input nodes:	4	
	No. of hidden layers:	1	
	No. of hidden nodes:	4	✓
	Activation functions:	Hyperbolic tangent and pure linear	
	Optimisation function:	Not specified	

2.7.6 Tree-based Ensemble Methods

While most model developers have adopted ANN-based approaches, one study proposed a Random Forest model (RF-Zei) for predicting UCS (Zeini et al., 2023). Random Forest is an ensemble learning algorithm that aggregates the predictions of multiple decision trees to improve accuracy and reduce over-fitting. Tree-based ensemble methods have shown effective capability in dealing with engineering datasets that often suffer from non-uniform and irregular distributions. Predictors included PI_0, $slag\%$, $ash\%$, M, A_to_B, $Na\_to\_Al\_mix$, and $Si\_to\_Al\_mix$. Two key hyperparameters were optimised: the number of trees in the forest and the number of predictors considered at each split. Using the `GridSearchCV` function from the Scikit-learn library in Python, along with five-fold cross-validation, the optimal configuration was found to consist of 500 trees and five predictors.

2.7.7 Comparison of Self-Reported Accuracy of Models

A summary of the performance metrics for the models discussed earlier is presented in Table 2.10. While most studies report R^2 scores for their models, various other metrics have been used to express prediction errors. This comparison should be qualified by the fact that, with a few exceptions mentioned below, different studies have used datasets of varying sizes for regression. Larger datasets, which are typically drawn from multiple sources, introduce more variability in soil conditions, mix designs, and curing conditions. This can potentially result in lower R^2 scores when compared to models based on a more homogeneous dataset derived from a single experimental study or soil type. Therefore, when comparing two models based on different datasets, the higher R^2 scores (or lower error) of one model over another do not necessarily indicate its superiority, particularly if the dataset used for that model is smaller. Nonetheless, such comparisons can provide valuable insights into model performance.

This comparison takes into account both the dataset size and the number of experimental studies used in training. Due to the availability and consistency of performance metrics reported for UCS models, only UCS models have been included in this comparison.

In Figure 2.5 (a), a gently downward-sloping trend of R^2 scores is observed as dataset size increases. A group of lower-performing models, including the conventional regression model RM-Sar-3 and two ANN models, ANN-Nis-3 and ANN-Nis-4, appear as outliers. However, this trend becomes even more evident in Figure 2.5 (b), where R^2 decreases consistently as the number of studies increases from 1 to 20, with only ANN-Nis-3 and ANN-Nis-4 slightly deviated from the regression line.

Two distinct groups can be identified for an evaluation of the modelling techniques: (A) models trained on less than 500 test records of data and less than 10 experimental studies, and (B) models trained on greater 500 test records of data and greater than 10 experimental studies. Among the models in group A, the tree-based ensemble model RF-Zei demonstrates the best performance, while the shallow ANN model ANN-Nis-4 performs the weakest. For group B, the evolutionary model GEP-Pha-2 shows the best performance, while another shallow ANN model, ANN-Nis-3, performs the weakest. The conspicuously lower performance of models ANN-Nis-3 and ANN-Nis-4, compared to the rest of the models, is likely due to the absence of important input variables related to soil characteristics namely water content, and curing temperature in model training.

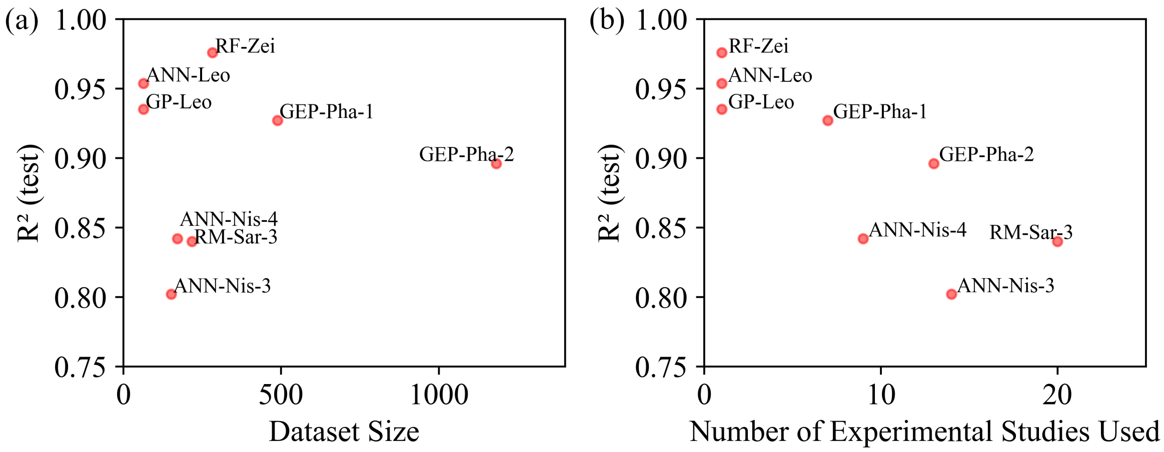


Figure 2.5. The number of experimental studies (a) and dataset size (b) plotted against R^2_{test} of the existing models

Two conclusions can be drawn from the evaluation of the above-mentioned modelling techniques: (i) machine learning has demonstrated better performance in modelling chemical soil stabilisation compared to conventional regression (Mozumder and Laskar, 2015), therefore, future models can focus mainly on machine learning approaches; (ii) ANN models have slightly outperformed evolutionary methods (Leong et al., 2018; Hanandeh et al., 2020). However, the greater transparency of evolutionary models may offset this slight advantage, especially in chemical soil stabilisation where understanding the effect of predictors on the predicted output is crucial. It is worth noting that in both studies where ANN marginally outperforms evolutionary methods, the size of the training data was small (i.e., < 130), making the comparison less relevant to scenarios where larger datasets are utilised.

Current knowledge on soil stabilisation modelling would benefit from a detailed evaluation and comparison of evolutionary methods, such as genetic programming, with (i) deep learning ANNs using a more comprehensive dataset covering multiple soil types and mix designs, and (ii) tree-based ensemble methods, which have demonstrated strong modelling capabilities, not only in the case of Zeini et al. (2023) but also in modelling challenging regression datasets (Zhuang et al., 2020).

Table 2.10. Performances of models

Model	Dataset Size	Target	Unit	Reported Performance Scores	Reported Errors
ANN-Moz	283	UCS	MPa	$r_{train} = 0.996$ $r_{test} = 0.982$	$MSE_{train} = 0.34$ $MSE_{test} = 1.50$
RM-Moz	283	UCS	MPa	$r_{train} = 0.910$ $r_{test} = 0.899$	$MSE_{train} = 7.24$ $MSE_{test} = 8.04$
ANN-Leo	64	UCS	MPa	$R^2_{test} = 0.9535$	Not reported
GP-Leo	64	UCS	MPa	$R^2_{train} = 0.9197$ $R^2_{test} = 0.935$	$RMSE_{train} = 2.01$ $MAE_{train} = 1.513$ $RMSE_{test} = 2.73$ $MAE_{test} = 1.754$
ANN-Han	125	M_r	Ksi	$R^2_{train} = 0.9741$ $R^2_{test} = 0.9532$	Not reported
GEP-Han	125	M_r	Ksi	$R^2_{test} = 0.9517$ $R^2_{test} = 0.9467$	Not reported
RM-Sar-1	180	LL	%	$R^2 = 0.92$	RME = 2.6
RM-Sar-2	180	PI	-	$R^2 = 0.83$	RME = 3.8
RM-Sar-3	218	UCS	kPa	$R^2 = 0.84$	RME = 6.5
RM-Sar-4	118	FSI	%	$R^2 = 0.97$	RME = 8.1
RM-Sar-5	118	MDD	kN/m^3	$R^2 = 0.87$	RME = 1.3
RM-Sar-6	118	OMC	%	$R^2 = 0.86$	RME = 1.8
ANN-Nis-1	112	OMC	%	$R^2_{train} = 0.973$ $R^2_{test} = 0.970$	$RMSE_{train} = 1.254$ $RMSE_{test} = 1.043$
ANN-Nis-2	112	MDD	g/cm^3	$R^2_{train} = 0.961$ $R^2_{test} = 0.975$	$RMSE_{train} = 0.047$ $RMSE_{test} = 0.381$
ANN-Nis-3	152	UCS	kPa	$R^2_{train} = 0.815$ $R^2_{test} = 0.802$	$RMSE_{train} = 80.578$ $RMSE_{test} = 90.896$
ANN-Nis-4	172	UCS	kPa	$R^2_{train} = 0.963$ $R^2_{test} = 0.842$	$RMSE_{train} = 474.136$ $RMSE_{test} = 231.908$
GEP-Pha-1	489	UCS	kPa	$R^2_{train} = 0.912$ $R^2_{test} = 0.927$	$RMSE_{train} = 156.4$ $RMSE_{test} = 204.6$ $MAE_{train} = 102.8$ $MAE_{test} = 118.4$
GEP-Pha-2	1183	UCS	kPa	$r_{train} = 0.952$ $r_{test} = 0.946$ $R^2_{train} = 0.907$ $R^2_{test} = 0.896$	$RMSE_{train} = 234.7$ $RMSE_{test} = 212.3$ $MAE_{train} = 170.2$ $MAE_{test} = 151.5$
RF-Zei	283	UCS	MPa	$R^2_{test} = 0.9757$	$RMSE_{test} = 0.9815$

2.8 Knowledge Gaps

The above literature review has identified several critical knowledge gaps, which are summarised next.

2.8.1 Strength Prediction

Several models have been proposed for predicting UCS of waste-stabilised soils but no review of their performance has been conducted to assess their consistencies, strengths, limitations, and the extent to which their predictor-response relationships align with empirical evidence.

In the context of strength development, a large body of experimental research has addressed many of the fundamental questions; however, critical research gaps remain in the area of predictive modelling. Most existing models have been developed using data derived from a single soil type or relatively small datasets compiled from published studies. While some recent models, such as GEP-Pha-1 and GEP-Pha-2 (Pham et al., 2021, 2022), have utilised larger datasets, they rely on cement and lime as stabilising agents rather than focusing on alkali-activated aluminosilicate wastes. Hence, a key limitation of the current literature is the under-utilisation of available experimental data and the absence of a predictive model trained on an expanded dataset made exclusively of waste-based stabilisation data.

The presence and concentration of key cementitious oxides, such as CaO, Al_2O_3, Fe_2O_3, and SiO_2, play a fundamental role in the geopolymerisation reactions described earlier. However, the existing predictive models have not explicitly considered these compounds as predictors (Consoli et al., 2016; Hanandeh et al., 2020; Leong et al., 2016; Mozumder and Laskar, 2015; Nissanka et al., 2021; Pham et al., 2021, 2022; Sari Ahmed et al., 2020; Zeini et al., 2023). Instead, models typically rely on the ASTM-based classification of ash into Class C and Class F, which broadly distinguishes ash based on high or low calcium content, respectively. Furthermore, sample preparation methods, which is another potentially influential factor affecting stabilisation, have largely been overlooked in model development (Dissanayake et al., 2017; Pham et al., 2021, 2022).

2.8.2 Swelling Stabilisation

Several critical aspects of swelling stabilisation using aluminosilicate-based binders remain insufficiently investigated. Further research is needed to identify the key factors that influence the effectiveness of swelling reduction, with particular attention to the required levels of alkaline activator concentration and curing time. Exploring practical strategies to reduce activator dosage is especially important, as it offers both cost and environmental benefits. This is supported by life cycle assessment studies, which highlight the significant impact of activator usage on the overall sustainability of alkali-activated systems (Habert et al., 2011).

To assess the range of its applicability, further investigation is required into the effectiveness and performance of alkali-activated systems across different soil types, particularly those with varying smectite content and fine particle fractions. In parallel with this, a better understanding of how geopolymer-based treatments influence the microstructural and chemical characteristics of stabilised soils is necessary. In addition, research is needed to evaluate the extent to which UCS is enhanced when the primary objective of stabilisation is to control swelling behaviour.

A better understanding of the durability of geopolymer-stabilised soils requires further investigation into their long-term chemical stability, particularly under varying environmental conditions such as cyclic wetting and drying. One important process influencing how durable gains from stabilisation are is the leachability of cementitious elements, such as calcium, aluminium, silicon, and magnesium, which play a central role in the time-dependant geopolymerisation process. Excessive leaching of these elements during early curing period can compromise the structural integrity and effectiveness of the stabilised soil.

Furthermore, monitoring changes in pH levels in the soils adjacent to the stabilised zone is essential for assessing potential impacts on nearby structure and sensitive natural ecosystems. High pH levels may adversely affect the surrounding ecosystem, groundwater quality, and even the long-term stability of nearby infrastructure. Despite its importance, the effect of cyclic wet–dry conditions on both the leaching behaviour and pH stability of geopolymer-treated soils remains largely unexplored. Addressing this gap is crucial for evaluating the long-term performance and environmental sustainability of alkali-activated soil stabilisation systems.

To the best of the authors' knowledge, no predictive models have yet been developed to estimate the swelling behaviour of expansive clays treated with geopolymer-based binders. This is unsurprising given the dearth of experimental studies on stabilisation for swelling. Nevertheless, the lack of models represents a significant gap in the current body of research.

2.8.3 Modelling Approaches

ANN models have demonstrated slightly better performance compared to evolutionary methods when using datasets with fewer than 130 test records collected from one study (Hanandeh et al., 2020; Leong et al., 2018). TbE's have also shown great capability in modelling complex regression datasets (Zeini et al., 2023; Zhuang et al., 2020). However, evolutionary methods, such as GP, offer clear advantages in terms of transparency, as they provide explicit mathematical expressions. This feature is particularly valuable in modelling chemical soil stabilisation that require interpretability. Given this advantage, further evaluation and comparison are necessary using datasets that adequately represent the complexities of chemical soil stabilisation accounting for greater variability in soil types, sample preparation methods, binder contents, and curing conditions (temperature and time). Also, existing comparative studies have predominantly focused on performance metrics such as R^2 or error rates, without sufficiently exploring the effectiveness of each approach in learning the predictor-response relationships, and their alignment with empirical findings.

Therefore, a systematic comparison of GP with (i) tree-based ensemble methods, and (ii) deep learning ANN, is needed. This evaluation should not only assess predictive performance but also consider the interpretability and consistency of the relationships between the modelled predictors and empirical evidence using reliable techniques such as SHapley Additive exPlanations analysis (Lundberg and Lee, 2017). Such an evaluation would provide valuable insights into the strengths and weaknesses of these three modelling approaches, helping to support more informed decision-making in the development of predictive tools for chemical soil stabilisation in the future.

2.8.4 Literature Gaps Mapped to Research Objectives & Questions

Table 2.11 relates the literature gaps identified above with the research questions investigated in this study and the three research objectives stated in Chapter 1. This alignment demonstrates how each aspect of the study is grounded in existing knowledge gaps and directed towards clearly defined research objectives.

Table 2.11. Mapping of research gaps to research questions and objectives

Literature Gaps	Research Questions *(as stated in Section 1.4)*	Research Objectives *(as stated in Section 1.3)*
1. Absence of a detailed review of existing UCS prediction models that critically evaluates their consistency, strengths, limitations, and alignment with empirical experimental findings.	**1.** How can insights from the capabilities and limitations of existing strength prediction models be used improve the development and design of reliable waste-based soil stabilisation systems?	ii.
2. Lack of understanding of the individual and interactive effects of experimental factors on swelling behaviour, including insufficient insight into the mechanisms governing swelling stabilisation under alkali activation. **3.** Lack of understanding of the durability and environmental impacts of stabilised soils, particularly relating to leaching of cementitious elements and elevated pH levels under wet–dry cycling.	**2.** What are the key physicochemical factors governing the swelling behaviour of alkali-activated aluminosilicate-stabilised soils, and how can new laboratory data address gaps in current experimental knowledge?	i.
4. Underutilisation of published waste-based UCS datasets in developing and assessing stabilisation models. **5.** Limited understanding of how cementitious chemical compound contents and sample preparation methods influence UCS when used as model predictors. **6.** Absence of a reliable and validated predictive model capable of estimating swelling stabilisation performance.	**3.** How can a unified and reproducible machine-learning framework be developed to predict the strength and swelling behaviour of soils stabilised with recycled aluminosilicates? **4.** How reliably can machine-learning models, trained on both literature-derived strength data and newly generated swelling data, predict stabilisation performance and identify the most influential mix-design parameters?	i. & iii.
7. Lack of detailed comparative evaluation of Genetic Programming (GP) against Artificial Neural Networks (ANN) and Tree-based Ensemble (TbE) methods for chemical soil stabilisation modelling.	**5.** How do most adopted methods such as Artificial Neural Networks and Genetic Programming compare with Tree-based Ensembles in their ability to model soil stabilisation using alkali-activated aluminosilicates?	ii.

CHAPTER 3

METHODS AND MATERIALS

The methods presented in this chapter are aligned with the objectives and research questions outlined in Section 1.4. This chapter covers a meta-analysis of the existing models, the experimental procedures used to investigate soil swelling, the statistical analyses conducted on the compiled UCS and swelling datasets, and the implementation and evaluation of machine learning techniques for developing new predictive models for strength and swelling, as illustrated in Figure 3.1.

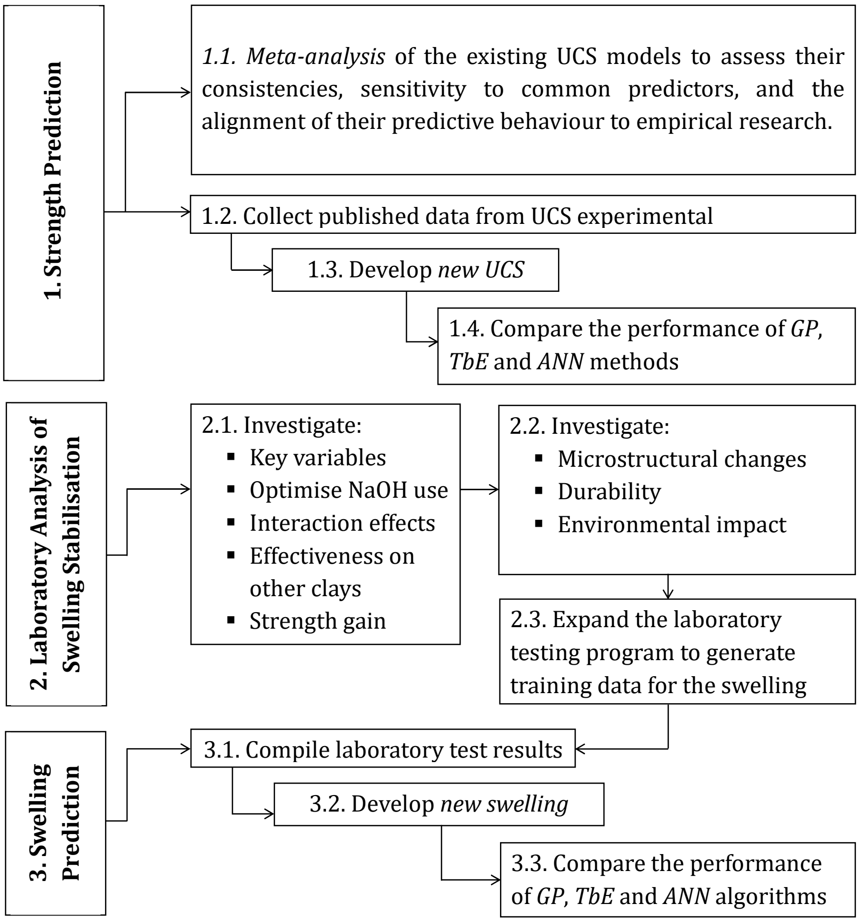


Figure 3.1. A mind map showing the key research components

3.1 Meta-analysis of Existing Models (Task 1.1)

This section outlines the literature search process used to identify UCS models for inclusion in the analysis, the criteria applied for model selection and classification, and the methodology for the subsequent meta-analysis. The meta-analysis consisted of three complementary components designed to evaluate model performance from different perspectives. Analysis 1 applied a global sensitivity analysis, which examines the influence of each predictor over its full range of possible values and accounts for interactions between predictors. This differs from local sensitivity analysis, which does not consider interaction effects among predictors. The global approach used here quantifies the overall contribution of each predictor within each model and allows meaningful comparison of predictor importance across models with similar structures. Analysis 2 was a *qualitative parametric analysis* comparing predictor–response relationships for common variables among similar models. In other words, Analysis 1 focussed on the relative importance of different predictors, while Analysis 2 was more concerned with the nature and direction of the relationship between a given predictor and UCS. Analysis 3 was an *empirical consistency assessment* examining the alignment of predictor–UCS trends with experimental evidence. Together, these analyses provide a structured framework for assessing the consistency, interpretability, and empirical validity of the selected models. Model classification was needed as analyses 1 and 2 require models to have close similarities in their capabilities.

3.1.1 Literature Search For Predictive Models

Literature was searched to identify existing UCS models of soils stabilised with aluminosilicate-based binders. The search was performed across the ScienceDirect, Scopus, and Google Scholar databases, using combinations of the following keywords: *“soil stabilisation” OR “soil stabilization”* AND *"unconfined compressive strength"* AND *“model” OR “prediction” OR “machine learning” OR “regression”* AND *“ash” OR “slag” OR “aluminosilicate” OR “alkali activation” OR “alkali activated”*. The search was restricted to studies published in peer-reviewed journals and conference proceedings in English up to 2023. Reference lists of relevant papers were also reviewed to identify additional models not captured in the initial database search. A list of previous models has already been presented in Table 2.4.

Models were included in following meta-analysis if they: (i) predicted UCS for soils stabilised with aluminosilicate wastes such as fly ash or slag; and (ii) provided sufficient methodological detail to enable assessment of their predictor variables and applicability.

3.1.2 Model Selection and Classification

Model selection was undertaken in three stages. First, models were identified from the literature through a systematic search for potential inclusion in the meta-analysis. These models have been presented in Table 2.4. Second, the models were classified according to their scope of applicability, particularly in relation to the role of alkaline activators and curing conditions, as discussed in subsequent paragraphs below. Finally, using the classification method developed in the second stage, suitable models were selected for the meta-analysis. As a result, some models identified in the initial stage were later excluded.

Empirical findings have shown that soil stabilisation via alkali-activated aluminosilicate wastes such as slag and ash is heavily influenced by the presence of an alkali activator as demonstrated by Abdila et al. (2022). Alkaline activators increase the rate of dissolution of alumina and silica in source materials, which is an important factor in geopolymerisation reactions. The process of geopolymerisation, that is needed for soil stabilisation, may not even begin in low-calcium mixtures in the absence of an alkaline activator. Although the presence of an alkali activator may not impact the ultimate soil strength, which depends on the chemical composition of the mixture, particularly silica and alumina, they do play a crucial role in the early strength development, which can be a benefit for projects with time constraints.

On the other hand, for shallow stabilisation using mechanical compaction, some applications may not require activators. In such cases, models such as RM-Con presented in Equation A.2 (Appendix A) can be considered, provided that the necessary conditions recommended by Consoli et al. (2016) are met. Experimental studies have shown positive results for similar approaches (Karami et al., 2021; Renjith et al., 2021; Gireesh Kumar and Harika, 2021). Therefore, model classification can begin by first grouping the existing models into two broad groups based on their consideration of alkaline activators. Some existing models do not incorporate the effect of alkaline activators (Consoli et al., 2016; Hanandeh et al., 2020; Pham et al., 2021, 2022) and they cannot be placed in the same group as the ones that do (Mozumder and Laskar, 2015; Leong et al., 2016; Zeini et al., 2023).

The literature also highlights the importance of considering curing temperature and curing time (t_curing) in the selection of appropriate models for soil stabilisation (Provis and Bernal, 2014). Consequently, models can be further categorised based on the ranges of T_curing and t_curing in the experimental dataset used for training/regressing the model.

The resulting categorisation yielded five distinct model classes denoted as MC1 to MC5 and presented in Table 3.1 and Figure 3.2. This classification system was used to select suitable models for this analysis.

Table 3.1. Model categories and their applicability

Model Class	Applicability
MC1	Suitable for scenarios where alkali activators and high temperatures are used. These models are not suitable for in-situ soil stabilisation that relies on in-situ curing conditions.
MC2	These models require alkali activators and are suitable for soil stabilisation that relies on ambient curing conditions. They are likely to be more accurate in estimating short-term strength as the experimental data used in their development had a maximum *t_curing* of 28 days.
MC3	These models do not require alkali activators. The experimental dataset used for their development includes a wide range of *t_curing* under ambient *T_curing*.
MC4	Similar to MC3, but with *t_curing* restricted to 28 days or less, and the model is hence likely to capture only short-term strength.
MC5	The model relies on *lime*% and compaction to achieve the required strength, which makes it most suitable for shallow stabilisation. The model does not require activators and does not heavily depend on *t_curing* and *T_curing*.

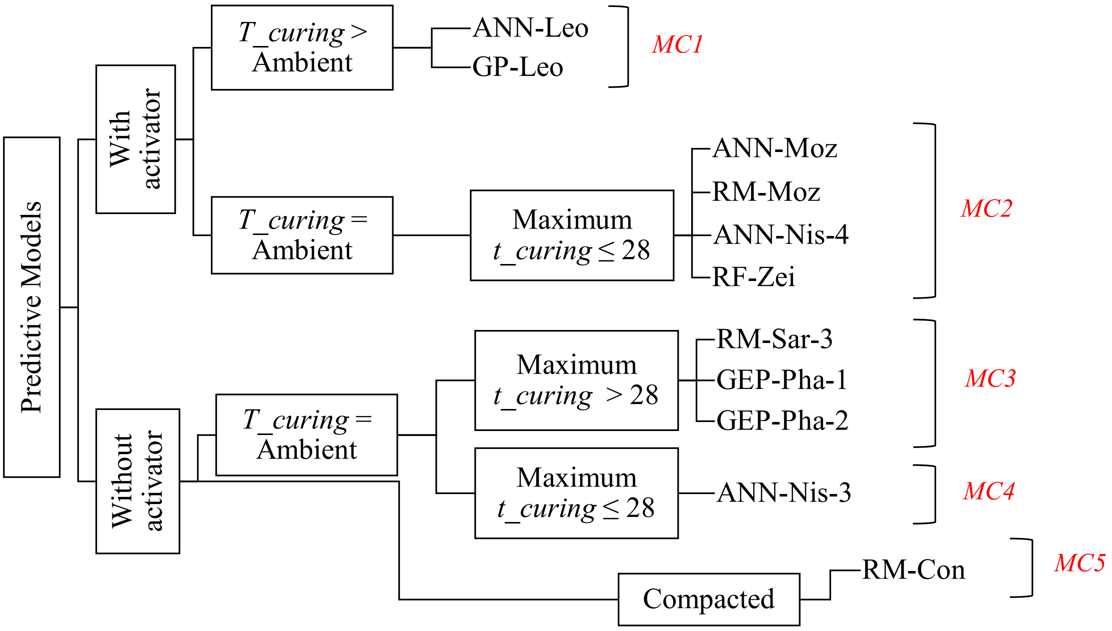


Figure 3.2. Classification of the existing models

3.1.3 Meta-analysis of Model Behaviour

The **global sensitivity analysis** aimed to (a) measure how strongly are UCS predictions influenced by predictors, and (b) determine whether similar models assign comparable importance to common predictors. This required considering models with similar sets of predictors. For example, models relying on the activation of binders with an alkaline solution could not be meaningfully compared with those that do not. Based on the UCS model classification above, the MC3 models, which include models RM-Sar-3, GEP-Pha-1, and GEP-Pha-2, share similar predictors and capabilities and were therefore selected for this analysis.

The analysis involved conducting a global sensitivity analysis based on the Sobol method (Iwanaga et al., 2022; Herman and Usher, 2017). Sobol is a variance-based method used to quantify the amount of output variance due to variance in individual predictors (Zhang et al., 2015). Total sensitivity analyses were carried out to capture sensitivity due to both individual predictors and their interactions with each other. Variance-based sensitivity analysis methods are heavily influenced by the variability space, i.e., the ranges of values of predictors, considered for input parameters. Two distinct ranges of predictor values were hence used: i) a range of values that extend beyond the maximum and minimum values in the training or development dataset, and ii) the range of values within the training data itself. These ranges are hereafter referred to as the "analysis range" and the "training range", respectively, and are presented in Table 3.2.

Table 3.2. Predictor fixed values, the range of predictor values used in this analysis, and the range of predictor values in their development/training datasets

Model ID	**Predictors**	**Units**	**Fixed Values**	**Analysis Range**	**Training Range**
RM-Sar-3	UCS_0	kPa	100	-	-
	$ash\%$	%	25	0 to 60	0 to 50
	t_curing	Days	28	0 to 365	0 to 90
GEP-Pha-1	PL_0	%	50	-	-
	LL_0	%	20	-	-
	$clay\%$	%	30	0 to 100	15 to 70
	$silt\%$	%	40	0 to 100	10 to 74
	$sand\%$	%	30	0 to 100	3 to 75
	$organic\%$	%	1.86	-	-
	$water\%$	%	100	0 to 300	22 to 265
	t_curing	Days	280	0 to 365	18 to 271
	$cement\%$	%	0	-	-
	$ash\%$	%	20	0 to 60	0 to 30
	$CaO\_ash\%$	%	12.82	-	-
	$Ca\_to\_Si\_ash\%$	%	0.38	-	-
	$LOI\_ash\%$	%	7.22	-	-
GEP-Pha-2	PI_0	-	30	-	-
	$clay\%$	%	30	0 to 100	5 to 71
	$silt\%$	%	40	0 to 100	8 to 74
	$sand\%$	%	30	0 to 100	1 to 75
	$water\%$	%	100	0 to 300	14 to 265
	t_curing	Days	280	0 to 365	3 to 360
	$lime\%$	%	0	-	-
	$cement\%$	%	0	-	-
	$ash\_F\%$	%	20	0 to 60	0 to 34
	$ash\_C\%$	%	20	0 to 60	0 to 30
	$slag\%$	%	0	-	-

The **qualitative parametric analysis** aimed to *qualitatively* compare how similar models respond to the common predictors. This analysis employed the same models as those used in Analysis 1. It focuses on identifying whether the direction (positive or negative) and shape (linear, non-linear) of the predictor–UCS relationships are consistent across models. This allows the detection of similarities that suggest agreement in modelling approaches and differences that may indicate inconsistencies or limitations in the way certain variables are represented. The parametric analyses were conducted in two steps. First, for each selected model, a set of base case scenarios were built, consisting of a set of reasonable input values. Common predictors were assigned the same values across the base case scenarios for the three models. Second, a given individual predictor value was varied while all other predictors were kept constant. Although this approach may not capture the complex multi-dimensional behaviour of models, it is easy to implement and interpret. Table 3.2 presents the values of the base case scenarios under the column titled 'Fixed Values'. The base case values contain '0' values for predictors that are not shared between the models e.g., *cement*% and *lime*%. The results were also used to identify the extent to which input-output relationships generated by the models agree with empirical evidence.

The **empirical consistency assessment** involved conducting a *parametric comparison* on a broader set of models to evaluate the extent to which the qualitative relationships between inputs and outputs align with empirical evidence. This did not require models to be similar, therefore, five models were used: RM-Con, RM-Sar-3, RM-Moz, GEP-Pha-1, and GEP-Pha-2. This comparison helps to determine which models reflect known physical and chemical behaviours in stabilised soils and which produce results that deviate from observed trends, potentially pointing to data limitations, model bias, or overfitting. In this analysis, the mean predictor values in the training data for each model were used as the fixed values for each model. Table 3.3 presents the fixed values (Fixed Values), the range of predictor values in this analysis (Analysis Range), and the range of predictor values in training data (Training Range). Similar to analyses 1 and 2, both analysis and training ranges were used.

Table 3.3. Predictor fixed values, the range of predictor values used in this analysis, and the range of predictor values in their development / training datasets

Model ID	Predictors	Fixed Values	Analysis Range	Training Range
RM-Con	*ash%*	10	0 to 30	0 to 25
	lime%	5	0 to 15	3 to 9
	DUW_mix	15	10 to 20	14 to 17
	DUW_lime	15	15 to 20	-
RM-Sar-3	*UCS_0*	100	0 to 2000	39 to 1514.9
	ash%	25	0 to 50	0 to 50
	t_curing 28	28	0 to 360	0 to 90
RM-Moz	*LL_0*	63.78	0 to 200	37.68 to 116.27
	PI_0	38.84	0 to 100	14.07 to 88.46
	slag%	15.83	0 to 50	0 to 50
	ash%	2.24	0 to 30	0 to 20
	M	12.45	0 to 20	4 to 14.5
	A_to_B	0.62	0 to 1	0.45 to 0.85
	Na_to_Al_mix	1.16	0 to 3	0.24 to 1.98
	Si_to_Al_mix	1.71	1 to 3	1.49 to 2.49
GEP-Pha-1	*PL_0*	33.4	0 to 50	1 to 53
	LL_0	88.65	0 to 200	31 to 119
	clay%	37.30	0 to 100	15 to 70
	silt%	35.60	0 to 100	10 to 74
	sand%	27.10	0 to 100	3 to 75
	organic%	1.86	0 to 30	0 to 27
	water%	135.99	0 to 300	22 to 265
	t_curing	25.70	0 to 365	18 to 271
	cement%	12.04	0 to 30	0 to 36
	ash%	14.37	0 to 30	0 to 30
	CaO_ash%	12.82	0 to 50	1 to 48
	Ca_to_Si_ash%	0.38	0 to 3	0 to 2.46
	LOI_ash	7.22	0 to 50	0 to 49
GEP-Pha-2	*PI_0*	38.17	0 to 100	5 to 78
	clay_0	30.92	0 to 100	5 to 71
	silt%	38.68	0 to 100	8 to 74
	sand%	30.40	0 to 100	1 to 75
	water%	93.79	0 to 300	14 to 265
	t_curing	35.84	0 to 365	3 to 360
	lime%	1.89	0 to 20	0 to 20
	cement%	8.27	0 to 100	0 to 100
	ash_F%	3.68	0 to 34	0 to 34
	ash_C%	0.94	0 to 34	0 to 30
	slag%	2.87	0 to 42	0 to 42

3.1.4 Reconstructing Published Models

The UCS models included in the meta-analysis were reproduced using reported information in their respective publications. For regression models, Where explicit regression expressions were available (e.g., RM-Con, RM-Sar-3, RM-Moz), the published formulations were directly used in this analysis. For GEP models, e.g., GEP-Pha-1, GEP-Pha-2, the symbolic expressions provided in the respective studies were reconstructed for use. In cases where data scaling or preprocessing of training data was explicitly reported, these steps were also implemented to ensure consistency with the original model development.

Due to the limited information available on the preprocessing and scaling techniques applied to the training datasets of ANN models and model GP-Leo, they were excluded from the analysis. Although the excluded models could be reconstructed in principle using published information on their architecture, their predictions would likely be unreliable without the required information on the preprocessing steps of their training datasets. This raises a larger issue in the literature, concerning reproducibility of reported results.

3.2 Data Collection (Tasks 1.2 and 3.1)

For the UCS models, published experimental data were compiled, whereas the swelling model was developed using the author's own experimental data. Details of these datasets and results of their preprocessing and refinement are provided in their respective chapters: Chapter 5 for UCS and Chapter 7 for swelling.

3.3 Model Development (Tasks 1.3 and 3.2)

The process of building the models was conducted in three key stages (see Figure 3.3). The first stage involved data collection and preprocessing. The second and third stages can be represented by two loops: i) an iterative process to check for and address covariance and multicollinearity, ii) predictor selection and model training and evaluation while continuously monitoring and mitigating training issues such as over-fitting and under-fitting. Python was used for data processing and refinement, modelling, and analyses.

The development of the UCS and swelling models required two separate datasets. A total of 1598 test records of UCS test data were collected from 23 published experimental studies to prepare the UCS dataset. Over 260 1-dimensional free swelling (1DFS) tests were conducted as part of this research and their results were compiled to prepare the swelling dataset. Data preparation steps and the statistical descriptions of the two datasets are presented in their respective chapters.

3.3.1 Assessing Covariance and Multicollinearity

The training datasets were thoroughly checked for anomalies, duplicates, and incomplete test records, all of which were removed prior to further analysis. Binder contents were standardised as percentages of dry soil mass, and the test records were randomised to prepare the dataset for analysis. The datasets were analysed and refined to reduce covariance and multicollinearity among variables, which can adversely impact model performance (Smith, 1993). This process involved using

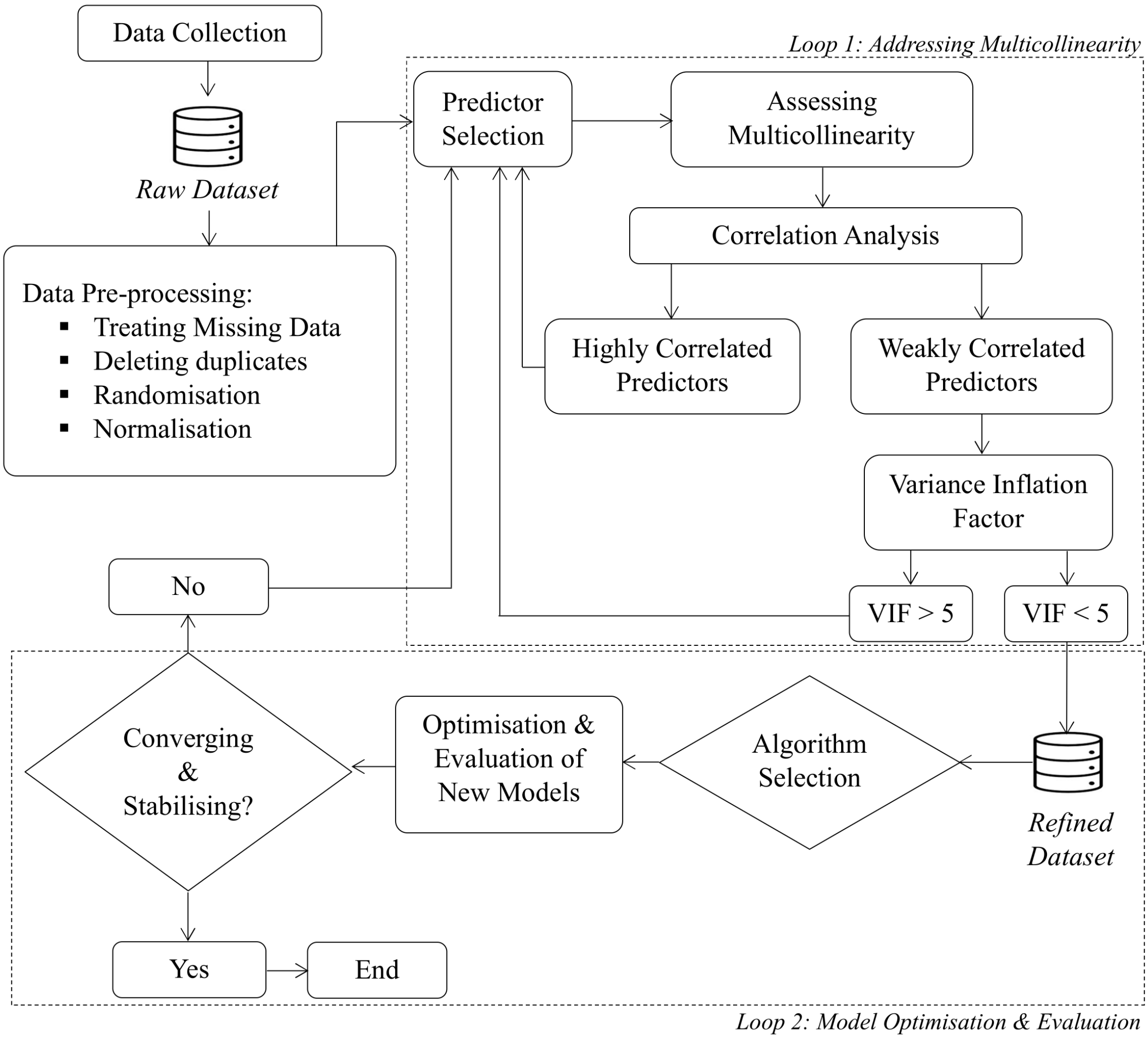


Figure 3.3. Schematic diagram illustrating the process of data preparation, pre-processing, refinement, analysis, and machine learning implementation

highly correlated variables to create new, less correlated predictors. Pearson correlation (r_s) was used to assess the relationships between predictors, as shown by Equation 3.1.

$$r_{P_i,P_j} = \frac{\sum_{k=1}^{n} \left(P_{i,k} - \bar{P}_i\right)\left(P_{j,k} - \bar{P}_j\right)}{\sqrt{\sum_{k=1}^{n} \left(P_{i,k} - \bar{P}_i\right)^2 \sum_{k=1}^{n} \left(P_{j,k} - \bar{P}_j\right)^2}} \tag{3.1}$$

where $P_{i,k}$ is the value of the i-th predictor for the k-th observation, $P_{j,k}$ is the value of the j-th predictor for the same observation, n is the total number of observations, and $\bar{P}_i$ is the mean of predictor P_i across all observations.

A key limitation of Pearson correlation is that it only assesses pairwise correlations and does not consider the effects of other predictors, i.e., multicollinearity. Multicollinearity was assessed by evaluating the combined influence of all predictors on a given predictor P_i using the Variance Inflation Factor (VIF), defined by the following formula:

$$\mathrm{VIF}_i = \frac{1}{1 - R_i^2}, \tag{3.2}$$

where R_i^2 represents the coefficient of determination, indicating how well P_i can be estimated using other predictors.

The ideal scenario, though difficult to achieve, is when R^2 equals 0, resulting in a VIF value of 1. A good dataset typically has VIF values below 10, indicating minimal to moderate multicollinearity (Akinwande et al., 2015). In this study, we used a threshold of 5 to further reduce the effect of multicollinearity.

3.3.2 Genetic Programming

Genetic Programming (GP), a subset of evolutionary algorithms, generates predictive models by evolving expression trees composed of mathematical functions (e.g., $-$, $+$, $\times$, $\sqrt{\cdot}$, $|\cdot|$, log, sin, cos), predictors, and constant values (Koza, 1992). This study has utilised a symbolic regression–based GP approach, implemented using the open-source `gplearn` Python library. The process begins by generating a population of random expression trees, forming a diverse set of potential solutions. It then iteratively improves this population using genetic operations inspired by evolution and natural selection. The hyperparameter `p_crossover` controls the crossover operation, where segments of two parent trees are exchanged to produce offspring trees. This promotes the combination of potentially effective substructures from different solutions. Similarly, the hyperparameter `p_mutation` governs mutation operations, which introduce variation by modifying parts of a tree, such as replacing an operator or a variable, to explore new regions of the solution space.

The evolutionary process retains high-performing expressions across generations, while poorly performing ones are discarded or assigned lower reproduction probabilities. Selection is often guided by a tournament-based strategy controlled by the `tournament_size` parameter. This iterative refinement leads to increasingly accurate and compact expressions, where less significant predictors are naturally excluded. The growth of expression trees is penalised using a `parsimony_coefficient`, which discourages overly complex formulas, thereby reducing overfitting and improving interpretability. The final optimised expression tree, as illustrated in Figure 3.4, can be expressed as a mathematical formula, with intermediate nodes representing operations and leaf nodes representing predictors or constants. This structural transparency makes GP particularly well-suited for applications in soil stabilisation, where understanding the influence of predictors is crucial for practical implementation.

A full memetic approach, which typically combines evolutionary search with subsequent local optimisation of numerical constant values, was not implemented in this study. Rather, constant values were integrated as leaf nodes and evolved directly within the expression trees alongside the other predictors. Hyperparameters

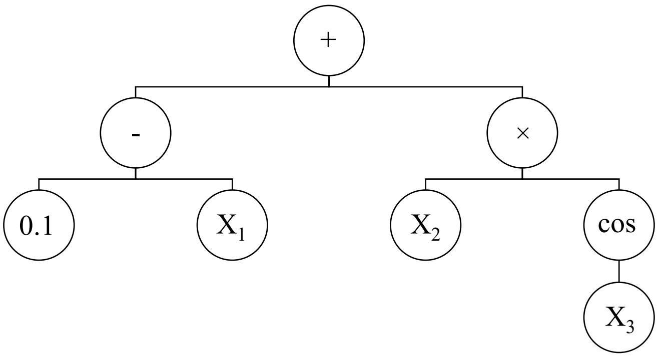


Figure 3.4. A simple example of an expression tree showing how predictors X_1 to X_3 interact to influence the output. The corresponding mathematical expression is $\hat{y} = [0.1 - X_1] + X_2 \cdot \cos(X_3)$.

were optimised in a separate pre-training phase rather than being tuned during the evolutionary process.

3.3.3 Tree-based Ensemble Methods

Tree-based ensemble (TbE) methods combine multiple models to create more accurate predictions by leveraging diversity to reduce overfitting and improve generalisation. These methods can be classified into bagging and boosting. Bagging involves training multiple models on bootstrapped subsets of data and aggregating their predictions. Examples include Random Forest (RF), which builds decision trees with random feature splits, Extra Trees (ET), which adds extra randomness for faster training and reduced variance, and Bagging Regressor (BR), a general framework for bagging any base estimator (Géron, 2019; Hastie et al., 2009). The final output of any bagging method is calculated by taking the average of the combined output of all trees as shown in Equation 3.3.

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} y_i(x), \tag{3.3}$$

where $\hat{y}$ is the predicted output, n is the number of decision trees in the ensemble, and $y_i(x)$ is the prediction of the i-th tree for an input vector x.

Boosting algorithms build 'weaker' models sequentially, with each model correcting errors of the previous ones. Popular boosting methods include Gradient Boosting (GB), which optimises a loss function by fitting residuals, and AdaBoost (AB), which adjusts instance weights to focus on harder-to-predict data Hastie et al. (2009). The final output of the boosting algorithms is calculated by taking the weighted sum of the trees as shown in Equation 3.4:

$$\hat{y} = \sum_{i=1}^{n} \alpha_i y_i(x), \tag{3.4}$$

where α_i is the learning rate (or weight) of the i-th tree.

The key difference between bagging and boosting methods is how they handle training. In bagging, multiple decision trees are trained independently on different bootstrap samples (random samples with replacement), while in boosting, trees are trained sequentially, with each model attempting to correct the errors of the previous one.

In this study, the performance of RF, ET, BR, GB, and AB in modelling chemical soil stabilisation data was evaluated by optimising key hyperparameters: the number of estimators in the forest (`n_estimators`), the maximum number of predictors to consider when finding the best split (`max_features`), the maximum depth of each estimator (`max_depth`), the minimum number of test records required to split a node (`min_samples_split`), the minimum number of test records needed at the leaf nodes (`min_samples_leaf`), the proportion of the data used to train each individual estimator (subsample), the weight of each estimator's contribution to the final prediction (`learning_rate`), the minimum reduction in impurity, i.e., entropy or uncertainty, required to make a split (`min_impurity_decrease`), and whether estimators are trained with sampling done with replacement (bootstrap) (Géron, 2019). Other hyperparameters not mentioned above were left at their default values. It is worth noting that while most of the hyperparameters mentioned above are relevant to all five algorithms, some (such as `learning_rate` and `bootstrap`) only apply to specific algorithms.

3.3.4 Artificial Neural Network

Artificial Neural Networks (ANN) with multiple hidden layers are classified as deep learning models (Xu et al., 2021) and have been widely applied in geotechnical engineering contexts (Onyelowe et al., 2023). Compared to most algorithms, ANN models are generally more complex and resource-intensive to train due to a larger number of optimisable hyperparameters. The number of hyperparameters increases as the architecture of the model expands. This requires greater computational effort and careful tuning to achieve optimal performance. Key trainable hyperparameters include the activation function, optimiser, loss function, learning rate, batch size, number of epochs, and the network's structural configuration such as the number of hidden layers and nodes per layer (Géron, 2019).

Figure 3.5 illustrates a typical architecture of an ANN used in deep learning, comprising an input layer, multiple hidden layers, and an output layer. Each neuron within the network performs two key operations: *aggregation*, where it computes a weighted sum of its inputs and adds a bias term; and *activation*, where a non-linear function, e.g., Rectified Linear Unit (ReLU) (Géron, 2019), is applied to the aggregated input to introduce model flexibility and enable learning of complex non-linear patterns. For instance, the hidden neuron X_i1 receives inputs from all nodes in the previous layer and processes them through these two steps. This transformation is repeated across all layers of the network, passing information forward until a final prediction $\hat{y}$ is produced at the output layer.

During training, $\hat{y}$ is compared with the actual target using a loss function, e.g., Mean Squared Error. The resulting error is propagated backward through the network using the backpropagation algorithm. Gradients of the loss with respect to the weights and biases are calculated and used to update the parameters via optimisation algorithms such as Stochastic Gradient Descent or Adam (Géron, 2019). This iterative training process continues until the model achieves satisfactory performance.

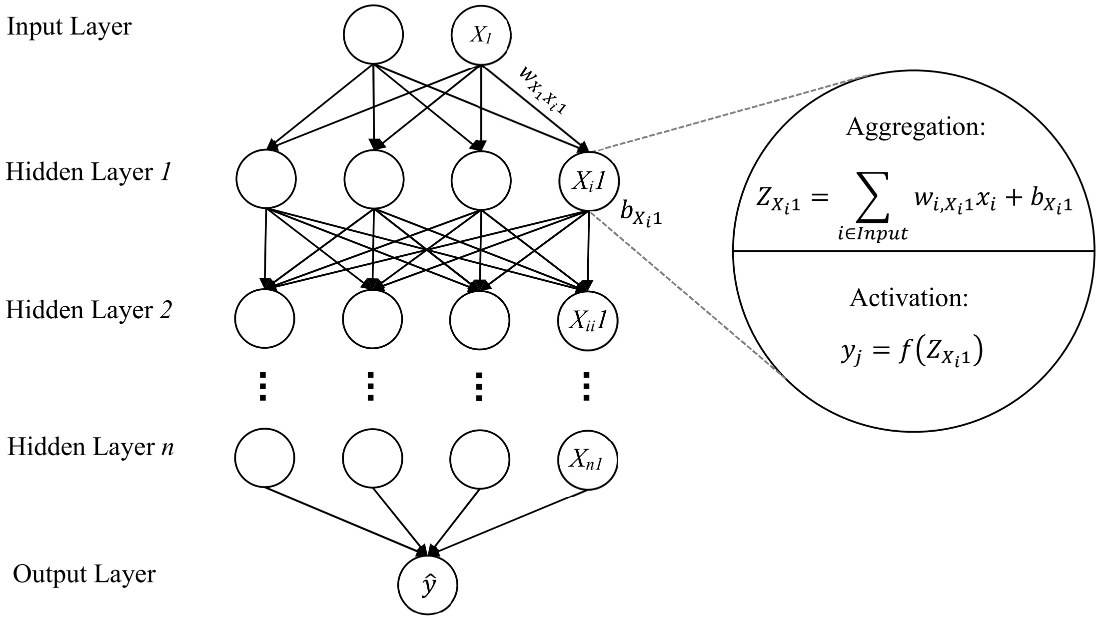


Figure 3.5. A typical architecture of an ANN-based model

3.3.5 Optimisation Method

Overfitting and underfitting are common challenges in machine learning that affect a model's ability to generalise to unseen data. Overfitting occurs when a model becomes overly complex and learns not only the underlying patterns in the training data but also noise and irrelevant details. This results in considerably higher performance on the training set than the unseen testing set. In contrast, underfitting happens when a model is too simple or lacks the capacity to capture the patterns in the data, leading to consistently poor performance on both the training and testing sets. Model optimisation aims to establish a balance between model complexity and generalisability to ensure that the best possible performance is achieved without overfitting.

Optimisation of hyperparameters was carried out using a two-stage approach. First, a random search was performed over a broad hyperparameter space to identify promising regions of hyperparameter values. These regions were then refined using a combination of automated grid search and manual refinement. Both random and grid search procedures employed cross-validation to ensure that the selected hyperparameters generalised well across different subsets of the data, thereby reducing the risk of overfitting.

Cross-validation (CV) is widely used and recognised as a fundamental statistical method in machine learning for model validation and generalisation (Géron, 2019). It has been effectively applied in previous studies to assess predictor influence (Cozzolino et al., 2023), evaluate model performance (Zhang et al., 2021), and guide model training (Tiwari and Satyam, 2021). The main objective of CV in this study was to facilitate unbiased model evaluation during training. Figure 3.6 illustrates how CV was implemented in hyperparameter optimisation in this study. Cross-validation involved the following steps:

i. Dataset was split into 80% and 20% for training and testing, respectively.
ii. The training set was further divided into 10 folds (the optimal number of folds for CV (Kohavi, 1995)). The model was trained on 9 folds and validated on the remaining fold. This was repeated 10 times (Rep 1 to 10), allowing each fold to be used as a validation set once. Specified hyperparameter values were then searched to find the best combination based on the CV score, calculated as the average R^2 score across all folds as shown Equation 3.5:

$$\text{CV Score} = \frac{1}{n}\sum_{i=1}^{n} R_i^2, \tag{3.5}$$

where n is the number of CV folds, ${R_i}^2$ is the R^2 score of the model trained on a given fold.
iii. The best-performing hyperparameter values were further refined through grid search and manual tuning.
iv. The entire training set was then passed through the model to update its parameters.
v. The final performance was evaluated on the unseen test set.

The benefits of CV are highlighted in step 2, where the training process reduces the dependency on a single train-test split commonly used, which can lead to biased and misleading results if the split is not representative of the underlying patterns in the data. Model performance was evaluated using CV scores and metrics from both training and testing phases. Overfitting and underfitting were monitored through CV results, scatter plots, and metrics including R^2, residual errors, mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).

The CV score reflects average model generalisability and training adequacy. Comparing it to the training score (R^2_{train}) helps detect overfitting and underfitting. A strong indicator for overfitting is when R^2_{train} is significantly higher than the CV

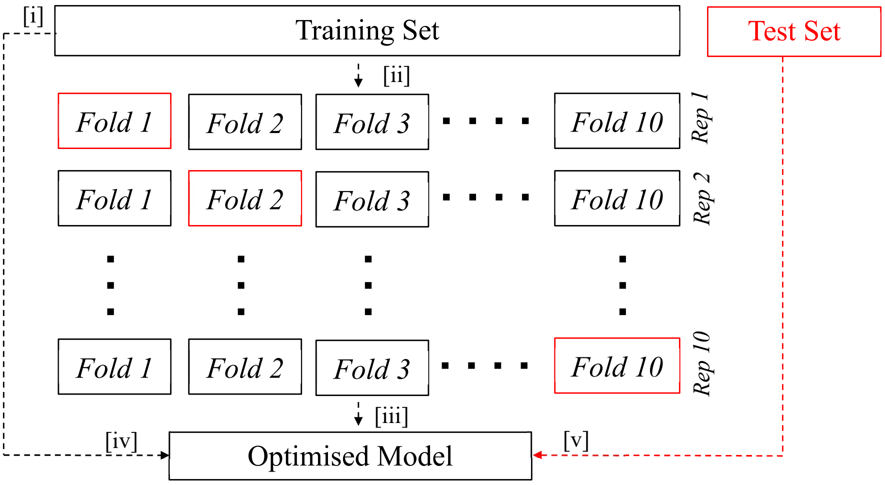


Figure 3.6. Illustration of how K-fold cross-validation was used in hyper-parameter optimisation

and test scores. The opposite is true for underfitting, when R^2_{train} is significantly lower than the CV and test scores. Ideally, closely aligned training, testing, and CV scores indicate a balanced model with minimal bias and variance.

A visual assessment of the effectiveness of hyperparameter optimisation is through the scatter plots. Scatter plots of predicted versus actual values visually represent model accuracy for training and testing sets. Ideally, points align along a 45-degree line, indicate accurate predictions, while significant deviations indicate potential overfitting or underfitting.

3.3.6 Model Interpretation via SHapley Additive exPlanations Analysis

To better understand the relative importance and effects of individual predictors on model predictions, SHapley Additive exPlanations (SHAP) analysis was conducted following the methodology proposed by Lundberg and Lee (2017).

In a given multivariate model, SHAP values $S_1, S_2, \ldots, S_n$ are computed for predictors $P_1, P_2, \ldots, P_n$ to quantify each variable's contribution to the predicted output. The SHAP value for a predictor P_i represents the marginal contribution of that variable to the model output, expressed as the change in the prediction from the model's average prediction $\bar{y}$, due to a change in P_i. $\bar{y}$ is calculated by taking the average of model predictions across all test records in the datasets. It acts as a baseline for SHAP analysis. The predicted value $\hat{y}$ for any test record in the data can be expressed as the sum of the model's mean prediction and the SHAP values of all predictors, as shown in Equation 3.6.

$$\hat{y} = \bar{y} + \sum_{i=1}^{n} S_i \tag{3.6}$$

SHAP summary plots were used to evaluate the overall influence of each predictor on UCS predictions. Positive SHAP values indicated that the corresponding predictor increased the predicted UCS, while negative values implied a reduction. Additionally, SHAP dependency plots were used to visualise the effect of each predictor across its range of values, providing insights into non-linear effects, detecting potential anomalies, and enhancing the interpretability of the model.

3.3.7 Assessing Model Uncertainty

Data issues such as skewness and irregular distributions may lead to learning difficulties and increase model uncertainty (Zhuang et al., 2020). To mitigate data-induced uncertainty and evaluate the impact of data variability on model performance, min-max normalisation and bootstrap resampling, a common tool for uncertainty analysis (Efron, 1979), were applied. Min-max normalisation, as presented by Equation 3.7, was used to scale all predictor values to a standard range of 0 to 1, ensuring that variables with large values do not disproportionately influence the model. This scaling balanced the contribution of all predictors, reduced the risk of bias, and improved the learning process.

$$X_{\mathrm{norm}} = \frac{X - X_{\mathrm{min}}}{X_{\mathrm{max}} - X_{\mathrm{min}}}, \tag{3.7}$$

where *X_norm* is the normalised predictor value for a given predictor, X is the original predictor value, *X_min* and *X_max* are the minimum and maximum predictor values.

Bootstrap resampling helps assess the reliability and stability of a model by repeatedly training it on different versions of the dataset and analysing how its performance changes. The following steps were carried out to study model uncertainty using bootstrap resampling technique:

i. The training dataset was randomly sampled with replacement to generate 20 bootstrap sets. Sampling the data with replacement allowed for some test records to be represented more than once while some may not be sampled at all enabling a better assessment of model performance when a slightly different version of the dataset is used to tune its hyperparameters.
ii. Models were trained on each bootstrap sample to assess their sensitivity to data variations.
iii. Models were evaluated using out-of-bag (OOB) samples, i.e., test records excluded during resampling, to estimate generalisation to unseen data. R^2 scores for training and testing phases were calculated.
iv. Finally, the mean and 95% confidence interval of these scores were calculated. Narrower intervals indicated greater stability and reliability, whereas wider intervals reflected increased uncertainty.

The top two best-performing models were selected based on their performance metrics, including bootstrap resampling results, to be further evaluated using Wilcoxon signed-rank test (Wilcoxon, 1945). The test was conducted on the paired test R^2 scores to assess whether the performance differences between the models were statistically significant or simply due to random variation.

While a larger number of bootstrap iterations would increase the effectiveness of the Wilcoxon signed-rank test, we limited our analysis to 20 resamples to balance detailed statistical analysis with the significant computational time required, particularly for deep Artificial Neural Networks and Genetic Programming. Similar number of iterations (observations) have been used for Wilcoxon tests in other studies (Turton et al., 2023; D'Alonzo and Cipriani, 2012).

3.4 Experimental Analysis of Swelling Stabilisation (Task 2.1 and 2.2)

3.4.1 Materials

The base soil used in this experiment was a residual high plasticity Silty CLAY (CH) recovered from a site in Western Sydney. Organic matter, rootlets, and ironstone gravels were removed from the soil before it was oven-dried and ground. Once in powder form, it was passed through a 425 µm sieve to prepare the base soil. Figure 3.7 shows photos of the base soil, ash, slag, and NaOH pellets used in the experiments. X-ray Diffraction (XRD) tests presented in Table 3.4 show that the base soil consisted primarily of quarts, kaolinite, montmorillonite, and illite minerals. The particle size distribution of the base soil is presented in Figure 3.8. Table 3.6 summarises the chemical compounds and their concentrations in the materials utilised in this study.

Figure 3.7. Materials used in testing (a) base soil, (b) ash, (c) slag, and (d) NaOH pellets used in this study

Table 3.4. Clay mineral composition and concentrations determined through XRD analysis

Minerals	**Concentrations (wt%)**
Quartz	65
Kaolinite	15.2
Montmorillonite	11.9
Illite	8

To study the effect of stabilisation on clays with different swelling potentials, the base soil was amended by adding known percentages of either fine sand (particle

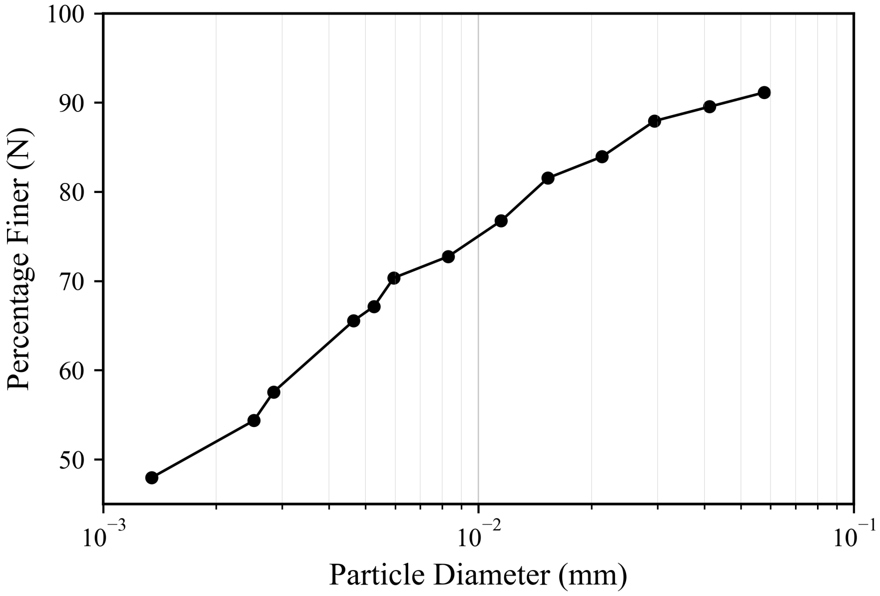


Figure 3.8. Particle size distribution of the base soil

sizes < 425 µm) or sodium bentonite soil (predominantly smectite) to prepare four additional soil types (amended_1 to amended_4). The sand and bentonite blends used to modify the base soil are detailed in Table 3.5. The plasticity index (PI) serves as an preliminary and approximate indicator of soil swelling behaviour, with higher PI usually indicating higher swelling (Patel, 2019). As shown in Figure 3.9, the PI values of the five soils used in this study range from 29.4 to 57.1, approximately evenly distributed, spanning from the upper low plasticity (CL) to the upper high plasticity (CH) region on the Casagrande chart, covering a range usually prone to swelling and shrinkage behaviour.

Table 3.5. Properties of the untreated soils including base soil, bentonite, and amended soils, i.e., soils amended with bentonite

Properties	**Base Soil**	**Bentonite**	**Amended 1**	**Amended 2**	**Amended 3**	**Amended 4**
Added sand (%)	0	0	5	0	0	0
Added bentonite (%)	0	100	0	5	10	15
Liquid Limit (%)	51	566	47	55	63	75
Plastic Limit (%)	16	40	18	18	18	18
Plasticity Index	35	526	29.36	36	45	57
USCS Classification	CH	CH	CL	CH	CH	CH
Swelling (%)	26	-	24	40	49	53

Atterberg limits tests were conducted in accordance with Australian Standard AS 1289 (Standards Australia, 2000). For each test, approximately 200 g of dry soil, sieved through a 425 µm mesh, was used. Distilled water was gradually added until the soil reached a plastic consistency. The moistened samples were then sealed in airtight bags and allowed to equilibrate for 24 hours to ensure uniform moisture distribution prior to testing.

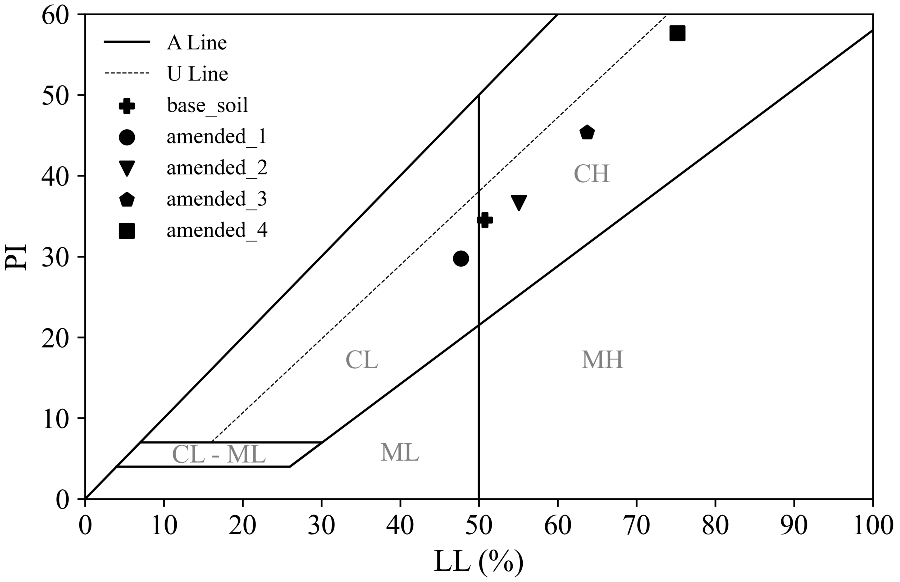


Figure 3.9. Locations of the five soils used in this experiment on the Casagrande plasticity chart

Binders included low-calcium ash supplied by a local coal power plant based in New South Wales, Australia. The slag was purchased from a local supplier. The alkaline activator liquid was prepared by dissolving the required mass of dry NaOH pellets in distilled water. NaOH pellets had a purity of 98%. The solution was stirred multiple times until all pellets were fully dissolved. The solution was left for 24 hours before use. The chemical compositions of the base soil, bentonite, ash, and slag were analysed through semiquantitative x-ray fluorescence (XRF) analyses. Figure 3.10 show the intensities versus 2-theta degrees for the detected elements showing high intensity peaks, indicating high concentrations, for silicon and aluminium in the base soil and ash, and high peak for calcium and relatively lower intensities for silicon and aluminium for slag. Table 3.6 presents the detected elements in their most common oxide forms and their contents.

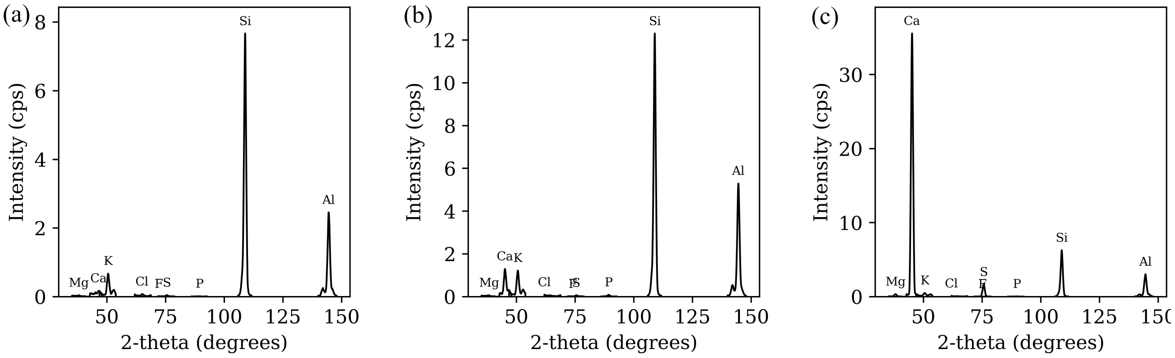


Figure 3.10. Elemental composition of the (a) base soil, (b) ash, and (c) slag based on XRF analysis

Table 3.6. Chemical composition and concentrations (% by mass) for base soil, bentonite, ash, and slag determined through XRF analysis

Compounds	**Base soil**	**Bentonite**[1]	**Ash**	**Slag**
Al_2O_3	17.82	13.84	23.30	13.33
SiO_2	64.61	52.35	62.70	31.38
CaO	0.09	1.58	1.31	38.68
Fe_2O_3	3.60	17.39	3.95	0.33
Na_2O	0.29	0.36	0.47	0.13
MgO	0.77	1.86	0.78	5.85
P_2O_5	0.04	0	0.31	0.06
SO_3	0.06	0	0.05	1.74
K_2O	1.28	0.16	1.47	0.32
TiO_2	0.78	1.35	0.98	0.53
MnO	0.01	0	0.06	0.23

[1] The compound concentrations for bentonite come from Taheri (2023) as the same bentonite was used

3.4.2 Sample Preparation and Curing

The dry ingredients of the mix, namely base soil, sand, bentonite, ash, and slag, were first mixed in a dry state. It was essential to ensure that the dry mix was thoroughly homogenised before any liquid was added. Once the dry mix was ready, the required amount of liquid (either distilled water or activator solution) was added to the mix, with continuous stirring using a flexible spatula to ensure a uniform consistency. The prepared mixture was then transferred into cylindrical PVC moulds with an internal diameter of 38 mm and a height of 50 mm. To reduce potential friction at the soil-wall interface, silicon grease was applied to the interior walls of the moulds. Paper filters were placed above and below the sample within the mould to prevent disturbance at the top and bottom of the sample during compaction.

The sample was compacted in the mould using a loading frame, with a static load of 177 N applied for 60 seconds. The dry densities of the prepared samples ranged from 0.0013 to 0.0014 g/mm^3. After unloading, the mould containing the sample was wrapped in cling wrap to maintain moisture during the curing process. Figure 3.11 shows prepared mixtures (a) and samples after compaction (b). The prepared samples were placed in sealed buckets and cured in a temperature-controlled environment at a constant temperature of 21°C.

In this study, binder contents are expressed as a percentage by dry soil mass. Liquid contents are expressed as a percentage of the total dry solids mass, i.e., the combined mass of soil, ash, and slag. The specific mixture designs were developed in alignment with the objectives of the testing program and are presented in Section 3.4.3.

Figure 3.11. (a) Sample mixture after addition of the alkaline activator and (b) prepared specimens prior to wrapping in cling wrap and curing

3.4.3 Testing Program

To address the research questions for swelling stabilisation, four types of tests were conducted, namely 1DFS under monotonic and cyclic wet-dry conditions, MIP and SEM imaging, UCS tests, and infiltration tests under cyclic wet-dry conditions to assess leachability. As discussed in Chapter 2, 1DFS method is among several other tests that can be used to assess soil swelling. 1DFS was selected in this study because, given its relative simplicity and low cost in resources and time, it allowed a large number of tests to be conducted and hence a wider range of variables to be explored. The tests provided sufficient data to address the research questions and generate training data for predictive modelling.

In this thesis, *untreated samples* refer to soil samples mixed with distilled water only (i.e., no added binder or activator and hence samples are not activated or stabilised); *unstabilised samples* are binder-containing soil samples mixed with distilled water (i.e., no added NaOH); and *stabilised samples* are binder-containing soil samples activated using a NaOH solution.

The 1DFS tests (under monotonic wetting) were carried out in four successive sets. Set 1 tests aimed to establish a baseline and confirm the role of activator in the stabilisation, especially given its relatively large contribution to the overall cost of the process. Swell tests were conducted on: (a) untreated base soil, (b) unstabilised base soil containing 5% ash, 5% slag and 20% distilled water, and (c) stabilised base soil with the same binder contents as (b) but distilled water replaced with 3M NaOH solution. Curing time was 30 days for all samples.

Set 2 aimed at making a preliminary assessment of the most important factors in stabilisation by evaluating the effects of large changes, between upper and lower limits, of four key variables on the swelling of stabilised samples of base soil. The variables in question are ash%, slag% and alkaline liquid contents (ALC%) as well as molarity of NaOH solution (M). Table 3.7 show the lower and upper values of these variables; given the paucity of literature on stabilisation for swelling, these values were selected based on a review of the literature on stabilisation for UCS. Table 3.8 shows details of set 2 tests.

Building on the outcomes of set 2, set 3 involved further tests on the base soil to evaluate the effects of a larger combination of variables and identify any interaction effects that were not captured in set 2. Finally, while tests in sets 1 to 3 were conducted on the base soil, in set 4, swell tests were carried out on the four amended soils to assess the ability of aluminosilicate stabilisation to reduce swelling of soils with variable plasticity and initial swell potential.

Based on the findings of 1DFS tests carried out under monotonic wetting conditions, MIP, SEM, and UCS tests were carried out. Rigid wall infiltration and additional 1DFS tests were also conducted under cyclic wet-dry conditions. The testing program is summarised in Table 3.9. Table 3.10 presents the types of samples analysed for each test.

Table 3.7. Upper (U) and lower (L) limit values considered for each variable

Variables	**Abbreviation**	**L (%)**	**U (%)**	**L & U Range in UCS Studies**
Ash content	ash%	5	40	0 to 50 (e.g., Pham et al. (2022); Odeh and Al-Rkaby (2022); Cristelo et al. (2011))
Slag content	slag%	5	30	0 to 50 (e.g., Abdullah et al. (2019); Mozumder and Laskar (2015); Singhi et al. (2016))
Alkaline liquid content	ALC%	20	60	6.35 to 147 (e.g., Singhi et al. (2016); Leong et al. (2018); Suksiripattanapong et al. (2021))
NaOH Molarity	M	3 M	14 M	3 to 15 (e.g., Singhi et al. (2016); Dissanayake et al. (2017))

Table 3.8. Set 2 free swell tests: two-factor specimen design, where one variable content was varied per specimen

ID	**ash%**	**slag%**	**ALC%**	**M**	**Comments**
baseline	L	L	L	L	Baseline - using the lower limits for all variables
ash%	U	L	L	L	Effect of ash by using its upper limit content
slag%	L	U	L	L	Effect of slag by using its upper limit content
ALC%	L	L	U	L	Effect of ALC% by using its upper limit content
M	L	L	L	U	Effect of NaOH molarity by using its upper limit concentration

Table 3.9. Summary of tests conducted in this study

Research Questions	Tests	Aims
1	Monotonic 1DFS – Set 1	To confirm the role of alkaline activator in stabilisation
1	Monotonic 1DFS – Set 2	To identify key variables affecting stabilisation
1	Monotonic 1DFS – Set 3	To identify effects of interaction between variables
2	Monotonic 1DFS – Set 4	To assess the ability to reduce swelling of various soils
3	MIP	To quantify changes in porous structure
3	SEM	To visualise changes in microstructure
4	UCS	To assess UCS gains
5	1DFS tests under wet-dry cycles	To assess the swelling behaviour of stabilised soils
5	pH tests under wet-dry cycles	To assess the change in pH of effluent samples
5	Infiltration tests under wet-dry cycles	To assess the leachability of cementitious elements

Table 3.10. The types of samples prepared and analysed for each test

Analysed Samples	Binder	Activator	Set 1 1DFS	Set 2 1DFS	Set 3 1DFS	Set 4 1DFS	MIP & SEM	UCS	W&D Cycles
Untreated samples	×	×	✓			✓		✓	✓
Unstabilised samples	✓	×	✓				✓		
Stabilised samples	✓	✓	✓	✓	✓	✓	✓	✓	✓

3.4.4 1DFS Test

The methodology for the 1DFS test followed ASTM Standard D4546-21, Method B, which is commonly used to estimate the one-dimensional swelling of expansive clays induced by wetting (ASTM International, Committee D18, 2021). Cured samples were unwrapped, and the mould height and initial sample thickness H_i were recorded. Moulds were then placed in a desiccator with water raised to about 10 mm above the base of the moulds, allowing it to soak the sample from the bottom. Weights were placed on top of the moulds during soaking to prevent them from being pushed up during swelling. A digital calliper was used to measure the sample thicknesses. The final thicknesses (H_f) were measured once swelling reached its maximum value. Swelling (swelling%) was calculated using Equation 3.8.

$$\text{swelling\%} = \frac{(H_f - H_i)}{H_i} \times 100, \tag{3.8}$$

where H_f is the maximum thickness.

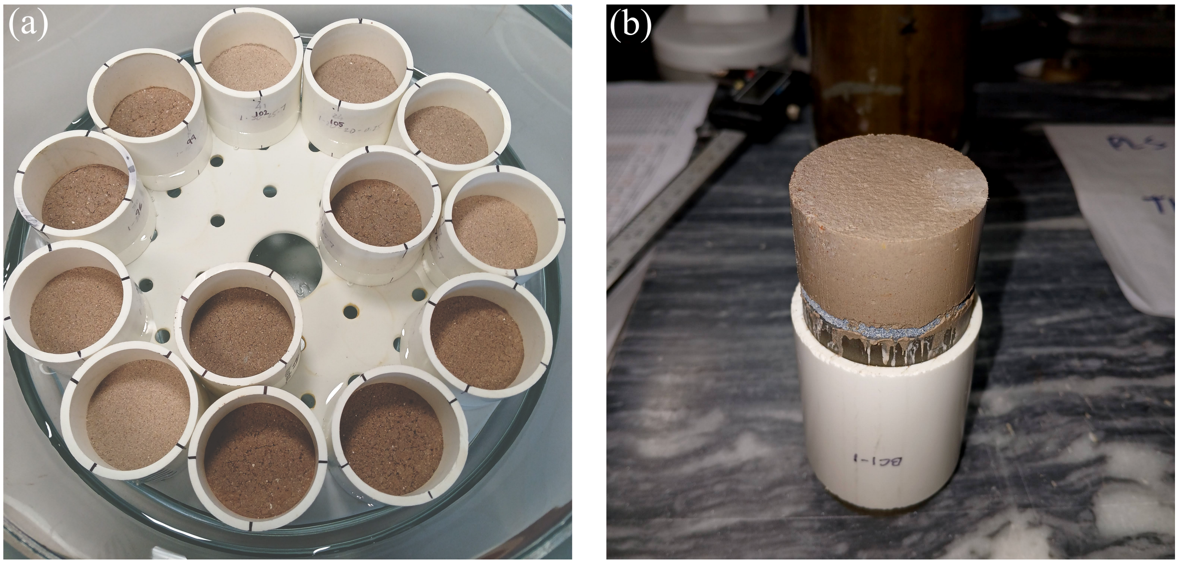


Figure 3.12. (a) Samples undergoing swelling inside a desiccator and (b) a sample extruded from its mould after completing a test

3.4.5 MIP and SEM Analyses

The base soil was amended with 20% ash and 10% slag and then mixed with 25% ALC to prepare two duplicate stabilised samples (Stabilised_A and Stabilised_B). The same proportions were mixed with 25% distilled water to prepare two duplicate unstabilised samples (Unstabilised_A and Unstabilised_B). Distilled water was used to ensure no activation occurs in the unstabilised samples. Ash and slag were included in the unstabilised samples to maintain consistent dry mix constituents, ensuring that any observed microstructural changes were solely attributable to stabilisation. Samples were prepared in accordance with the above sample preparation method and cured for 7 days under 21ºC. After curing, samples were cut into smaller pieces. MIP samples were approximately 15 mm × 5 mm × 3 mm in size. After cutting, samples were immediately frozen using liquid nitrogen then placed in a freeze-dryer for 48 hours at -57ºC and a vacuum of 0.1 mBar to ensure samples are completely dry before analysis. MIP analysis was conducted to quantify the pore size distribution and investigate changes in pore structure due to stabilisation. The AutoPore IV 9500 manufactured by Micromeritics was used for this analysis. The test procedure using this equipment is semi-automatic and follows the standard ASTM D4284-12 (ASTM International (D32 Committee), 2017).

SEM imaging was carried out to examine the microstructural changes induced by stabilisation, including the morphology, particle arrangement, and pore structure of the samples. SEM images were captured using the Phenom XL G2 Desktop SEM, supplied by Thermo Fisher Scientific. They allowed morphological evolution of the samples to be examined and provided a visual reference against which MIP results could be assessed.

3.4.6 UCS Tests

As mentioned earlier, although UCS is not the primary focus of this part of the research, knowledge of UCS gains as a result of stabilisation for swelling is useful, especially since optimal conditions for the latter may not be so for the former. UCS tests were performed in accordance with ASTM Standard D2166-16 (ASTM International, 2024). Sample preparation followed the same method as described in section 3.2 above. Mixtures were poured and compacted in layers in 38 mm diameter moulds to achieve similar dry densities as the swelling samples. Sample height to diameter ratio was 2:1. Two batches of samples were prepared for analysis.

For the first batch, compacted samples were extruded from their moulds, wrapped in clingwrap, then placed in zip-lock bags, and stored for curing under the same curing conditions as the swelling samples. Fourteen samples were prepared as part of the first batch to study the effects of initial dry density (initDD, i.e., dry density of the mix before stabilisation), ash%, slag%, ALC%, M, and t_curing on strength.

The second batch of samples aimed to investigate the strength of samples after exposure to water. Samples were prepared as above except that they were not extruded from their moulds. Sample moulds were placed in zip-lock bags for curing. After 7 days of curing, sample moulds were placed in a glass container with distilled water rising to about 50 mm from the base of the moulds. Distilled water was allowed to infiltrate the samples from the bottom for a period of ten days before being extruded from their moulds and tested. Weights were placed on top of moulds during soaking to stop them from being pushed up due to swelling. Figure 3.13 presents an image of a sample after being prepared and compacted (a) and another image showing a sample after completing a UCS test (b).

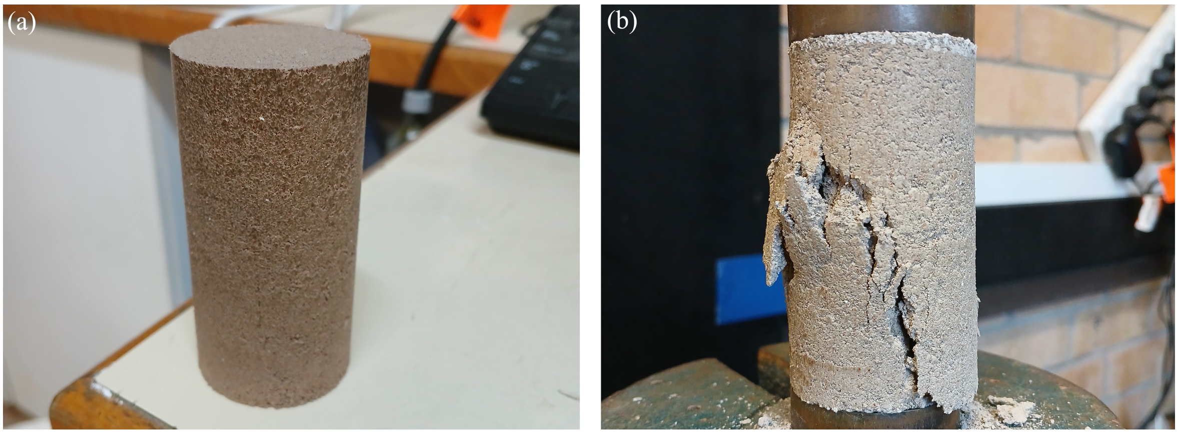


Figure 3.13. (a) UCS sample after being compacted and (b) UCS sample after completing its test

3.4.7 Infiltration and 1DFS Tests Under Wet-Dry Cycles

Infiltration tests were conducted with a rigid wall permeameter, to assess leachability of stabilised samples and compare it with that of the untreated base soil. For each test, the volume of effluent collected per sample container was equivalent to one pore volume. Pore volumes (PV) of stabilised and unstabilised samples were measured by using consolidation apparatus. Samples were prepared and cured for 7 days. After curing, the initial heights (H_i) were measured to calculate the total volume (V). Distilled water, approximately twice the liquid limit of the base soil, was then mixed with samples to create a slurry consistency. The slurry was consolidated in an 38 mm diameter odometer mould under gradually increasing vertical loads until H_i was reached. Once the sample was consolidated to H_i height, the saturated mass was recorded, followed by oven drying to obtain the dry mass. The mass of pore water was determined as the difference between the saturated and dry mass. Finally, PV was calculated by dividing the mass of pore water by the density of water (998 kg/m^3). While this method differs from the fundamental approach, which requires additional tests for each sample to determine their specific gravity values, it offered a time-efficient and consistent means the estimation of PV across samples within the scope and constraints of this study.

The infiltration samples were prepared in accordance with the above sample preparation method. Prepared samples were placed inside an empty mould containing a porous disk and a paper filter at the bottom. After compaction, another paper filter was placed on top of the sample. The paper filters acted as a barrier to retain soil particles while allowing water to pass through, whereas the porous quartz disks provided a stable, permeable support for the sample, facilitating uniform water infiltration and outflow.

After 7 days of curing, samples were air dried for 24 hours before being oven dried for an additional 48 hours under a temperature of 50ºC. After drying, the dry thicknesses and masses were measured before commencing the wetting cycle. A total of 10 PV's of effluent was collected per wetting cycle, which ranged from 28 to 33 days. The time taken to complete three wetting and drying cycles was about 5 months in total. 1DFS tests were conducted using the same method described above, except that water was allowed to soak the sample from the top rather than from the bottom.

Chemical analyses included measuring pH levelas and conducting XRF tests to determine chemical compositions and their concentrations in effluent samples. The XRF tests were performed using the Rigaku Supermini 200.

Figure 3.14 (a) illustrates the apparatus showing how the sample rig was connected to the pressure-volume controller. Due to the low permeability of the untreated based soil, a pressure of 5 kPa was applied from the top to facilitate the flow of distilled water through the sample. In contrast, the stabilised samples with relatively higher permeability did not require any pressure and a slightly different apparatus was hence used for effluent collection from the stabilised soils. As shown in Figure 3.14 (b), chromatography columns were used to control the flow of water into the sample mould in order to keep the standing water level in sample moulds at approximately 10 mm above the sample. The effluent collected during these tests was subsequently analysed to assess pH levels and chemical composition.

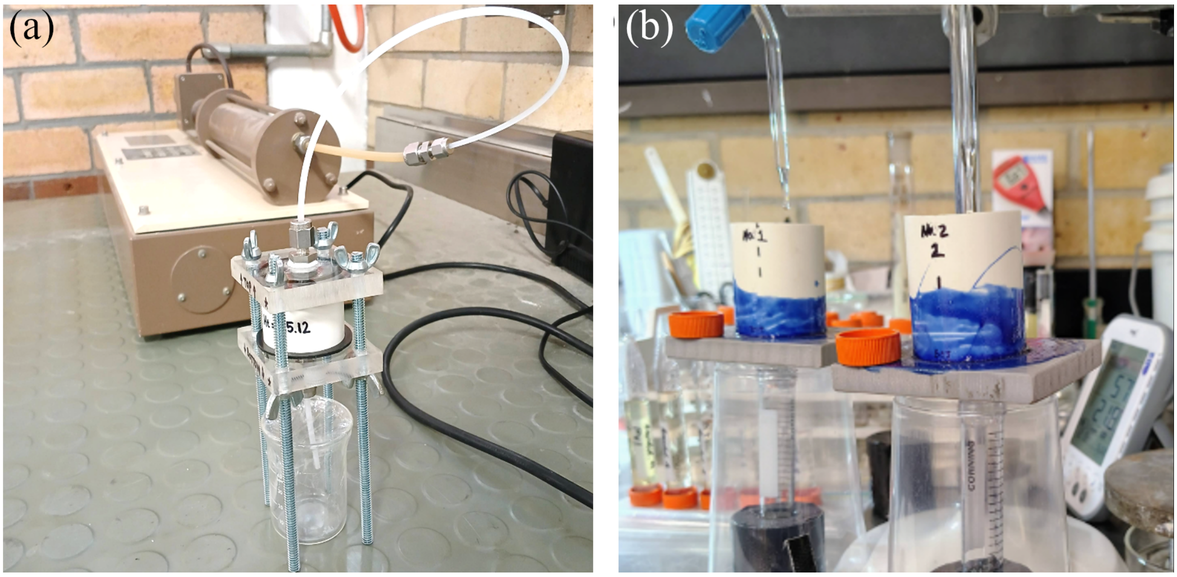


Figure 3.14. (a) Constant head permeability apparatus used for estimating K_{sat} and effluent collection for the untreated base soil, and (b) the testing apparatus for stabilised samples using chromatography column with effluent samples collected in centrifuge tubes at the bottom of the set up

3.4.8 Repeatability

To assess repeatability, all 1DFS samples in Sets 1 and 2 of the experiments were duplicated. For Sets 3 and 4, one out of every four samples were duplicated. The unstabilised samples in Set 4 were each tested in duplicate. Two duplicate stabilised samples were prepared for microstructural (MIP and SEM). Multiple specimens were extracted from each sample for SEM and MIP analysis to ensure representativeness and repeatability of the results. One out of every four UCS samples was duplicated, and for infiltration tests, the stabilised samples were duplicated.

CHAPTER 4

META-ANALYSIS OF EXISTING STRENGTH PREDICTION MODELS

This chapter presents the results of a meta-analysis comprising three complementary analyses. As illustrated in Figure 4.1, the first analysis is a *global sensitivity analysis* that quantifies the relative influence of individual predictors within each model and compares similar models. The second is a *qualitative parametric analysis* that compares predictor–UCS relationships for common variables across similar models. The third is an *empirical consistency assessment* that evaluates the extent to which a broader set of models aligns with established experimental evidence. Together, these analyses provide a comprehensive evaluation of model performance, highlighting both strengths and limitations in their predictive capabilities.

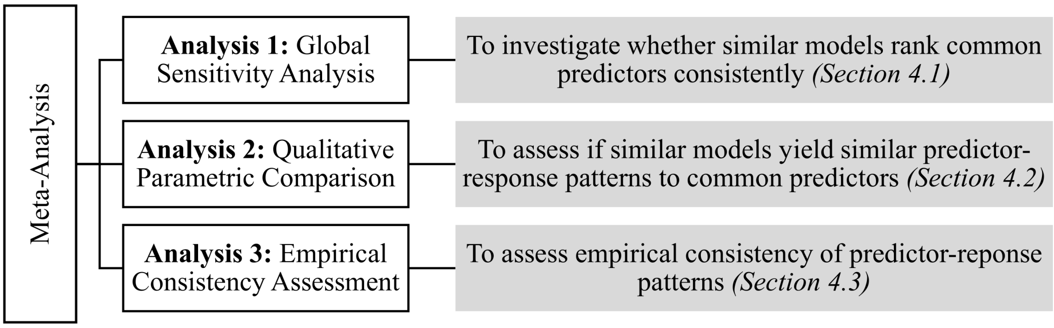


Figure 4.1. Overview of the structure and key components of Chapter 4

4.1 Analysis 1: Global Sensitivity Analysis of Predictor Influence

The results of the sensitivity analysis conducted on the RM-Sar-3 model are presented in Figure 4.2. Results indicate that the ranking order of predictors based on their importance remains the same for both training and analysis ranges. It is clear that the predictor variable *ash*% dominates in both cases, with total sensitivity indices of 70% for the training range and 77% for the analysis range. Similarly, UCS_0 exhibits reasonably high sensitivity, with sensitivity indices of approximately 46% for the training range and 42% for the analysis range. On the other hand, t_curing displays the lowest sensitivity, accounting for only 2% in the training range and 1% in the analysis range.

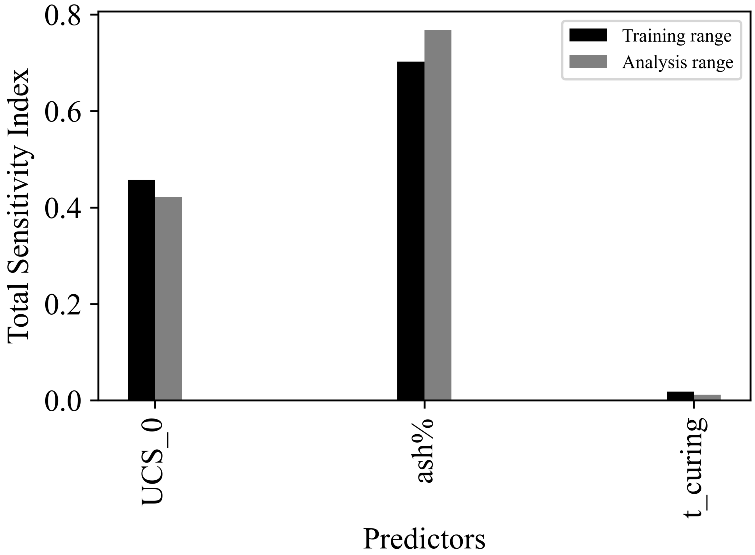


Figure 4.2. Predictor importance for the model RM-Sar-3

Model GEP-Pha-1 comprises a larger number of predictors than RM-Sar-3, including $cement\%$, which is a dominant cementitious binder in soil stabilisation. Therefore, in this case, the interactions between inputs are more complex. Results are shown in Figure 4.3. Considering the results obtained from the training range of values, the model exhibits much lower sensitivity to $ash\%$, compared to RM-Sar-3. This is likely due to the significant impact of $cement\%$, which emerges as the most important predictor in the analysis contributing to over 40% of the output variability in the model. The model demonstrates high sensitivity to PL_0 and $organic\%$ and, to a lesser extent, t_curing and $LOI\_ash\%$. $CaO\_ash\%$, $sand\%$, $silt\%$, LL_0, and $water\%$ appear to be significantly less influential in this model.

Applying the analysis ranges of predictor values changes the ranking order of predictors. The predictor $sand\%$ (Figure 4.3), which was among the least important predictors (in the training range of values) emerges as the most important predictor in the analysis range even though the increase in its variability space is not significant. The predictor $cement\%$ is the second most important predictor gaining a total sensitivity index of about 18%, which is less than half that of $sand\%$. The model exhibits considerable sensitivity to PL_0, LL_0, $organic\%$, t_curing, and LOI_ash. Predictors including $clay\%$, $silt\%$, $water\%$, $ash\%$, $CaO - ash\%$, and $Ca\_to\_Si\_ash$ retain their positions as the least important predictors, as observed in both the training and analysis ranges.

Model GEP-Pha-2 introduces new additives, such as $lime\%$ and $slag\%$, and differentiates between class C (ash_C) and class F (ash_F) fly ash. As shown in Figure 4.4, the ranking order of predictors remains the same for both training and analysis

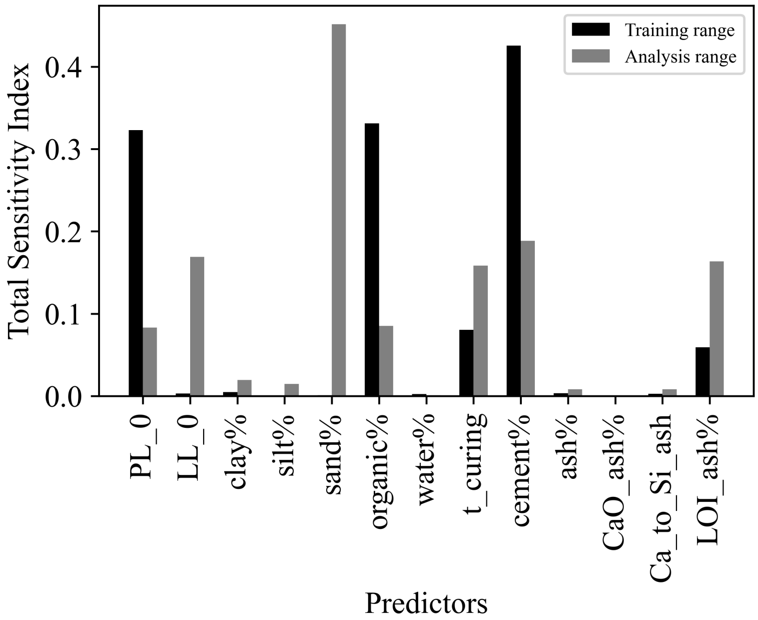


Figure 4.3. Predictor importance for the model GEP-Pha-1

ranges of predictor values. The highest importance is now attributed to *water*%, which contrasts with its lowest sensitivity index in the model GEP-Pha-1. Predictors *water*% and *cement*%, identified as the two most influential predictors in this model, achieved total sensitivity indices of 63% and 42% within the training range of predictor values, and 79% and 59% within the extended analysis ranges, respectively. Unlike the previous model, GEP-Pha-2 displays considerable sensitivity to *silt*%. The model also exhibits some sensitivity to *sand*%, *t_curing* and *ash_C* and the least sensitivity to *clay*%, *lime*%, *slag*%, and *ash_F*%.

Predictors *water*% and *cement*%, identified as the two most influential predictors in this model, achieved total sensitivity indices of 63% and 42% within the training range of predictor values, and 79% and 59% within the extended analysis ranges, respectively.

Overall, the above findings show that while models RM-Sar-3 and GEP-Pha-2 demonstrated stable predictor rankings, GEP-Pha-1 exhibited shifts when applied beyond the ranges of their training dataset. This highlights the importance of not only identifying key predictors but also assessing the robustness of predictor influence when models are applied to broader soil stabilisation contexts.

4.2 Analyses 2: Qualitative Parametric Comparison of Common Predictors

This analysis examines the common predictors used in models RM-Sar-3, GEP-Pha-1, and GEP-Pha-2, evaluating the extent to which these models produce consistent qualitative relationships between inputs and outputs. Figure 4.5 to Figure 4.7 show the results of this set of analyses. Ash content is a key predictor common to models RM-Sar-3, GEP-Pha-1, and GEP-Pha-2, even if quantified with different variables

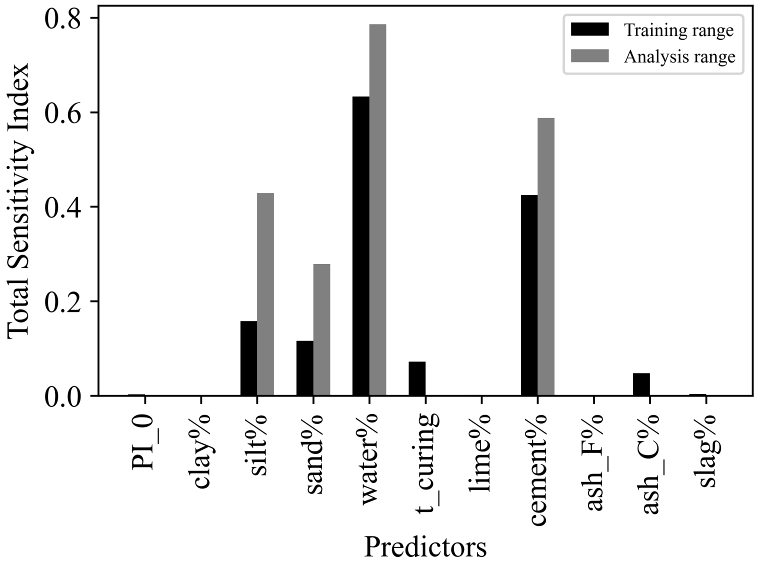


Figure 4.4. Predictor importance for the model GEP-Pha-2

(*ash_C*% and *ash_F*% in model GEP-Pha-2, and *ash*% in models RM-Sar-3 and GEP-Pha-1). Results, as shown in Figure 4.5, show a remarkably consistent set of relationships between UCS and ash content amongst the three models, all yielding positive, linear relationships, with GEP-Pha-1 exhibiting a slight non-linearity at lower *ash*%.

Curing time is another predictor that is common to the three models. Figure 4.6 shows a positive relationship between *t_curing* on *UCS*. The relationships appear to be linear in models RM-Sar-3 and GEP-Pha-2, with a slight non-linearity in model GEP-Pha-1, especially at low *t_curing*.

Further comparisons can be made by assessing the consistency of the behaviour of models GEP-Pha-1 and GEP-Pha-2 with respect to the remaining shared predictors, namely *clay*%, *silt*%, *sand*%, and *water*%. The results are presented in Figure 4.7. In terms of the direction of the relationship (positive or negative), the models exhibit similar behaviour for *silt*%, *sand*%, and *water*%, but not *clay*%. However, even where there is agreement in the direction of the relationship, the relationships themselves are different as can be seen in the figure. The direction of the relationship demonstrated by the model GEP-Pha-2 appears to better align with experimental findings reported by Odeh and Al-Rkaby (2022).

Based on the above findings, the three models exhibit comparable responses to significant predictors such as *ash*%, *silt*%, *sand*%, *water*%, and *t_curing*. The two most conspicuous inconsistencies are observed in the relationships between UCS and *clay*% and *silt*% yielded by GEP-Pha-1 and GEP-Pha-2. The first set of sensitivity analyses has shown that UCS has low overall sensitivity to *clay*% in both models, which makes this particular inconsistency less consequential. For *silt*%, GEP-Pha-2 and GEP-Pha-1 exhibit linear and highly non-linear relationships, respectively, with UCS. While GEP-Pha-1 has low overall sensitivity to *silt*% (see Figure 4.3), this variable is found to be significant for GEP-Pha-2 (see Figure 4.4).

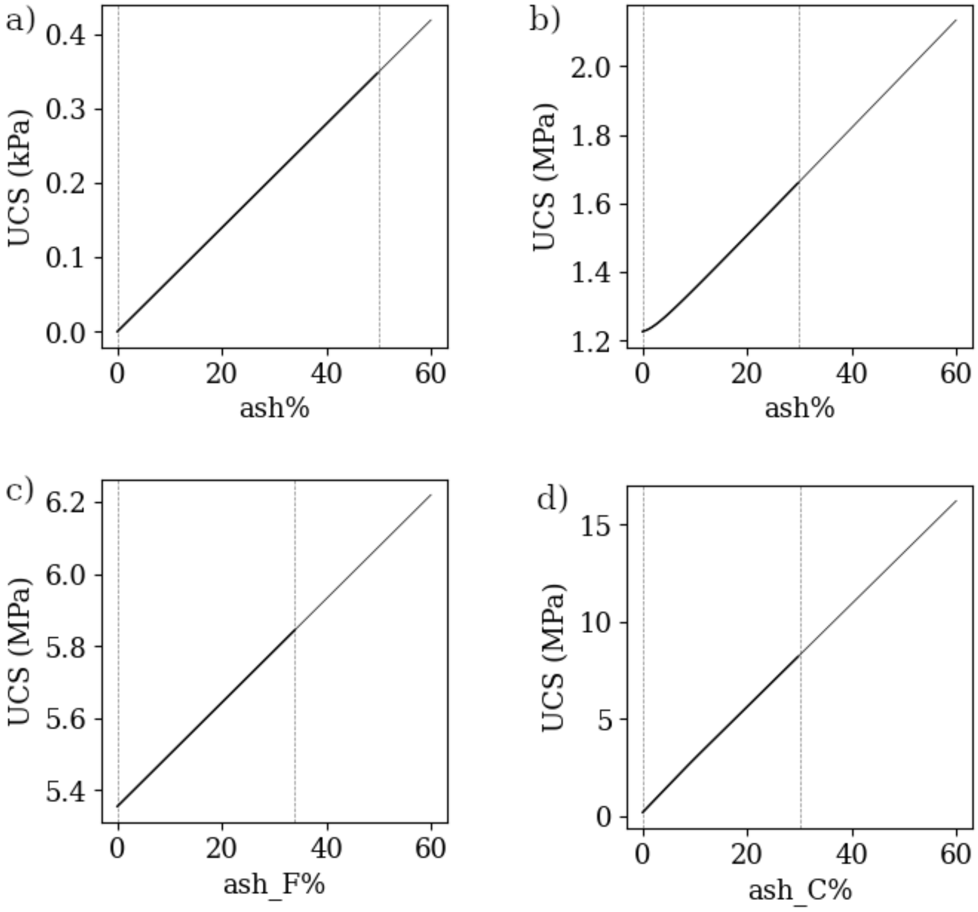


Figure 4.5. Behaviour of (a) RM-Sar-3, (b) GEP-Pha-1, and (c & d) GEP-Pha-2: UCS versus *ash*% within the range of values in their training data (shown within the two dashed lines) and outside

4.3 Analyses 3: Empirical Consistency Assessment Across Five Models

The relationships between UCS predictions of the model RM-Con and its three input variables appear to be non-linear, as shown in Figure 4.8(a) to (d). As discussed above, Consoli et al. (2016) recommend specific conditions for their model to produce reliable UCS estimates. Figure 4.8 (a) presents the effect of *DUW_mix* on *UCS*. Increasing *DUW_mix* yields a slow exponential increase in *UCS* within and beyond the maximum recommended value for *DUW_mix* (17 kN/m^3). Based on these results, *DUW_mix* values of less than the minimum recommended value (14 kN/m^3) did not provide significant strength development. Figure 4.8 (b) demonstrates the positive influence of *ash*% on UCS, up to an optimum *ash*% beyond which UCS declines. This is consistent with experimental findings reported by Mathiluxsan et al. (2016). Figure 4.8 (b) and Figure 4.8 (c) illustrate the effect of *lime*% and *DUW_mix* on UCS predictions. A positive non-linear relationship exists between these predictors and UCS within the recommended ranges of predictor values. *DUW_lime* has a negative effect on UCS as presented in Figure 4.8 (d).

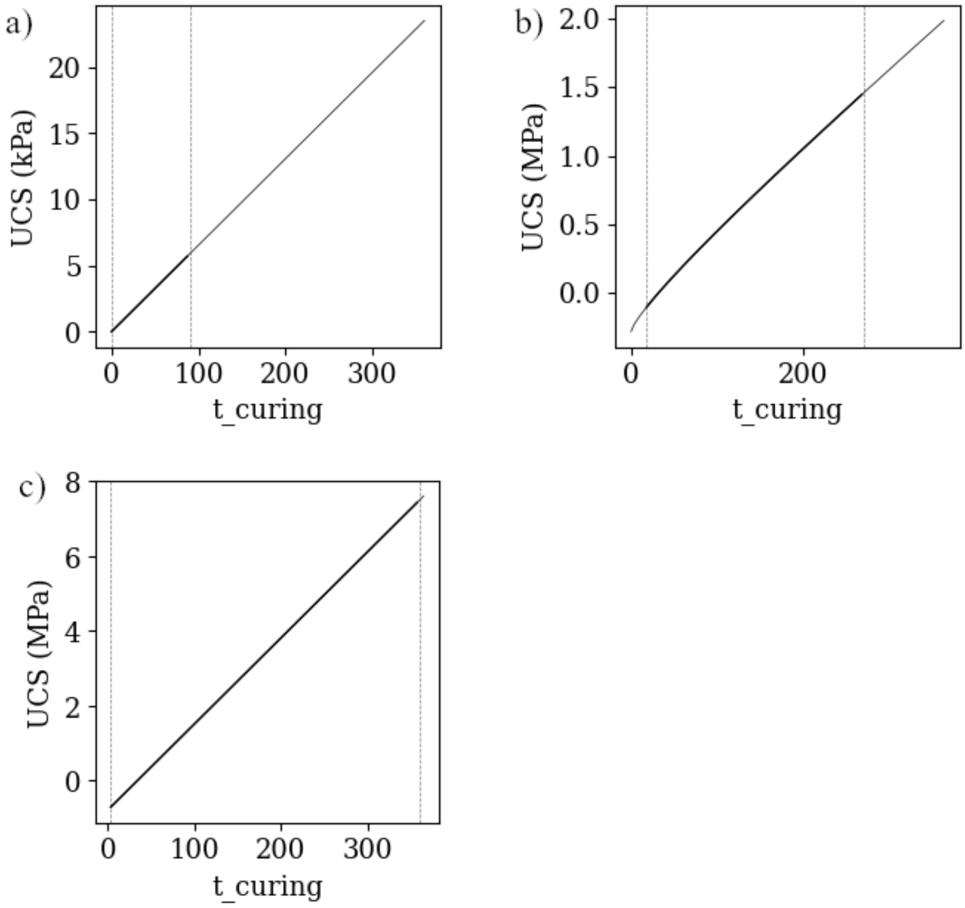


Figure 4.6. Behaviour of (a) RM-Sar-3, (b) GEP-Pha-1, and (c) GEP-Pha-2: UCS versus *t_curing* within the range of values in their training data (shown within the two dashed lines) and outside

Model RM-Sar-3 shows a positive linear relationship between UCS and its predictors including *ash*%, *UCS*_0, and *t_curing* as shown in Figure 4.9 (a) to Figure 4.9 (c). While the positive influence of *UCS*_0 and *t_curing* on the model outputs aligns with experimental findings, the effect of *ash*% does not appear to be adequately captured. The model suggests a continuous, monotonic increase in UCS with increasing *ash*%, which contradicts empirical evidence. Studies such as Ünsever and Diallo (2019) have shown that UCS increases only up to a threshold *ash*% (approximately 25% in that study), beyond which strength declines, indicating a non-monotonic relationship. This behaviour has been consistently observed in both activated and non-activated systems (Mathiluxsan et al., 2016; Ünsever and Diallo, 2019).

Figure 4.10 (a) to Figure 4.10 (h) show that, for model RM-Moz, all relationships between inputs and output are linear, with positive correlations for *LL*_0, *slag*%, *M*, *Na_to_Al_mix*, *A_to_B*, and *Si_to_Al_mix*, and negative ones for *PI*_0 and *ash*%. Simpler models with fewer variables can behave differently due to *ash*% when compared with more complex models allowing for the introduction of more cementitious additives. This is especially the case when high calcium content binders such

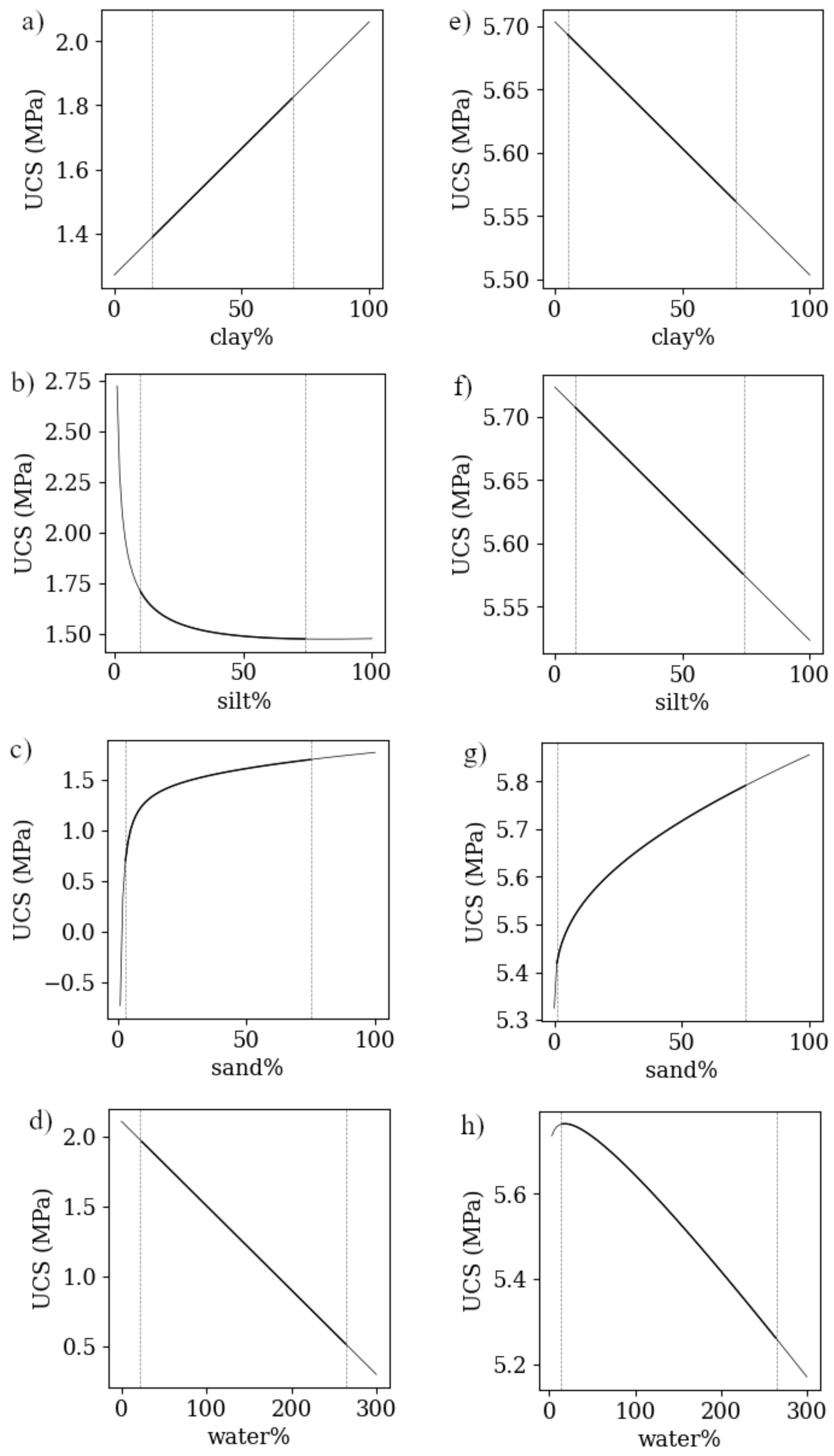


Figure 4.7. The behaviour of models GEP-Pha-1 (a to d) and GEP-Pha-2 (e to h) due to *clay*%, *silt*%, *sand*%, and *water*% within the range of values in their training data (shown within the two dashed lines) and outside

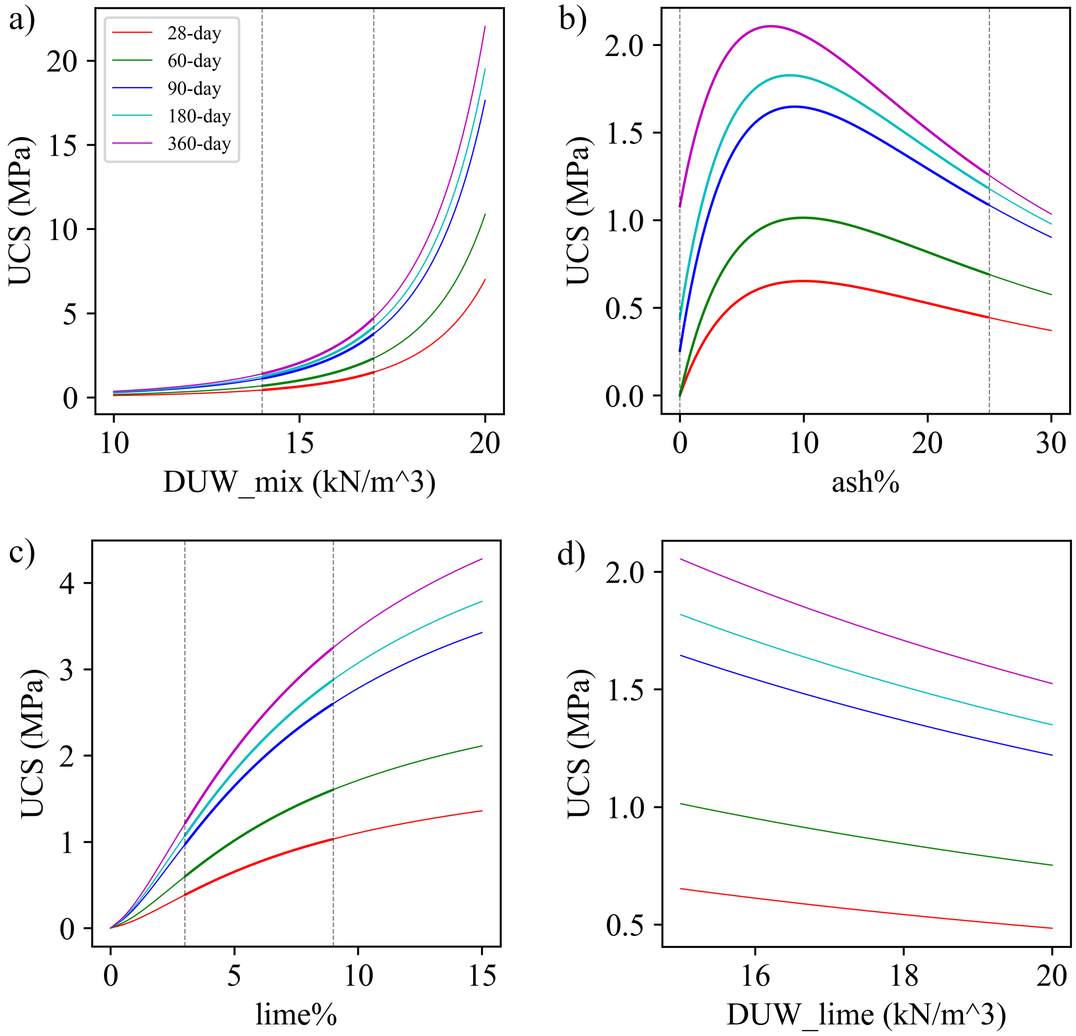


Figure 4.8. Model RM-Con: The effect of (a) DUW_mix, (b) DUW_lime, (c) ash%, and (d) lime% on UCS within the range of values in their development data (shown within the two dashed lines) and outside

as *slag*% and *cement*% are introduced to the mix. Due to the rapid strength development facilitated by high calcium binders, models may develop a negative correlation with *ash*%, which contributes to a much slower rate of strength development. The negative effect of *ash*% on the model RM-Moz may be a reflection of this phenomenon triggered by the strong influence of high calcium *slag*%.

On the other hand, lower values of *slag*% and LL_0 and higher values of PI_0 tend to produce negative UCS values in RM-Moz, pointing to an inconsistency in the model. Another inconsistency is related to Atterberg limits associated with soil composition, especially clay content (Chen, 1975). From the Atterberg limits relationship, $PI = LL - PL$, we know that there is a positive correlation between LL and PI for a given PL. On this basis, we can expect that changes in LL and PI

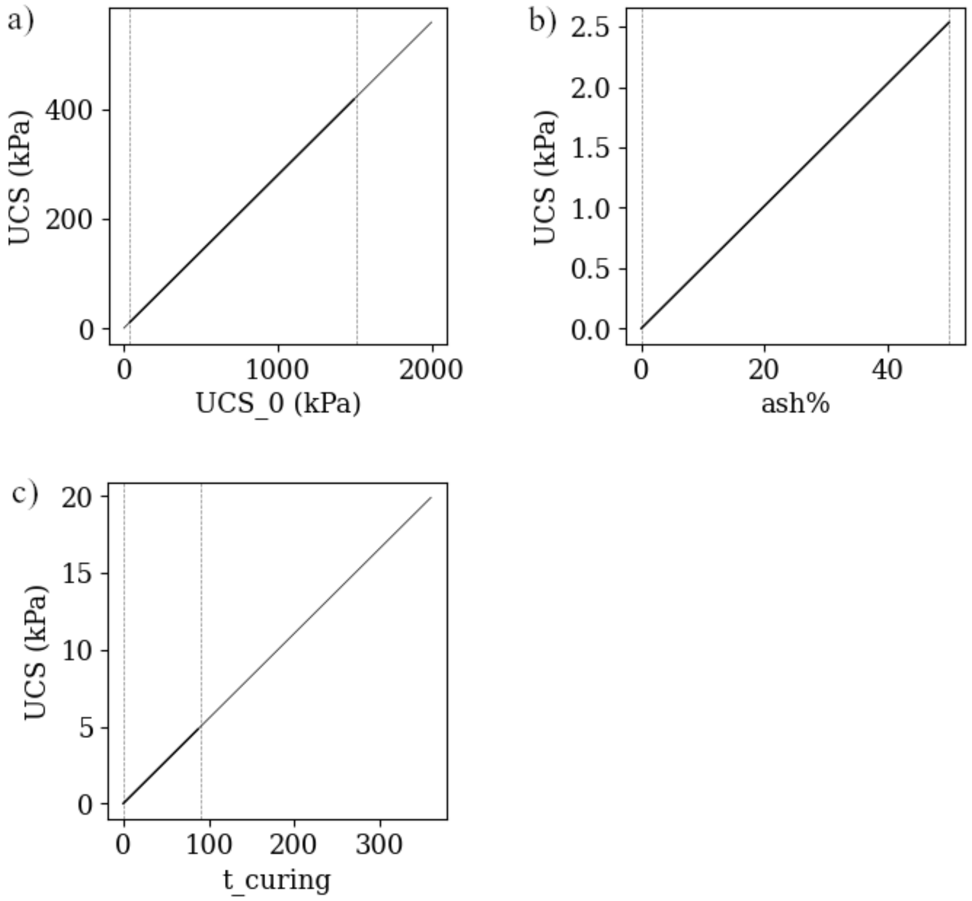


Figure 4.9. Model RM-Sar-3: The effect of (a) ash%, (b) ash%, and (c) t_curing on UCS within the range of values in their training data (shown within the two dashed lines) and outside

should yield similar responses, hence pointing to another inconsistency in RM-Moz which does not seem to have captured the effect of soil composition.

Model GEP-Pha-1 yields a positive linear relationship between predictors PL_0, $clay$%, $organic$%, CaO_ash, $Ca\_to\_Si\_ash$, and LOI_ash and UCS, and a positive non-linear relationship one with $cement$%, ash% and t_curing. However, an increase in LL_0 and $silt$% leads to an exponential decrease in UCS up to a certain value beyond which there appears to be a slow but positive linear relationship between LL_0 and UCS. Conversely, $sand$% leads to a rapid exponential increase in UCS up to a certain $sand$%, beyond which the rate of strength development slows down. The predictor $water$% has a negative linear relationship with the model outcome, which is consistent with the literature (Sharma and Kumar, 2020).

While most of the correlations described above are in agreement with experimental results (Odeh and Al-Rkaby, 2022; Mathiluxsan et al., 2016; Sharma and Kumar, 2020), it is important to note that experimental findings from several studies show a negative correlation between $clay$% and $organic$% on UCS (Toufigh et al., 2022; Tastan et al., 2011; Hebib and Farrell, 2003). Therefore, the positive effects of $clay$% and $organic$% on the GEP-Pha-1 model require further investigation.

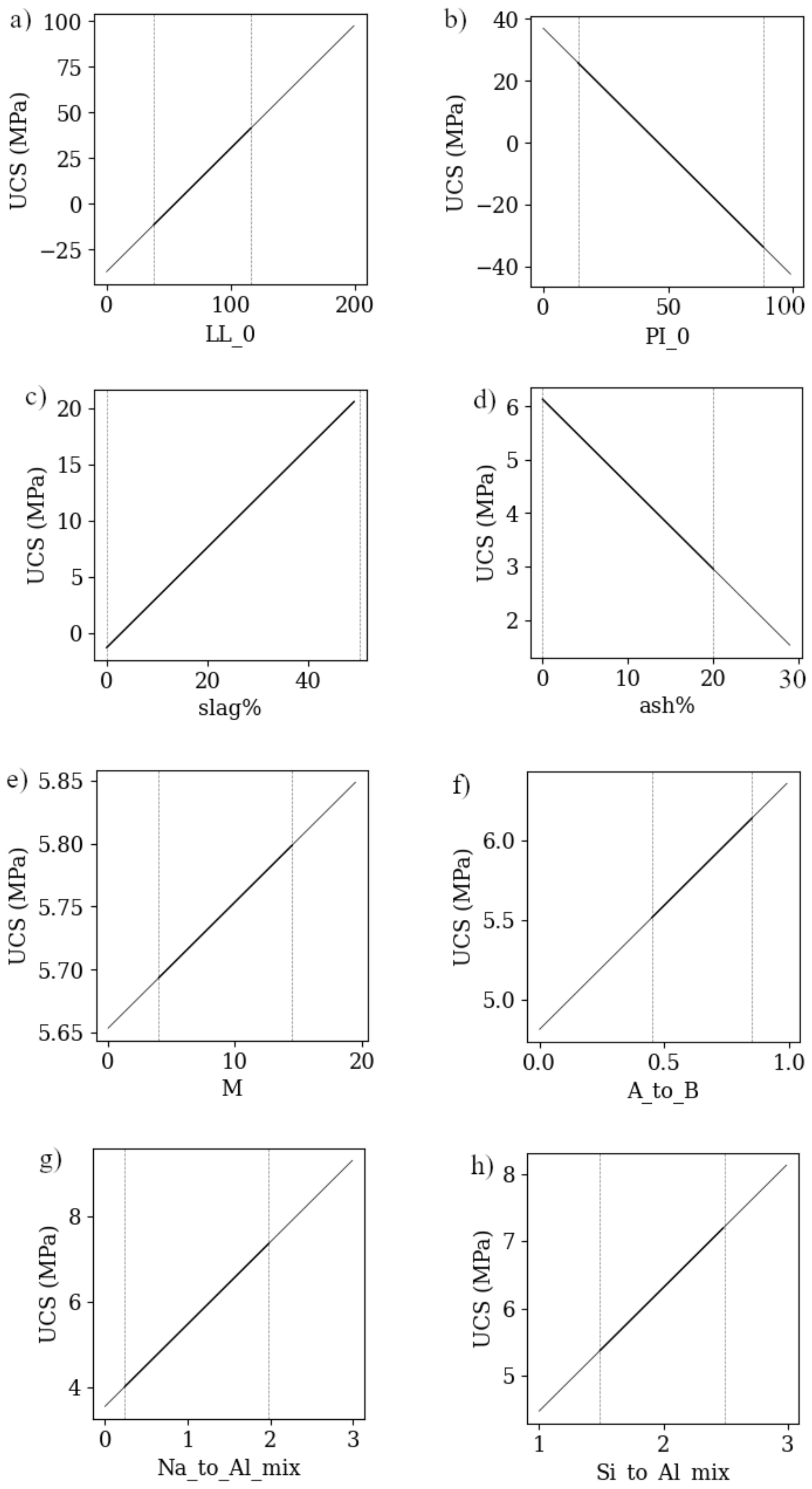


Figure 4.10. Model RM-Moz: The effect of (a) LL_0, (b) PI_0, (c) salg%, (d) ash%, (e) M, (f) A_to_B, (g) Na_to_Al_mix, (h) Si_to_Al_mix on UCS within the range of values in their training data (shown within the two dashed lines) and outside

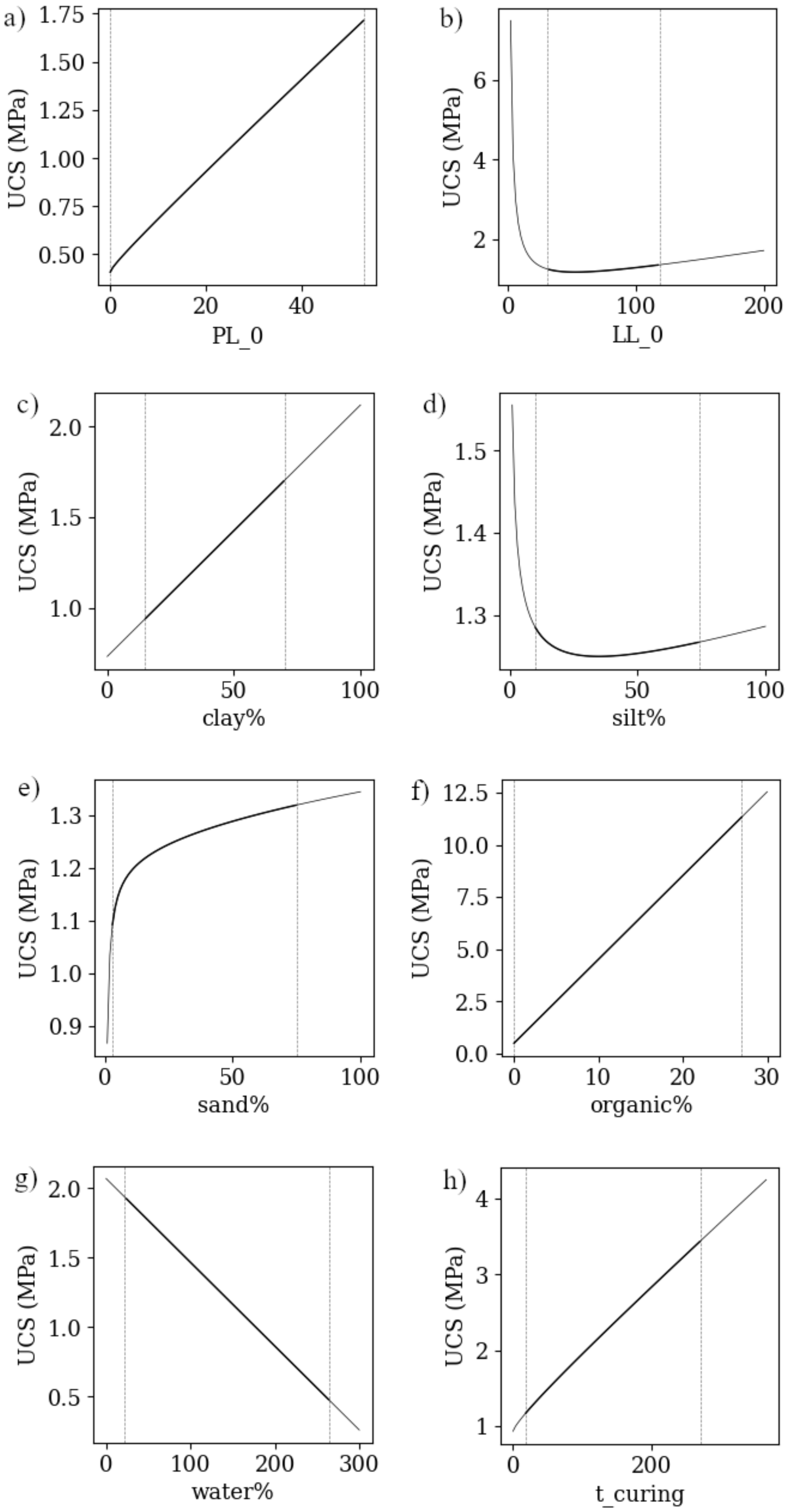


Figure 4.11. Model GEP-Pha-1: The effect of (a) PL_0, (b) LL_0, (c) clay%, (d) silt%, (e) sand%, (f) organic%, (g) water%, and (h) t_curing on UCS within the range of values in their training data (shown within the two dashed lines) and outside

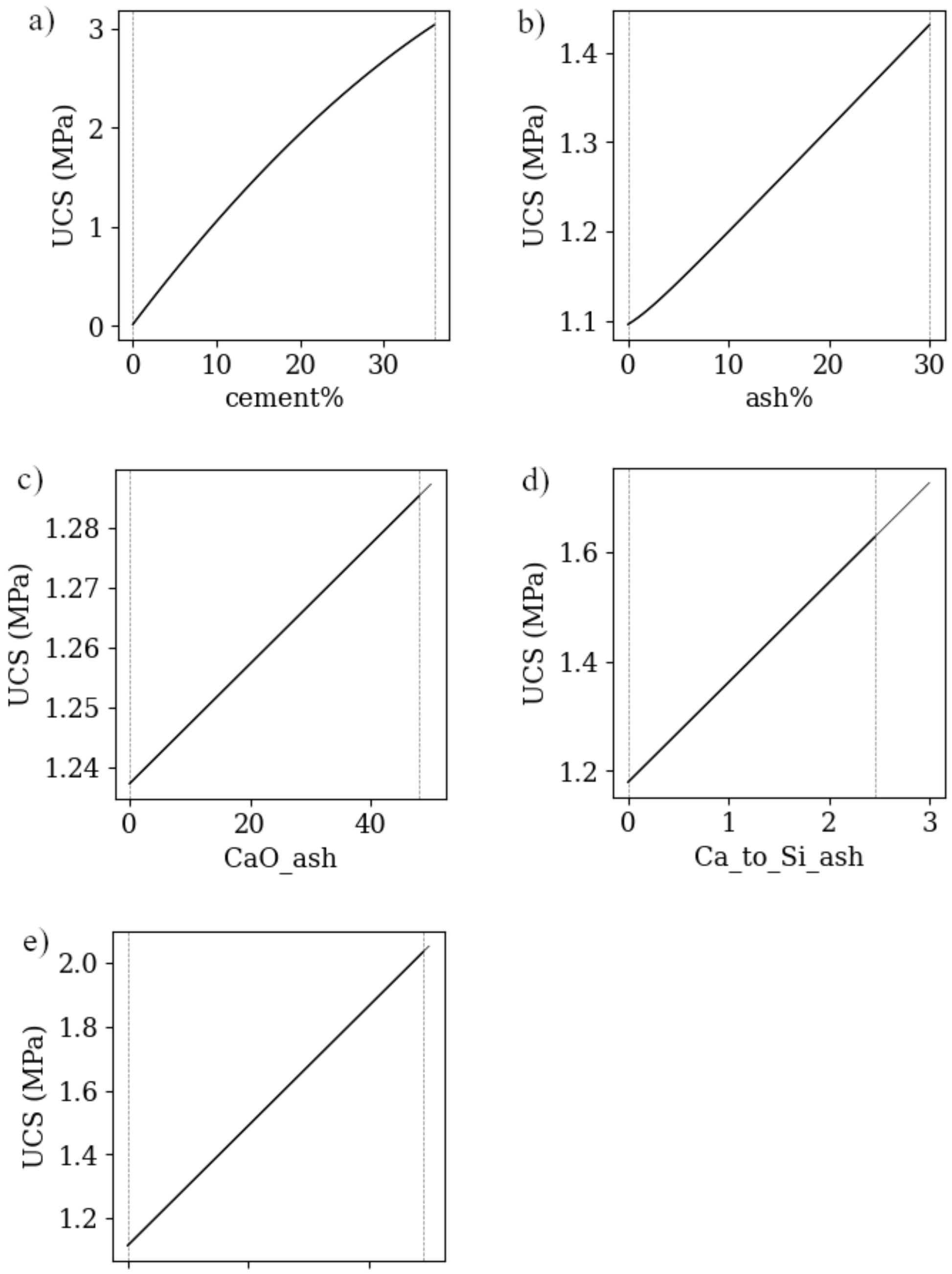


Figure 4.12. Model GEP-Pha-1: The effect of (a) cement%, (b) ash%, (c) CaO_ash, (d) Ca_to_Si_ash, and (e) LOI_ash% on UCS within the range of values in their training data (shown within the two dashed lines) and outside

GEP-Pha-2 yields a positive linear relationship between UCS and *cement*%, *ash_F*%, *PI*_0, and *silt*%, and a positive non-linear one with *ash_C*%, *lime*%, *sand*%, and *t_curing*. A negative linear relationship between model output and *clay*% is observed, which is consistent with experimental findings reported by Odeh and Al-Rkaby Odeh and Al-Rkaby (2022). As shown in Figure 4.13 (e), increasing *water*% leads to a non-linear decrease in UCS, which is consistent with the findings of Sharma and Kumar Sharma and Kumar (2020). Results also suggest that UCS increases linearly with increasing $slag\%$ up to a certain point, after which the relationship reverses linearly as shown in Figure 4.14 (c).

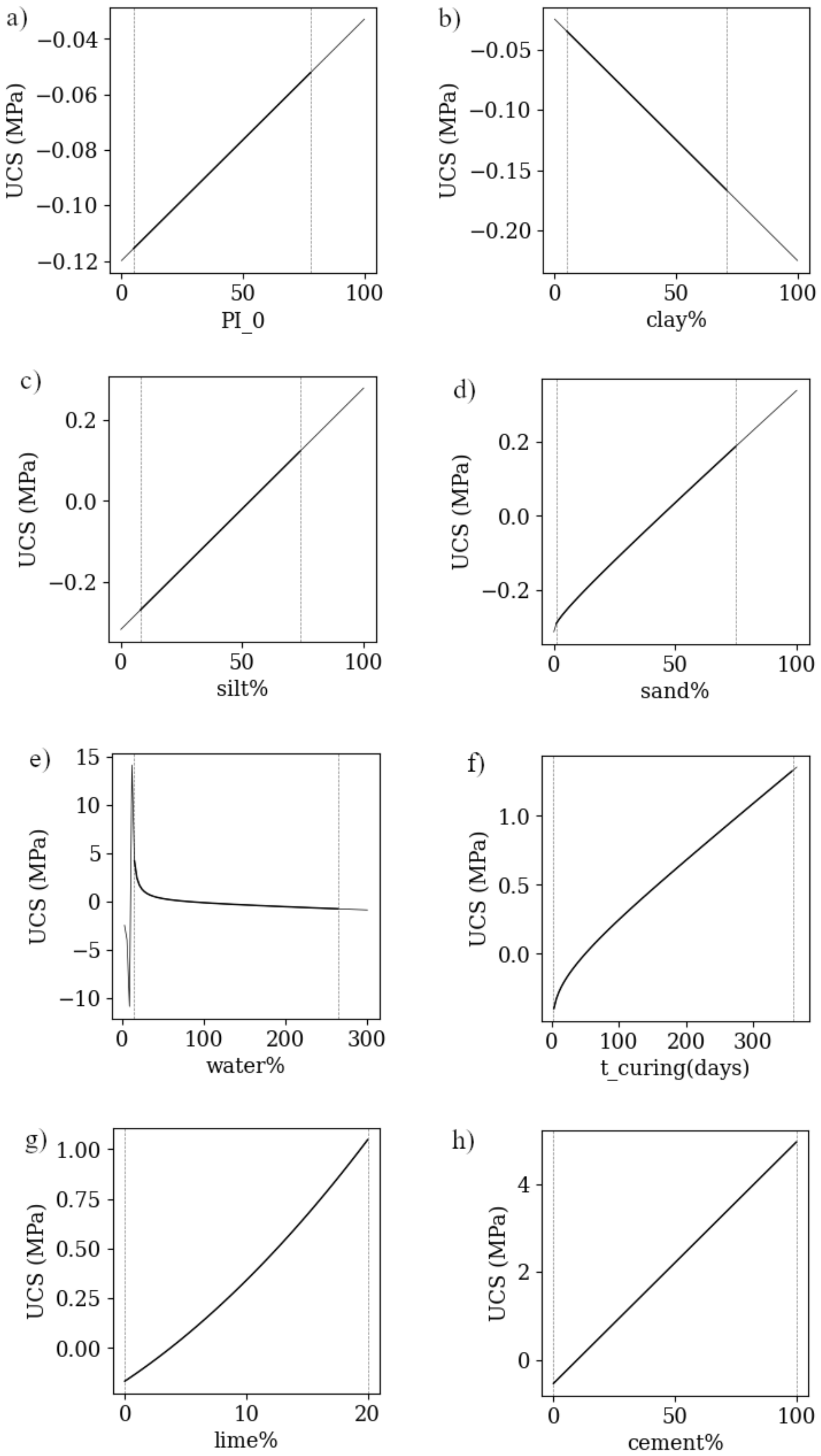


Figure 4.13. Model GEP-Pha-2: The effect of (a) PI_0, (b) clay%, (c) silt%, (d) sand%, (e) water%, (f) t_curing, (g) lime%, and (h) cement% on UCS within the range of values in their training data (shown within the two dashed lines) and outside

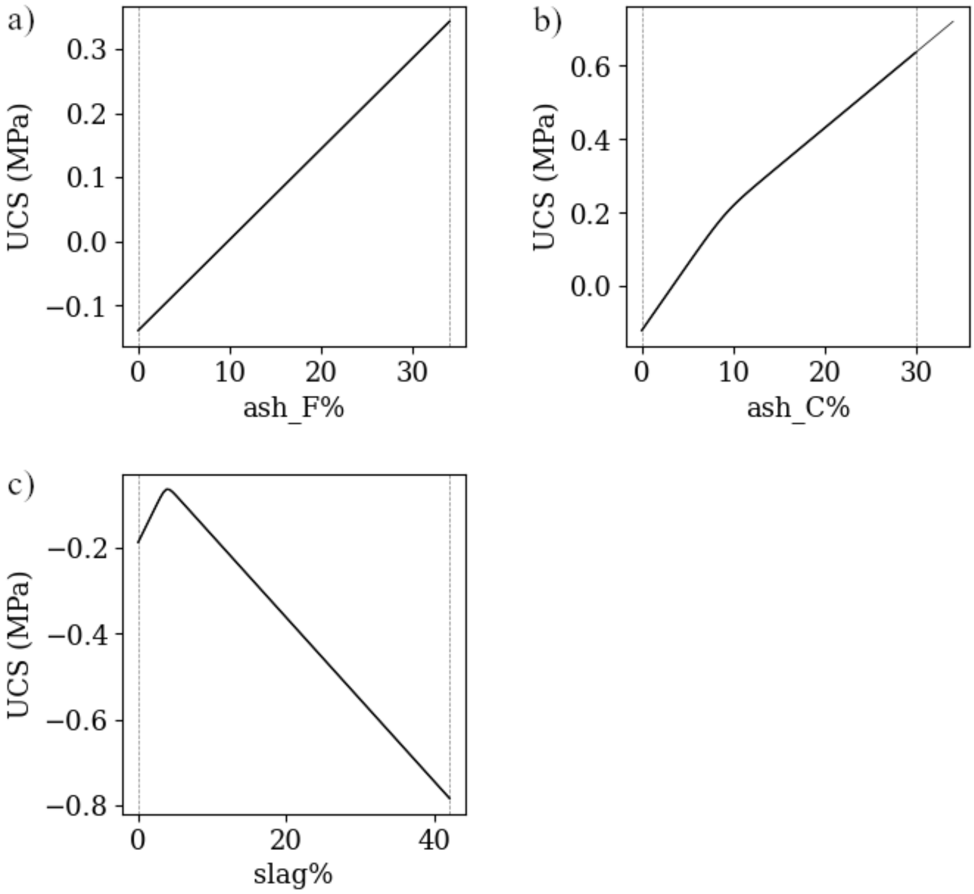


Figure 4.14. Model GEP-Pha-2: The effect of (a) ash_F%, (b) ash_C%, (c) slag% on UCS within the range of values in their training data (shown within the two dashed lines) and outside

4.4 Conclusions

The results of the three analyses collectively provide a systematic evaluation of existing UCS models for soils stabilised with aluminosilicate waste. While these models capture key aspects of soil stabilisation behaviour, their performance and applicability are strongly influenced by the predictors included and the range of their values in the training data. Consistency in predictor influence across models depends largely on the similarity of their predictor selection. For example, in GEP-Pha-1 and GEP-Pha-2, where *cement*% is a critical predictor, aluminosilicate content (e.g., *ash*%) plays only a minor role in UCS estimation, whereas in RM-Sar-3, which does not incorporate cement as a binder, *ash*% emerges as the most important predictor. The above analyses revealed that models mostly align reasonably well with the literature, reinforcing confidence in many of the relationships identified by the models.

Results also highlighted several limitations potentially in data quality, model optimisation, and generalisability that led to results that did not fully align with empirical evidence. For binder content, RM-Sar-3 predicted a continuous increase in UCS with increasing *ash*%, without the optimum content typically observed in experiments, while RM-Moz predicted a negative *ash*%–UCS relationship, likely due to the dominant influence of high-calcium and rapidly reacting slag. For soil properties, GEP-Pha-1 predicted positive effects of $clay$% and $organic$% on UCS, contrary to the negative correlations consistently reported in experimental studies. RM-Moz also exhibited inconsistent Atterberg limits behaviour, with LL_0 and PI_0 producing opposite effects on UCS despite being positively correlated and typically influencing UCS in the same (negative) direction for a given PL. Collectively, these findings highlight the need for developing new models trained on larger and more diverse datasets already available in the literature, to improve both versatility and generalisability.

CHAPTER 5

PREDICTING STRENGTH OF STABILISED SOILS

As illustrated in the mild-map below, this chapter presents the development of new machine learning models for predicting the strength of soils stabilised with alkali-activated aluminosilicate by-products. It first outlines data preparation, refinement, and the assessment of new predictors, followed by a preliminary evaluation of fourteen algorithms. Based on this screening, Extra Trees (ET), Genetic Programming (GP), and Artificial Neural Networks (ANN) were selected for detailed model development. The chapter then evaluates model accuracy, robustness, and performance uncertainty, and compares the two better-performing models (ET and GP) to highlight their relative performance. Finally, SHAP-based interpretability analyses reveal how key predictors influence unconfined compressive strength, forming a unified and transparent framework for reliable strength prediction in waste-based soil stabilisation.

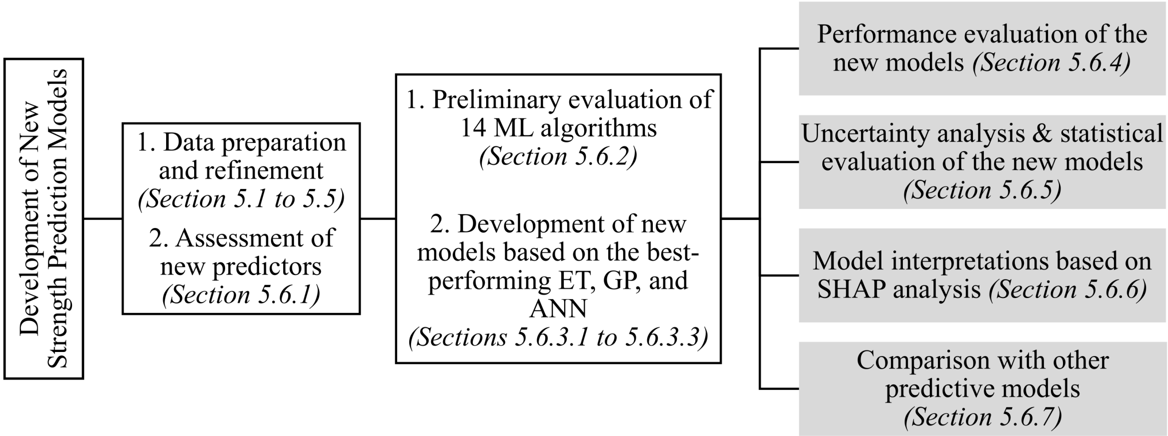


Figure 5.1. Mind-map summarising the structure and major analytical components of Chapter 5

5.1 Data Preparation

A total of 1,598 test records of data were collected from 23 peer-reviewed experimental studies that clearly reported key variables, sample preparation methods, testing procedures, and curing conditions. Before collecting data, a survey of the empirical literature was carried out to identify key predictors as listed in Table 5.1. Table 5.2 presents the details of the dataset. Sample preparation methods resulted in three distinct classes ($sample\_type$): (i) fluid-like samples (non-compacted), (ii) compacted samples below maximum dry density (MDD), and (iii) samples compacted to MDD, represented as 1, 2, and 3, respectively. Figure 5.2 presents the histograms of the collected data in raw form, i.e., before any analysis and refinement. Table 5.1 presents a list of variables and the number of data points collected.

5.2 Treating Missing Data

As shown in Table 5.1, some of the test records in the data contained missing points, which were related to either the clay chemical composition or water content values. Various methods can be used to handle missing data points, such as filling in missing data points with the average values of their respective columns, deleting test records that contain missing data points, or applying domain knowledge to develop mathematical relationships. In this case, handling missing data points involved searching for a predictor with a reasonable correlation with the missing data point to derive mathematical relationships. Simple linear and exponential expressions were developed with linear expressions presented in Figure 5.3 achieving higher R^2 scores. The following mathematical relationships were developed between clay% and variables with missing data point namely CaO_clay%, Fe2O3_clay%, SiO2_clay%, Al2O3_clay%. Another relationship was developed between sand% and water%. These relationships were used to fill in the missing data points.

To check for potential bias introduced by filling the missing data, the performances of two distinct scenarios were compared: i) preparing a dataset with test records containing missing values removed, and ii) using the above mathematical relationships to fill in the missing values. The evaluation was conducted using cross-validation. Results showed that both datasets yielded similar R^2 scores and no noticeable signs of bias. Hence, the dataset with estimated missing data points was used in this study.

Table 5.1. List of collected variables, their units, acronyms, and the number of data points collected

Variables	Units[1]	Acronym	Data Points
Properties of the host soil:			
Liquid limits	%	LL	1598
Plastic limit	%	PL	1598
Plasticity index	-	PI	1598
Clay fraction	%	clay%	1598
Silt fraction	%	silt%	1598
Sand fraction	%	sand%	1598
Properties of the mixture:			
Ash content	%	ash%	1598
Class C ash content	%	ash_C%	1598
Class F ash content	%	ash_F%	1598
Slag content	%	slag%	1598
Gravimetric water content of mixture	%	water%	1498
NaOH content	%	NaOH%	1598
NaOH Molarity	M	M	1598
Na_2SiO_3 content	%	Na2SiO3%	1598
Clay chemical compounds:			
CaO content in clay%	%	CaO_clay%	1162
Fe2O3 content in clay%	%	Fe2O3_clay%	1168
SiO2 content in clay%	%	SiO2_clay%	1168
Al2O3 content in clay%	%	Al2O3_clay%	1168
Ash chemical compounds:			
CaO content in ash%	%	CaO_ash%	1598
Fe2O3 content in ash%	%	Fe2O3_ash%	1598
SiO2 content in ash%	%	SiO2_ash%	1598
Al2O3 content in ash%	%	Al2O3_ash%	1598
Slag chemical compounds:			
CaO content in slag%	%	CaO_slag%	1598
Fe2O3 content in slag%	%	Fe2O3_slag%	1598
SiO2 content in slag%	%	SiO2_slag%	1598
Al2O3 content in slag%	%	Al2O3_slag%	1598
Sample preparation methods:			
Types of samples	Integer	sample_type	1598
Curing conditions:			
Curing temperature	°C	T_curing	1598
Curing time	Days	t_curing	1598
Target variable:			
UCS	MPa	UCS	1598

---

[1] All content variables represent the percentages of mass relative to the mass of dry soil.

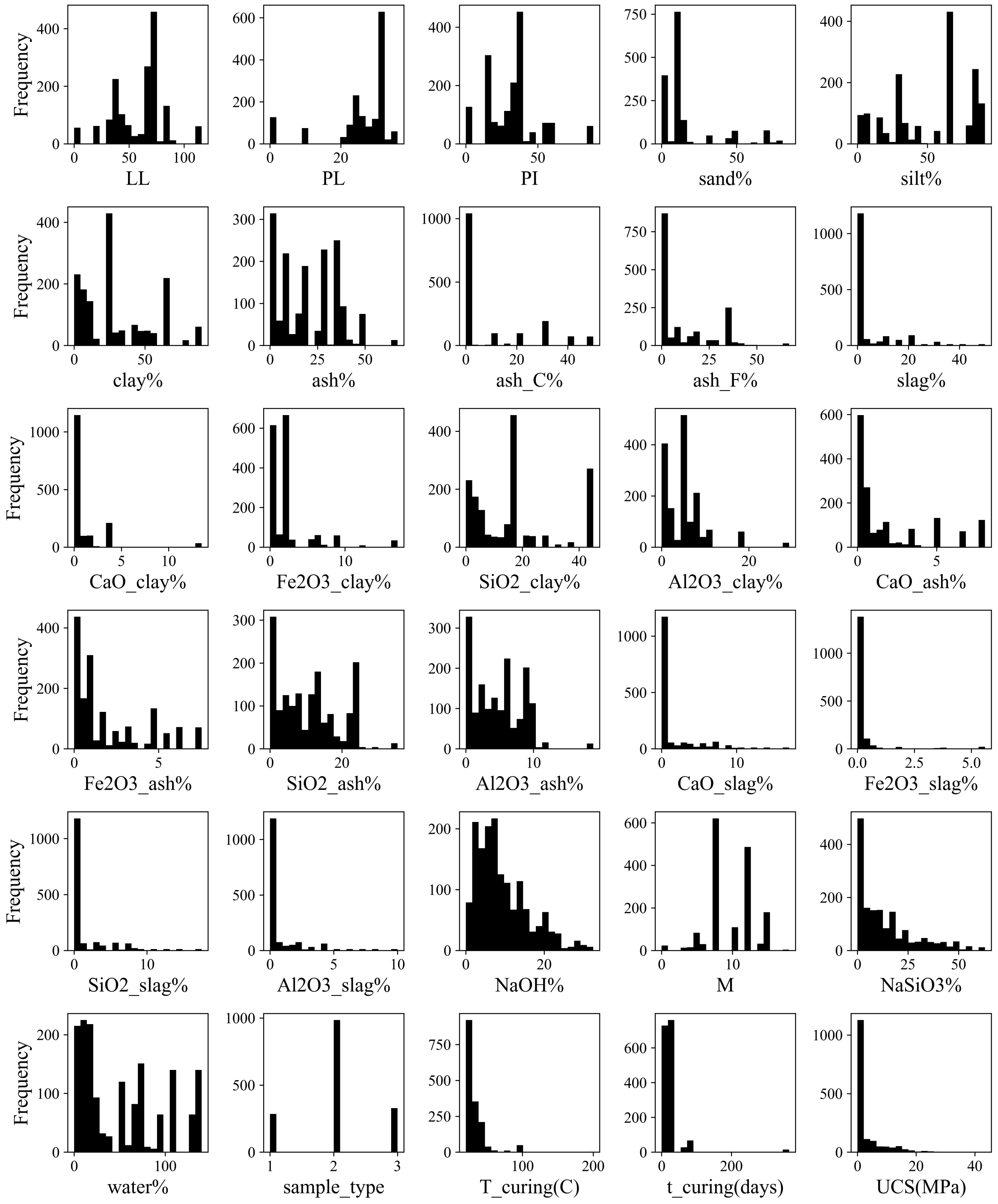


Figure 5.2. Histograms of the collected data

5.3 Assessing Covariance and Multicollinearity

The highly correlated predictors namely LL, PL, PI, sand%, silt%, and clay% were replaced with a new and better-performing predictor, i.e., sand_to_fines ratio. Predictors CaO% of clay, ash, and slag were aggregated into a single predictor called total_CaO%. Similarly, Fe2O3%, SiO2%, and Al2O3% of clay, ash, and slag were also combined and named total oxides%. Table 5.3 presents the final selection of predictors and contains brief clarification on how new predictors were created. The

Table 5.2. Sources of data, the number of test records collected from each source, types of additives used, and sample types

No.	Publications	Test Records	Soil	Ash Class	Slag	NaOH	Na_2SiO_3	Sample Type
1	Suksiripattanapong et al. (2021)	419	✓	C	×	✓	✓	2, 3
2	Mozumder and Laskar (2015)	286	✓	F	✓	✓	✓	2
3	Phetchuay et al. (2016)	210	✓	F	×	✓	✓	3
4	Vu et al. (2020)	79	✓	F	×	✓	✓	2
5	Varma et al. (2021)	72	✓	F	×	✓	✓	1
6	Leong et al. (2018)	61	✓	F	×	✓	✓	2
7	Singhi et al. (2016)	60	✓	C	✓	✓	✓	2
8	Toufigh et al. (2022)	40	✓	F	✓	✓	✓	1
9	Murmu et al. (2020)	40	✓	F	×	✓	×	1
10	Yaghoubi et al. (2018)	40	✓	F	✓	✓	✓	2, 3
11	Cristelo et al. (2011)	38	✓	F	×	✓	✓	2
12	Cristelo et al. (2013)	37	✓	F	×	✓	✓	3
13	Phummiphan et al. (2018)	36	✓	C	✓	✓	✓	1
14	Sukprasert et al. (2021)	34	✓	C	✓	✓	×	1
15	Odeh and Al-Rkaby (2022)	31	✓	C	×	✓	✓	2
16	Turkane and Chouksey (2022)	24	✓	F	×	✓	✓	1
17	Abdullah et al. (2019)	21	✓	F	✓	✓	✓	1
18	Abdullah and Ahmad (2017)	18	✓	F	✓	✓	✓	1
19	Phummiphan et al. (2016)	16	✓	C	×	✓	✓	1
20	Ahmed Hisseini et al. (2021)	12	✓	F	✓	✓	✓	1
21	Cristelo et al. (2012)	9	✓	F	×	✓	✓	3
22	Mypati and Saride (2022)	9	✓	F	×	✓	✓	1
23	Rios et al. (2016)	6	✓	F	×	✓	✓	1

table also provides a statistical summary of the dataset after analysis and refinement. Test records with T_curing(C) > 60°C were removed from the dataset as high curing temperatures are not representative of in-situ curing conditions. Similarly, test records with t_curing(days) < 1 day were removed due to their negligible practical significance. Also, test records with identical input variables were grouped, retaining only one representative test record per group, which included the average UCS value for that cluster. The removal of these test records was validated through CV, which confirmed an improvement in model accuracy. Overall, 288 test records were removed from the dataset reducing the size of the refined data to 1,310 test records.

Correlation matrices of the predictors are presented in Figure 5.4 (a). The Pearson correlation coefficients in these figures show that feature engineering successfully

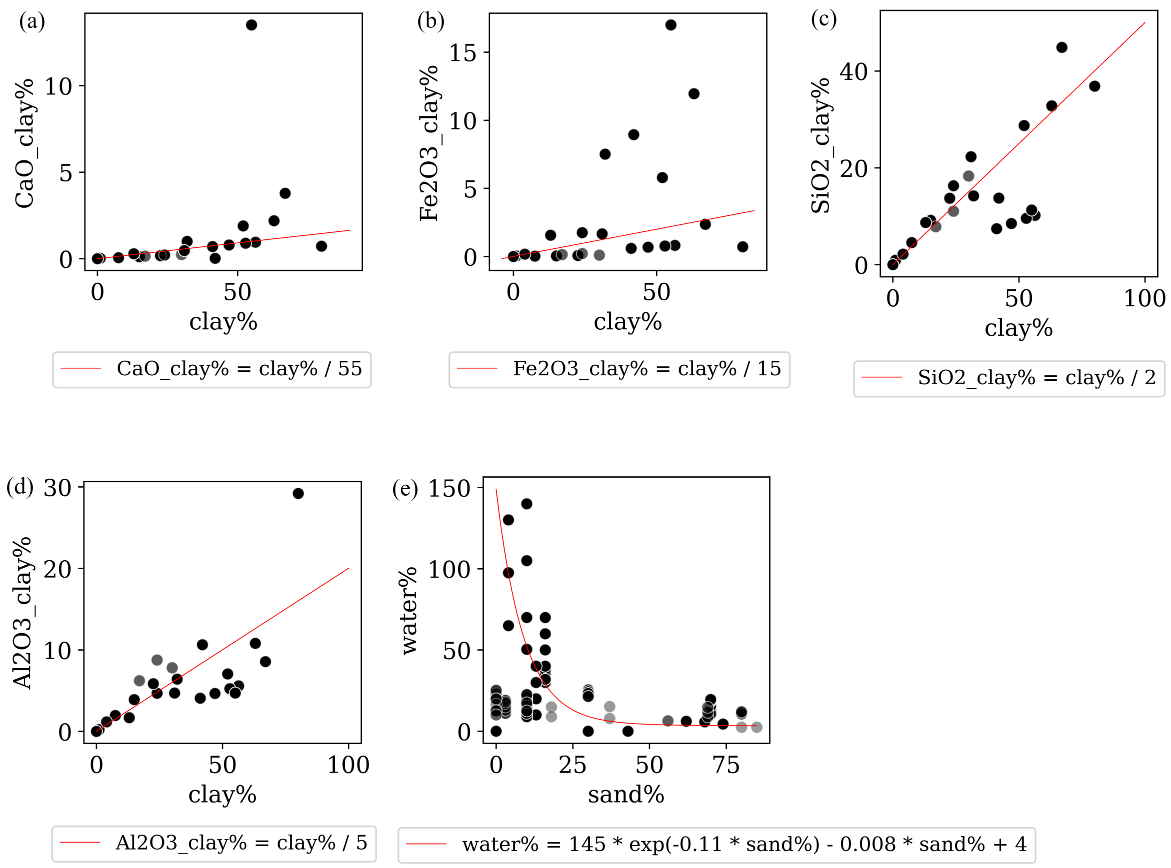
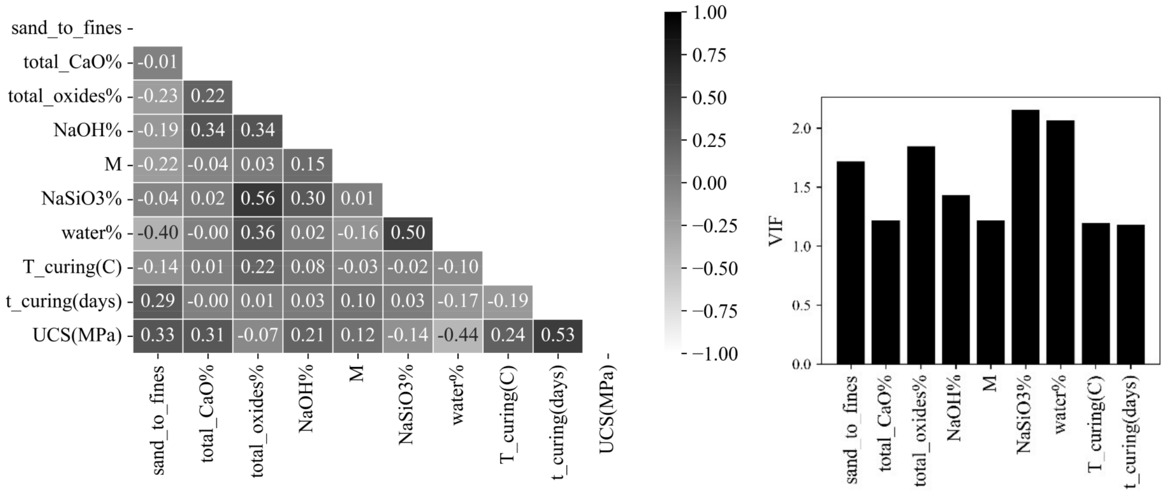


Figure 5.3. Mathematical relationships developed to complete missing data

reduced correlations among selected predictors to below 0.3, with only a few exceptions showing moderate correlation, peaking at 0.56. Figure 5.4 (b) show acceptable levels of multicollinearity, with VIF values < 3. These results show that the impact of multicollinearity has been effectively minimised.



(a) (b)

Figure 5.4. Predictor Correlations (a) and their Variance Inflation Factors (b)

Table 5.3. Optimal set of predictors, their statistical descriptions, and predictor descriptions

Predictors	**Test Records**	**Min**	**Mean**	**Max**	**Predictor description and notes on feature engineering**
sand_to_fines	1310	0.00	0.16	0.85	The ratio of sand fraction to fines in the host soil
total_CaO%	1310	0.02	4.67	22.88	The CaO contents in the mixture were combined by calculating the weighted sum of CaO_clay%, CaO_ash%, and CaO_slag%. This was achieved by multiplying the representative percentages of CaO in their respective source materials by the content of their source material and then dividing by 100. For instance, the calculation for CaO_ash% is as follows: CaO_ash% = (CaO_ash × ash%) / 100, where CaO_ash is the amount of CaO in ash (in percentage) and ash% is the amount of ash in the mixture.
total_oxides%	1310	5.64	43.94	98.25	Combined oxides of clay%, ash%, and slag%. Oxides considered include Fe2O3%, SiO2%, and Al2O3%. The same approach as presented in Equation (6) was followed to quantify individual oxides for each constituent.
NaOH%	1310	0.00	9.24	32.50	Sodium hydroxide content
M	1310	0.00	10.01	18.00	Sodium hydroxide molarity
NaSiO3%	1310	0.00	14.14	63.00	Sodium silicate content
water%	1310	0.00	53.23	140.00	Total water content
T_curing(C)	1310	20.00	28.54	60.00	Curing temperature
t_curing(days)	1310	1.50	24.84	365.0	Curing time
UCS(MPa)	1310	0.04	2.97	43.17	Unconfined compressive strength

5.4 Addressing Performance Uncertainty

Histograms in Figure 5.5 show that the predictors have varying value ranges and are strongly affected by skewness and irregular distributions, which may lead to learning difficulties and increased model uncertainty (Zhuang et al., 2020). To mitigate data-induced uncertainty and evaluate the impact of data variability on model performance, min-max normalisation and bootstrap resampling, a common tool for uncertainty analysis (Efron, 1979), were applied. Min-max normalisation was used to scale all predictor values to a standard range of 0 to 1, ensuring that variables such as $t\_curing(days)$ and $water\%$ do not disproportionately influence the model due to their larger values. This scaling balanced the contribution of all predictors, reduced the risk of bias, and improved the learning process.

Bootstrap resampling helps assess the reliability and stability of a model by repeatedly training it on different versions of the dataset and analysing how its performance changes. We used the following steps to study model uncertainty using bootstrap resampling. First, 20 bootstrap datasets were generated by randomly sampling test records with replacement from the original dataset. Second, individual models were trained on each bootstrap sample to assess sensitivity to data variations. Third, each model was evaluated using out-of-bag (OOB) samples, i.e., test records excluded during resampling, to estimate generalisation to unseen data, recording R^2 scores for training and testing phases. Finally, the mean and 95% confidence interval of these scores were calculated; narrower intervals indicated greater stability and reliability, whereas wider intervals reflected increased uncertainty.

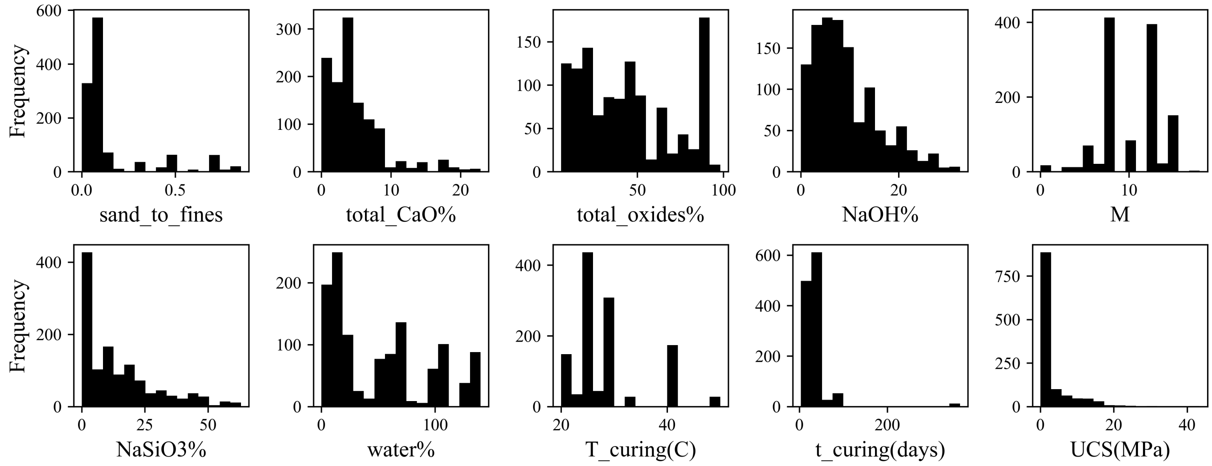


Figure 5.5. Mathematical relationships developed between predictors with missing data and clay% and sand%

5.5 Refined Dataset Compared with Other Studies

The refined dataset contained 1,310 test records from 23 experimental studies, which is significantly larger than the largest dataset of existing models in the literature, including cement and lime-based models, on both counts: RM-Sar with 218 test records from 20 papers and GP-Pha-2 with 1183 test records from 13 papers. A larger number of papers generally indicates a wider range of soil types, sample preparation methods, and curing conditions. A broader range indicates greater variability within the dataset, which is subsequently learned by the model.

Table 5.4 presents a comparison of the predictor value ranges in our dataset against those used by other models. The variability ratio V_r values provided in brackets in the table are calculated using the following equation:

$$V_r = \frac{\text{Max}_i - \text{Min}_i}{\text{Max}_o - \text{Min}_o}, \tag{5.1}$$

where Max_i and Min_i represent the maximum and minimum values of the predictors in our dataset, and Max_o and Min_o are the corresponding values in the datasets used by other models.

A V_r value greater than or equal to 1 indicates that the variability range in our dataset is at least as wide as that of the other dataset. Except for total water content (%) in GEP-Pha-1 and GEP-Pha-2, all V_r values in Table 5.4 are equal to or greater than 1. For some predictors, V_r is often much larger than 1, highlighting the broader coverage provided by our dataset.

For the UCS dataset, 60 test records were set aside as a holdout set for final evaluation of model performance, while the remaining data were split into an 80% training subset and a 20% testing subset.

Table 5.4. The range of predictor values in the refined dataset compared with other models' datasets

Predictors	**ANN_Moz**	**RM_Moz**	**RM_Con**	**ANN_Leo**	**GP_Leo**	**RM-Sar-3**	**ANN-Nis-3**	**ANN-Nis-4**	**GEP-Pha-1**	**GEP-Pha-2**	**RF-Zei**	**Refined Dataset**
$ash\_C\%$										0 - 30 (1.67)		0 - 50
$ash\_F\%$										0 - 34.5 (1.94)		0 - 67
$slag\%$	0 - 50 (1)	0 - 50 (1)								0 - 42.5 (1.18)	0 - 50 (1)	0 - 50
$NaOH\%$				6.35 - 13.4 (4.61)	6.35 - 13.4 (4.61)							0 - 32.5
M	4 - 14.5 (1.71)	4 - 14.5 (1.71)						3 - 15 (1.5)			4 - 14.5 (1.71)	0 - 18
$Na2SiO3\%$				4.23 - 11.4 (8.79)	4.23 - 11.4 (8.79)							0 - 63
$water\%$				4.07 - 15.9 (11.83)	4.07 - 15.9 (11.83)				22 - 265 (0.58)	14 - 265 (0.56)		0 - 140
t_curing							1 - 90 (4.04)	0 - 28 (12.98)	3 - 270 (1.36)	3 - 360 (1.02)		0 - 365
T_curing												20 - 60
$sand\_to\_fines$												0 - 0.85
$total\_CaO\%$												0.02 - 22.88
$total\_oxides\%$												5.63 - 98.25

Note: Empty cells indicate predictors not considered in the respective models. Values in parentheses represent variability ratio V_r.

5.6 Results and Discussions

5.6.1 Impact of New Predictors on Performance

The impact of new predictors related to sample preparation methods and cementitious chemical compound contents was evaluated based on the performance of the ET method. This was done by conducting a comparative analysis of model performance across four distinct predictor sets (Sets 1 to 4) used to train four models. Table 5.5 presents the list of predictors selected for each set, their multicollinearity analysis results using r and VIF scores, the tuned hyperparameter values, CV scores, and performance metrics for both training and testing phases. Set 1 is the same set of predictors in Table 5.3. All four models were tuned using 15 `estimators`, i.e., the number of decision trees in the forest, with slight variations in the values of `max_features`, `max_depth`, and `min_impurity_decrease` values across models. The `bootstrap` resampling method was incorporated in all cases.

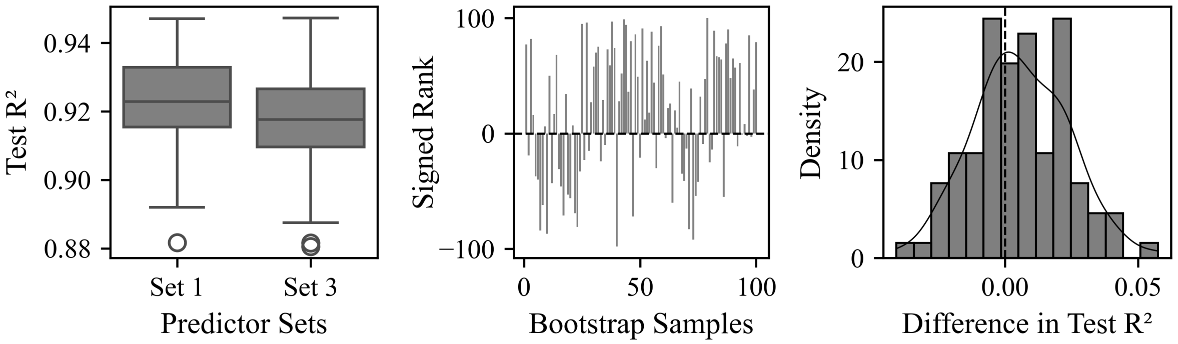
All predictor sets yielded similar performance with nearly identical CV scores (0.92 or 0.93). The R^2 scores were consistently 0.93 for both training and testing across all four scenarios. Evaluation of MAE, MSE, and RMSE also showed similar performances but with slight variations in model's ability to reduce errors. Results showed that inclusion of sample preparation methods, as defined in this study (not compacted, compacted to less than maximum density, compacted to maximum density), does not enhance model performance. Representing *sample_type* as a binary variable (compacted vs. non-compacted) produced similar results. However, replacing the commonly chosen predictors by other modellers *ash_C*%, *ash_F*%, and *slag*% with new predictors representing cementitious chemical compound contents (*total_CaO*% and *total_oxides*%) without *sample_type*, i.e., Set 1, yielded lower errors across all metrics for both training and testing phases.

A Wilcoxon signed-rank test was applied to ET's performance across 100 bootstrap resamples, comparing predictor Sets 1 and 3 (the latter being the most commonly used in the literature). Results are presented in Figure 5.6. The R^2_{test} score distributions for Sets 1 and 3 (Figure 5.6a) are broadly similar, with Set 1 only slightly higher. However, the Wilcoxon signed-rank test confirmed a statistically significant difference between Sets 1 and 3 ($D = Set1 - Set3$). As shown in Figure 5.6b, test results reveal a larger sum of positive ranks ($W^+ = 3401$) than negative ranks ($W^- = 1649$), indicating Set 1 outperforms Set 3 in magnitude more often. Similarly Figure 5.6c presents a histogram of the paired R^2_{test} differences, which is skewed toward positive values, favouring Set 1, with a significant p-value of 0.0026, surpassing the conventional significance threshold of 0.05 (Wilcoxon, 1945). This confirms that the better performance of Set 1 is unlikely to be due to chance, and accordingly, Set 1 was adopted for all subsequent analyses and modelling.

Table 5.5. Hyperparameter set up and performance of the ET model on the four predictor sets under investigation

Model Details	**Set 1**	**Set 2**	**Set 3**	**Set 4**
Predictor Selection:				
total_CaO%	✓	×	×	✓
total_oxides%	✓	×	×	✓
ash_C%	×	✓	✓	×
ash_F%	×	✓	✓	×
slag%	×	✓	✓	×
sample_type	×	✓	×	✓
The rest as listed in Table 5.3	✓	✓	✓	✓
Multicollinearity Analysis:				
Maximum r_s/VIF	+0.56/2.2	-0.58/4.1	+0.53/4.5	+0.56/2.8
Tuned Hyperparameters:				
`n_estimators`	15	15	15	15
`max_features`	8	8	9	9
`max_depth`	8	8	7	8
`min_samples_leaf`	3	3	3	3
`min_impurity_decrease`	0.03	0.02	0.02	0.05
`bootstrap`	✓	✓	✓	✓
Cross-Validation Scores:				
Mean R^2 ± Std. Dev.	0.93 ± 0.024	0.93 ± 0.032	0.92 ± 0.035	0.93 ± 0.026
Performance Scores:				
R^2_{train}	0.93	0.93	0.93	0.93
R^2_{test}	0.93	0.93	0.93	0.93
$\text{MAE}_{\text{train}}$	0.84	0.85	0.89	0.84
MAE_{test}	0.73	0.78	0.81	0.79
$\text{MSE}_{\text{train}}$	2.08	2.29	2.31	2.16
MSE_{test}	1.68	1.93	1.98	1.81
$\text{RMSE}_{\text{train}}$	1.44	1.51	1.52	1.47
$\text{RMSE}_{\text{test}}$	1.29	1.39	1.41	1.34

✓ indicates inclusion of a predictor.
× indicates exclusion of a predictor.



(a) (b) (c)

Figure 5.6. Comparison of predictor Sets 1 and 3 based on the performance of ET trained on bootstrap resamples showing (a) the distribution of R^2_{test} scores, (b) Wilcoxon signed-ranks for all non-zero paired differences ($\Delta = R^2_{\text{Set1}} - R^2_{\text{Set3}}$), and (c) histograms of paired R^2_{test} differences with a significant p-value of 0.0026 in favour of Set 1

5.6.2 Preliminary Evaluation and Comparison of Machine Learning Algorithms

A preliminary performance evaluation of a range of algorithms have been conducted based on the commonly used statistical tool in machine learning known as the k-fold cross-validation method. Its implementation is described by Hastie et al. (2009). Scikit-Learn's k-fold cross-validation (as described by Géron (2019)) was used in this evaluation. It is adopted when the performance of machine learning algorithms on a new dataset is not known. It helps determine the best candidate out of a group of selected algorithms, through a time-efficient and data-driven estimation of model performance before conducting full hyperparameter optimisation on the selected algorithm. This evaluation involved: (i) randomly dividing the training dataset into 10 distinct subsets, (ii) training the selected algorithms on nine of the subsets with *fixed hyper-parameters* as listed in Table 5.6, (iii) validating the trained algorithms on the remaining subset, (iv) calculating their R^2 scores, and (v) repeating steps (ii) to (iv) ten times, each time using a different subset for validation. Through k-fold cross-validation, the average R^2 scores and the standard deviation (σ) of R^2 scores for each algorithm were calculated. High average R^2 scores with low σ, in this case, indicated better generalisation performance and predictive stability, respectively. The entire dataset was used in this evaluation.

The performance evaluation included Linear Regression (LR), Least Absolute Shrinkage and Selection Operator (LASSO), Ridge Regression (RR), Elastic Net (EN), Support Vector Regression (SVR), K-Nearest Neighbours (KNN), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), AdaBoost (AB), Extra Trees (ET), Bagging Regressor (BR), Artificial Neural Networks (ANN), and Genetic Programming (GP). Linear and regularised regression models (LR, LASSO, RR, EN) yielded relatively weaker performance, reflecting their limited ability to capture the strong non-linearity inherent to alkali-activated soil systems. Distance-based and single-tree methods (SVR, KNN, DT) also underperformed, showing high variance and instability across folds. In contrast, TbE methods including Random Forest (RF), Gradient Boosting (GB), Bagging Regressor (BR), and particularly Extra Trees (ET), demonstrated higher accuracy and narrower variability in R^2 scores. Artificial Neural Networks (ANN) also performed strongly, confirming their capability to learn complex interactions in the dataset. Genetic Programming (GP) achieved similarly accuracy. Overall, TbE’s demonstrated greater potential for further optimisation and evaluation to identify the most effective TbE method for comparison with ANN and GP models.

Table 5.6. Algorithms and fixed hyperparameter settings used in the preliminary cross-validation screening

Category	**Method**	**Hyperparameters**	**Fixed Values**
Linear	LR	`fit_intercept`	True
Regularised Linear	LASSO	`alpha`, `max_iter`, `tol`	1.0; 1000; 0.0001
	RR	`alpha`, `max_iter`	1.0; None
	EN	`alpha`, `l1_ratio`, `max_iter`, `tol`	1.0; 0.5; 1000; 0.0001
Kernel / Instance-Based	SVR	`C`, `epsilon`, `gamma`, `kernel`	1.0; 0.1; ’scale’; ’rbf’
	KNN	`n_neighbors`, `weights`, `algorithm`	5; ’uniform’; ’auto’
Tree-Based Ensembles	DT	`criterion`, `max_depth`, `min_samples_split`, `min_samples_leaf`	’squared_error’; None; 2; 1
	RF	`n_estimators`, `max_features`, `max_depth`, `min_samples_split`, `min_samples_leaf`	50; 7; None; 2; 1
	GB	`n_estimators`, `learning_rate`, `max_depth`, `min_samples_split`, `min_samples_leaf`	50; 0.1; 3; 2; 1
	AB	`n_estimators`, `learning_rate`	50; 1.0
	ET	`n_estimators`, `max_features`, `max_depth`, `min_samples_split`, `min_samples_leaf`	50; 7; None; 2; 1
	BR	`n_estimators`, `max_samples`, `bootstrap`	10; 1.0; True
Neural Network	ANN	`hidden_layer_sizes`, `activation`, `solver`, `learning_rate_init`, `max_iter`	100; ’relu’; ’adam’; 0.001; 200
Symbolic Regression	GP	`population_size`, `generations`, `tournament_size`, `stopping_criteria`, `metric`, `parsimony_coefficient`, `function_set`	1500; 400; 10; 0.01; ’rmse’; 0.001; {sqrt, add, mul, sub, abs}

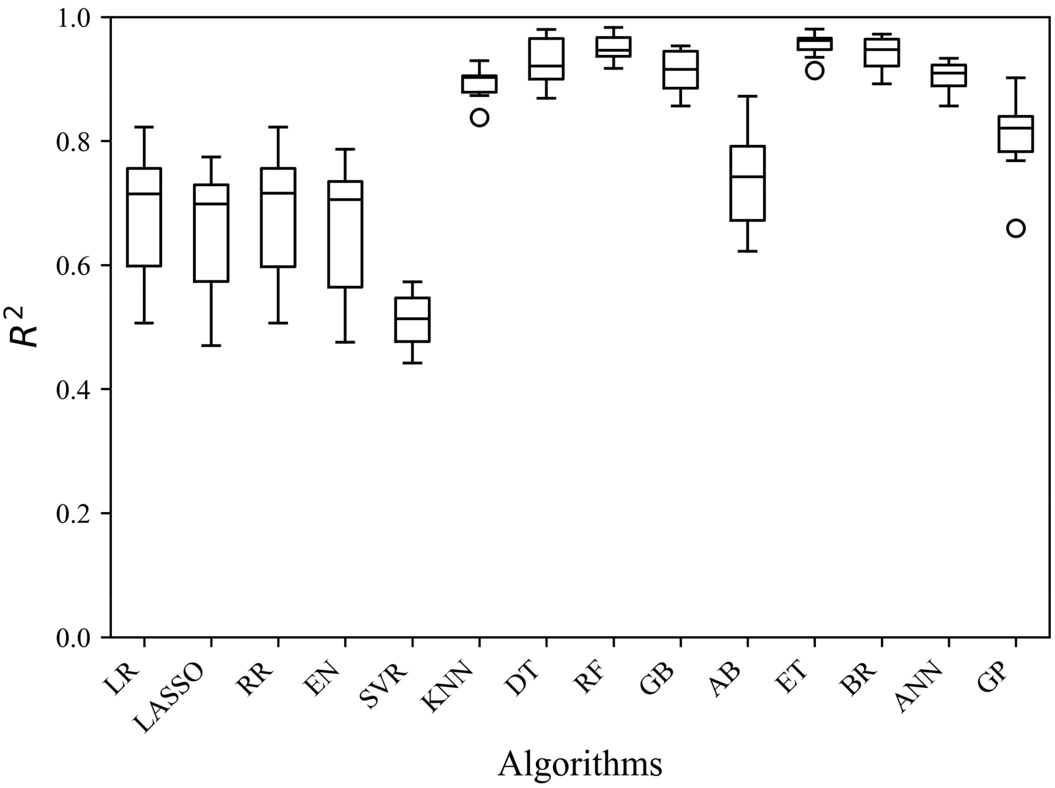


Figure 5.7. Preliminary performance evaluation of machine learning algorithms

5.6.3 Model Optimisation and Results

5.6.3.1 Tree-based Ensemble Methods

In the previous screening stage based on cross-validation, hyperparameters were fixed. Next, the performance of the TbE methods have been assessed by conducting hyperparameter optimisation. Table 5.7 presents the hyperparameter search spaces, their optimal values, and the CV scores along with their standard deviations of CV score for the assessed TbE methods. TbE methods including RF, GB, ET, and BR showed better performance by achieving CV scores of either 0.92 or 0.93 with reasonably small standard deviations (SD) of ≤ 0.024. AB exhibited the weakest performance by achieving a CV score of 0.77 with a SD of 0.050. The ET model achieved the highest score, tied with GB, but had the smallest SD of 0.017, suggesting marginally better performance consistency.

The training and testing performance metrics are presented in Table 5.8. ET stands out with the highest R^2 score of 0.93, both in training and testing, and relatively low error values on the testing phase. RF, GB, and BR also show strong performance, each achieving $R^2{}_{test}$ scores of 0.92, but with slightly higher MAE, MSE, and RMSE values on the testing phase. The AB models showed the weakest performance among the models by achieving R^2 scores of 0.81 and 0.82 for training and testing, respectively. These results show that the ET model, hereafter referred to as model ET-23, is a better candidate to be compared with the GP and ANN approaches.

Table 5.7. Methods used, hyperparameter search spaces and their optimised values, and CV scores for each method

Category	Methods	Hyperparameters	Hyperparameter Search Space	Optimal Values	CV Score
Boosting	GB	`n_estimators`	10 to 25	13	0.93 ± 0.028
		`max_features`	4, 6, 7, 8, 10	7	
		`learning_rate`	0.1 to 0.9	0.14	
		`max_depth`	5, 7, 10	7	
		`subsample`	0.1 to 0.7	0.3	
Boosting	AB	`n_estimators`	10 to 25	20	0.77 ± 0.050
		`learning_rate`	0.1 to 0.9	0.15	
Bagging	RF	`n_estimators`	10 to 25	13	0.92 ± 0.023
		`max_features`	4, 6, 7, 8, 10	7	
		`max_depth`	5, 6, 7, 8, 9, 10	7	
		`min_samples_split`	2, 4, 6, 9	9	
		`min_samples_leaf`	2, 3, 5	3	
Bagging	ET	`n_estimators`	10 to 25	15	0.93 ± 0.024
		`max_features`	4, 6, 7, 8, 10	8	
		`max_depth`	4, 6, 8, 10	8	
		`min_samples_leaf`	2, 3, 5	3	
		`min_impurity_decrease`	0.001 to 0.5	0.03	
Bagging	BR	`n_estimators`	10 to 25	15	0.92 ± 0.024
		`max_samples`	1, 0.5, 0.25	0.25	
		`bootstrap`	True, False	True	

Table 5.8. Performance scores and errors on the training and testing scores for ensemble methods

Methods	R^2 Scores		MAE		MSE		RMSE	
	Train	Test	Train	Test	Train	Test	Train	Test
GB	0.93	0.92	0.88	0.85	2.10	1.99	1.45	1.41
AB	0.81	0.82	1.55	1.37	5.90	4.79	2.43	2.19
RF	0.92	0.92	0.75	0.80	2.15	2.61	1.47	1.62
ET	0.93	0.93	0.84	0.73	2.08	1.68	1.44	1.29
BR	0.92	0.92	0.76	0.67	2.55	2.01	1.60	1.42

5.6.3.2 Genetic Programming Model

While all hyperparameters presented in Table 5.9 were important in model optimisation, population size, number of generations, and the choice of mathematical functions were particularly influential in the evolutionary process. An effective approach involved first identifying the most suitable combination of mathematical functions, followed by investigating the interactions between population size and the number of generations. Hyperparameter search space, the optimal hyperparameter values, and the CV score of the model GP-23 are presented in Table 5.9. Mathematical functions such as 'log' and '÷', were excluded because they could not effectively handle predictor values close to zero. Additionally, the '-ve' function was removed due to its tendency to produce negative values, which decreased performance stability.

Table 5.9. Hyperparameter search space and optimised hyperparameter values for the GP model

Hyperparameters	**Search Space**	**Optimal Values**
`population`: Number of individuals in each generation	100 to 2000	1500
`generation`: The number of cycles the algorithm runs	20 to 300	200
`tournament`: The number of models (individuals) randomly selected from the population to compete in a "tournament" for survival and reproduction	5, 10, 11, 12, 15	11
`stopping_criteria`: Conditions that terminate the evolutionary process before reaching the maximum number of generations	0.01	0.01
`fitness`: The performance measure used to evaluate and compare candidate models during the evolutionary process	MSE, RMSE	RMSE
`parsimony_coefficient`: To add a penalty to bigger expressions to stop them growing too large without real benefit, with higher values meaning stronger penalties.	0.001, 0.002, 0.05	0.001
`functions`: The set of metathetical functions (building blocks) used to construct polynomial terms in candidate models	$-$, $+$, $\times$, $\sqrt{}$, abs(), $-ve$, inverse	$+$, $\times$, inverse
Others	Default	Default

The optimised genetic expression tree generated by the evolutionary process is presented in Figure B.1 in Appendix B. To improve transparency and interpretability, the tree is decomposed into 4 sub-trees (ST1 to ST4), which are presented in Appendix B. These sub-trees are mathematically formulated in Equation B.2 to Equation B.5 (Appendix B), providing a structured breakdown of the complete GP model, expressed in Equation B.1. The variables in the sub-tree expressions are defined as follows: s: sand_to_fines, C: total_CaO%, O: total_oxides%, Na: NaOH%, M: M, N: NaSiO3%, W: water%, and t: t_curing (days). The evolutionary process dropped the predictor T_curing due to its minimal contribution to the predictive accuracy of the model. The model requires new inputs to be normalised using the statistics of its training dataset presented in Table 5.3.

5.6.3.3 Artificial Neural Network Model

The hyperparameters used, their search spaces, and the final tuned values for the ANN model are presented in Table 5.10. While complex interactions exist among all tunable hyperparameters, however, two interactions proved particularly influential and time-consuming to capture: (i) the interaction between the number of hidden layers and the number of nodes per hidden layer, and (ii) the interaction between learning rate and batch size. The optimal architectural set up of the ANN model is illustrated in Figure 5.9. The performance of the ANN model improved by increasing the number of hidden layers to two, containing 9 and 8 nodes respectively. Adding the third hidden layer did not yield any improvement.

Table 5.10. Hyperparameter search space and optimised hyperparamter values for the model ANN-23

Hyperparameters	**Search Space**	**Optimal Values**
No. of input nodes	9	9
No. of hidden layers	1, 2, 3	2
No. of nodes in hidden layer 1 (connection type = `dense`)	4 to 15	9
Activation function for layer 1 nodes	ReLU	ReLU
No. of nodes in hidden layer 2 (connection type = `dense`)	4 to 15	8
Activation function for layer 2 nodes	ReLU	ReLU
Output Layer with a single node (connection type = `dense`)	1	1
Activation function for the output node	`Linear`	`Linear`
Optimisation method (`solver`)	`Adagrad`	`Adagrad`
Metric used to calculate error (`loss`)	MAE, MSE	MSE
A specified rate (`learning_rate`) to allow weights and biases to be updated based on the calculated `loss`	0.01 to 0.3	0.18
No. of examples processed together before updating model parameters (`batch_size`)	1 to 10, 16, 32, 64	8
No. of times the entire training set was used to tune weights and biases (`epoch`)	0 to 1000	100

A useful feature of ANN-based models is their learning curves that enable real-time monitoring of the training effectiveness and checking for potential overfitting or underfitting. In well-optimised models, training curves for training and testing tended to converge and stabilise as training progresses, indicating strong generalisation to unseen data. As shown in Figure 5.8, the learning curves for model ANN-23 show that the validation set starts at a relatively low MSE (higher accuracy), while the training set starts at a much higher MSE, but quickly experiences a sharp reduction. Both training and validation sets converge and stabilise after about 70 epochs. The absence of any divergence between the two curves indicates that under and overfitting issues have been effectively mitigated.

ANN-based regressors are often thought of as *black boxes* due to lack of transparency relating model output to inputs and vice versa. However, the transparency of these relationships can be improved by representing the model in a complete mathematical form, as demonstrated in Appendix C.

5.6.4 Performance Evaluation & Comparison

Scatter plots comparing predicted and measured UCS values and probability distribution curves of residual errors for the models are presented in Figure 5.10. In the

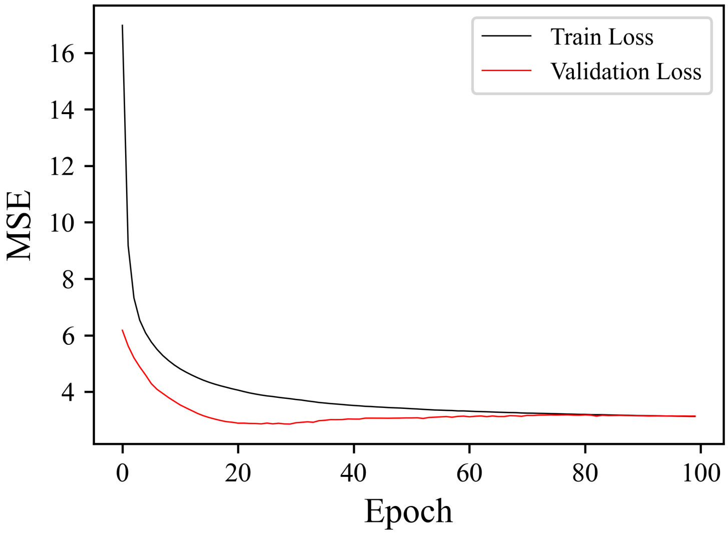


Figure 5.8. ANN-23 learning curve

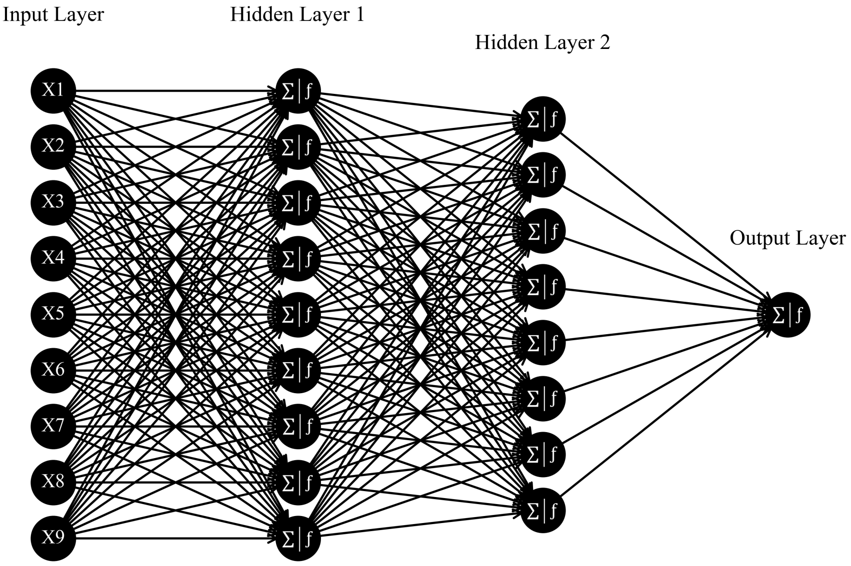


Figure 5.9. The optimal architectural setup of the ANN model consisting of 9 input nodes connected to the first hidden layer with 9 nodes, a second hidden layer with 8 nodes, with a single-node output

scatter plots, the predicted values closely align with the 45-degree reference line, indicating a reasonably strong correlation between measured and predicted UCS values across train, test, and holdout sets for models. However, some deviations can be observed in the holdout set scores, particularly in the case of the ANN model, which achieved the lowest R^2 score of 82%. The slightly reduced peaks in the testing distributions indicate a minor increase in error dispersion, but overall, the models demonstrate reasonably high accuracy and stability.

The residual error distributions for the train and tests sets are mostly narrow, generally symmetric, and centred around zero, indicating low bias and reasonably high predictive accuracy. However, the distributions for the holdout sets are relatively wider, especially in the case of the ANN model (Figure 5.10f), which shows a noticeable weakness in its generalisability.

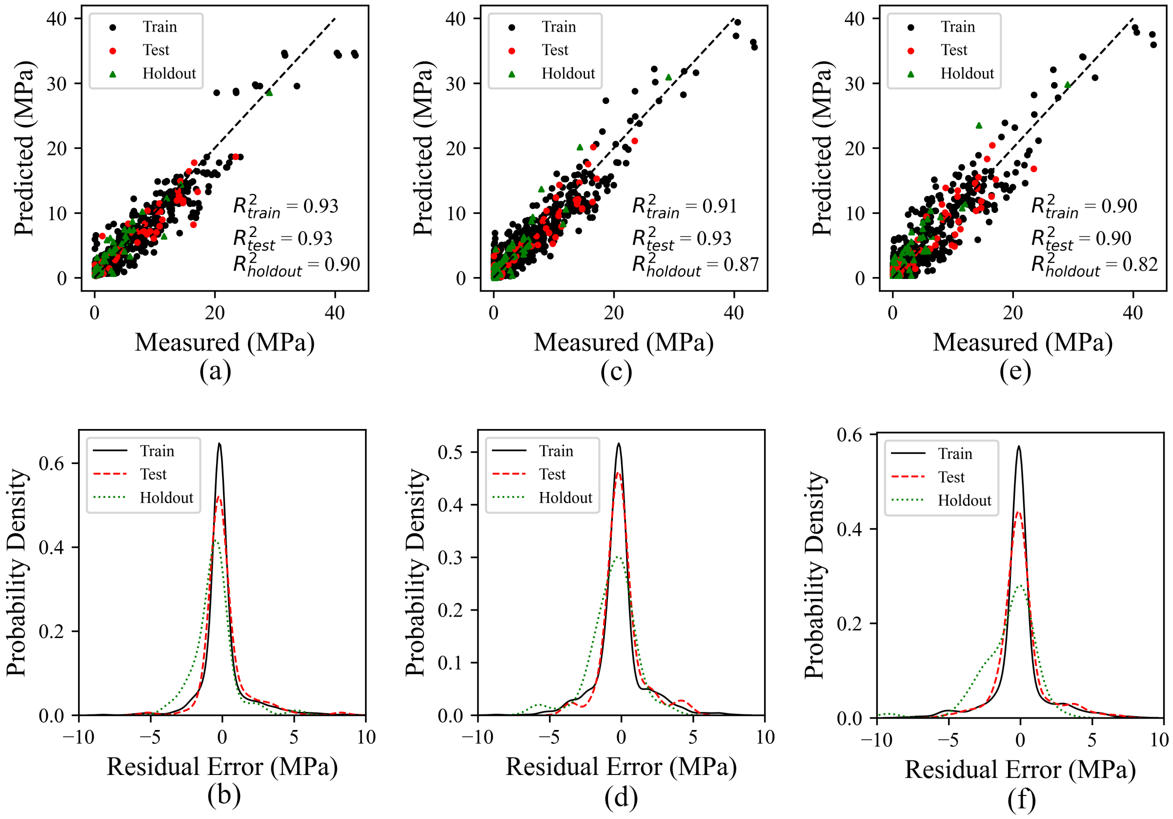


Figure 5.10. Scatter plots of the predicted versus measured UCS values and probability density curves of residual errors for model ET-23 (a & b), GP-23 (c & d), and ANN-23 (e & f) models

As summarised in Table 5.11, all three models achieve low MAE, MSE, and RMSE values, with slight variations in performance across the three models. Each model's error metrics remain stable across the training, testing, and holdout sets, indicating that none exhibits signs of overfitting or underfitting. ET-23 demonstrated the most consistently strong performance across all three models. It achieved the highest R^2 on training, testing, and CV (0.93), and a R^2 score of 0.90 on the holdout set. It also achieved the lowest MAE and RMSE on test data (0.76 and 1.39 respectively). A similarly strong performance was demonstrated by GP-23, matching ET-23's test R^2 (0.93) and outperforming it on training MAE, but recording a larger drop in holdout accuracy (R^2 = 0.87, RMSE = 1.67). On the other hand, ANN-23 achieved the lowest R^2 scores and the highest errors across training, testing, and holdout sets. Its holdout R^2 falls to 0.82, with the highest MAE (1.23) and RMSE (1.95). Therefore, the two better-performing models, ET-23 and GP-23, have been analysed and evaluated further in subsequent sections.

Table 5.11. Performance scores and errors on the training, testing, and holdout sets for models ET-23, GP-23, and ANN-23

	ET-23			**GP-23**			**ANN-23**		
Performance Metric	**Train**	**Test**	**Holdout**	**Train**	**Test**	**Holdout**	**Train**	**Test**	**Holdout**
R^2	0.93	0.93	0.90	0.91	0.93	0.87	0.90	0.90	0.82
MAE	0.81	0.76	1.01	0.99	0.85	1.12	0.94	0.90	1.23
MSE	2.14	1.94	2.12	2.59	1.87	2.80	3.04	2.52	3.80
RMSE	1.46	1.39	1.46	1.61	1.37	1.67	1.74	1.59	1.95
CV R^2 Scores	0.93 ± 0.024			0.92 ± 0.020			0.90 ± 0.014		

5.6.5 Model Uncertainty

The uncertainty of the three models was evaluated using bootstrap resampling technique. The distribution, mean, and 95% confidence intervals (CI) of the R^2 scores are shown in Figure 5.11. The ET-23 model achieved a mean R^2 of 0.94 (95% CI: 0.93–0.95) in the training phase (Figure 5.11a). On the test set (Figure 5.11b), ET-23 had a slightly lower mean R^2 of 0.92 (95% CI: 0.89–0.94), reflecting stable and consistent predictive performance. The small CI ranges suggest low variability in model performance for training and testing. For model GP-23 (Figure 5.11c & d), the training phase yielded a mean R^2 of 0.92 (95% CI: 0.91–0.94). The test phase achieved the same mean R^2 of 0.92 (95% CI: 0.90–0.95). As shown in Figure 5.11e and f, the widest CI's for both training and testing sets, and the smallest average mean R^2_{test} of 0.90 was achieved by the model ANN-23.

All distributions are roughly bell-shaped and approximately centred around their respective mean R^2 values (dashed lines). Curves are tightly distributed within the 95% CI represented by dotted lines. The training (Figure 5.11a, c, & e) and testing (Figure 5.11b, d, & f) distributions show narrow spreads, indicating that R^2 scores are consistently high across bootstrap resamples. All models demonstrated similar and reasonably low uncertainty, however, the ET and GP models demonstrated marginally better performance as shown by their narrower CI's across both training and testing sets compared to the ANN model.

A Wilcoxon signed-rank test on the paired R^2_{test} scores of the two better performing models, ET-23 and GP-23, produced a p-value of 0.53, indicating no significant difference between the two models. In Figure 5.12a, signed ranks for each bootstrap sample fall on both sides of zero, encompassing strong, moderate, and weak outcomes, without one model consistently outperforming the other. Figure 5.12b reinforces this by displaying a near-symmetrical distribution of score differences around the zero line (marked by the dashed vertical). Overall, these results confirm that ET-23 and GP-23 share comparable generalisation levels of performance.

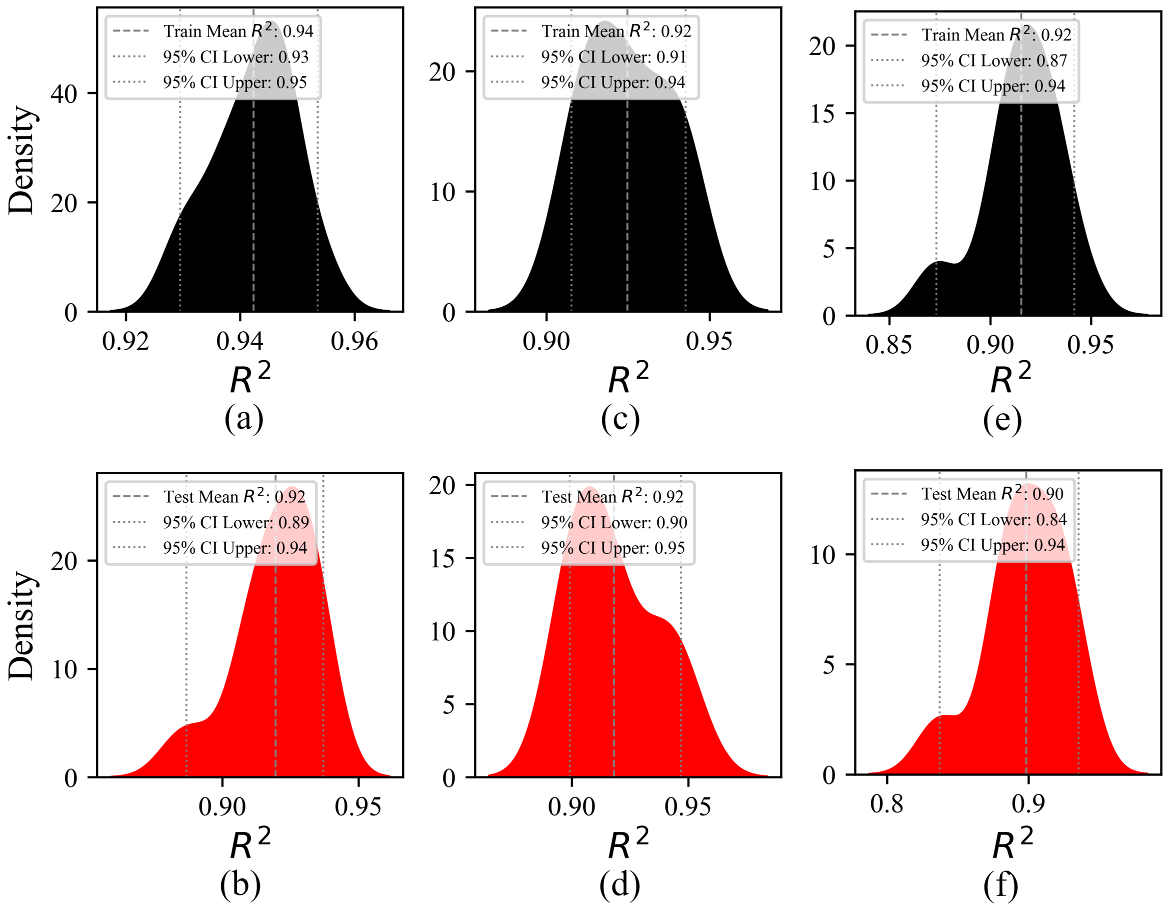


Figure 5.11. The distribution and mean of R^2 scores of the training and testing sets obtained through bootstrap resampling for ET-23 (a & b), GP-23 (c & d), and ANN-23 (e & f) models. The region within the dotted lines represents the 95% confidence intervals (CI)

5.6.6 Interpreting Model Performance

The SHAP summary plots for ET-23 (Figure 5.13a) and GP-23 (Figure 5.13b) present the predictors in descending order of importance along the y-axis, illustrating the relative contribution of each predictor to the predicted outcome. It highlights the differences in how each model is influenced by various predictors. In both models, *total_CaO*% emerges as the most influential predictor, exhibiting the highest absolute SHAP values, resulting in strong contribution to UCS estimation. Predictors *water*%, *t_curing*, and *total_oxides*% also play significant roles. However, their relative influence varies between the two models. The rest of the predictors have moderate to minimal influence on the two models. The strong influence of *total_CaO*% captured by the models aligns well with experimental findings as discussed below. The dependency plots for *total_CaO*% (Figure 5.14a & f) shows a strong positive quasi-linear impact on UCS, with SHAP values increasing steadily up to around *total_CaO*% of 18% in both models. Beyond this, the influence seems to taper off. Both models capture well the importance of CaO content, which is the key driver of fast-forming and strong C-A-S-H gels (Cristelo et al., 2012; Wardhono, 2018).

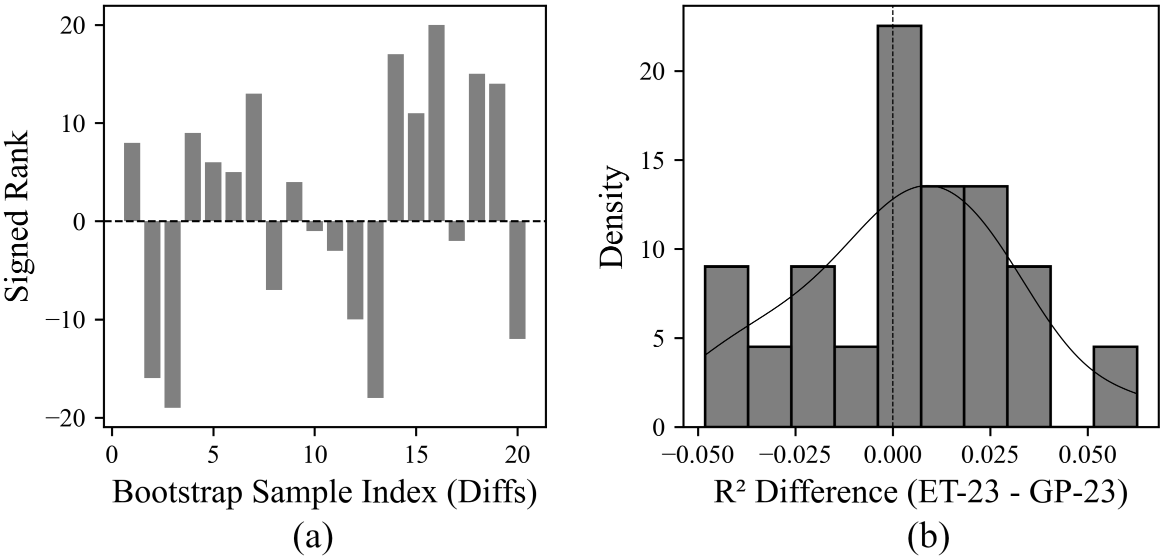


Figure 5.12. Wilcoxon signed-rank analysis of the paired R^2_{test} differences ($\Delta = R^2_{ET-23} - R^2_{GP-23}$), (a) Signed ranks for all non-zero differences, illustrating a roughly equal number of positive and negative values, (b) Histogram of the paired R^2_{test} differences, displaying an approximately symmetric distribution around zero and a non-significant p-value (0.53)

Dependency plot presented in Figure 5.14 (b & g) reveals a non-monotonic relationship between *water*% and UCS. Plots show that *water*% of less than about 20% yield positive SHAP values, indicating a positive influence on UCS, while higher water contents show a negative impact. This suggests the presence of an optimal moisture range, beyond which additional water reduces UCS. A similar trend was observed experimentally by Vu et al. (2020), who reported that UCS increased with water% up to a threshold of 32%, after which strength declined. Although the specific thresholds differ, likely due to differences in soil type, mix design, or curing conditions, the overall pattern of an optimal water range followed by strength loss is consistent. The vertical alignment of SHAP values at certain *water*% values (around 70%, 105%, and 140%) in Figure 5.14 (b & g) suggests that the effect of *water*% on UCS is influenced by interactions with other predictors, such as *ash*%, t_curing, or M.

The predictor $t\_curing(days)$ (Figure 5.14d & i) shows a generally positive influence on UCS, with longer curing times resulting in higher SHAP values, reflecting the progressive hardening and crystallisation of geopolymer gels. SHAP values increase rapidly up to approximately 28 days and continue to rise thereafter, albeit more gradually. Curing time is a critical predictor positively influencing UCS as it allows formation of more geopolymer gels with time as reported by (Hoy et al., 2016).

Predictor $sand\_to\_fines$ ratio is among the top four most influential predictors. Increased $sand\_to\_fines$ ratio reduces the required amount of alkaline solution and improves workability (Odeh and Al-Rkaby, 2022). As shown in Figure 5.14 (e &

j), it has a moderately positive impact, with increasing values improving UCS in ET-23 while its impact decreases in GP-23 (see Figure 5.13). SHAP values are widespread at both ends of the graph, particularly in the ET model (Figure 9 d), with a quasi-linear increase in between, generally indicating a positive influence on UCS of higher $sand\_to\_fines$ ratio, i.e., higher sand content. This is affirmed by experimental findings (Odeh and Al-Rkaby, 2022; Toufigh et al., 2022). The results also show that as the $sand\_to\_fines$ ratio approaches zero, SHAP values transition into the negative region, effectively capturing ET-23's ability to reflect the negative impact of reduced sand content on UCS as shown by Toufigh et al. (2022). This observation is also noticeable in the GP model, but it is less pronounced.

The predictor $total\_oxides$% is ranked as the second and fourth most important predictor in GP-23 and ET-23 models respectively. The dependency plots show a positive relationship between SHAP values and $total\_oxides$% up to about 25% $total\_oxides$%. Both models show that as $total\_oxides$% values exceed this threshold, SHAP values exhibit greater variability, with a trend towards negative SHAP values. The overall trend suggests an inverse relationship between $total\_oxides$% and the predicted UCS in both models beyond this threshold.

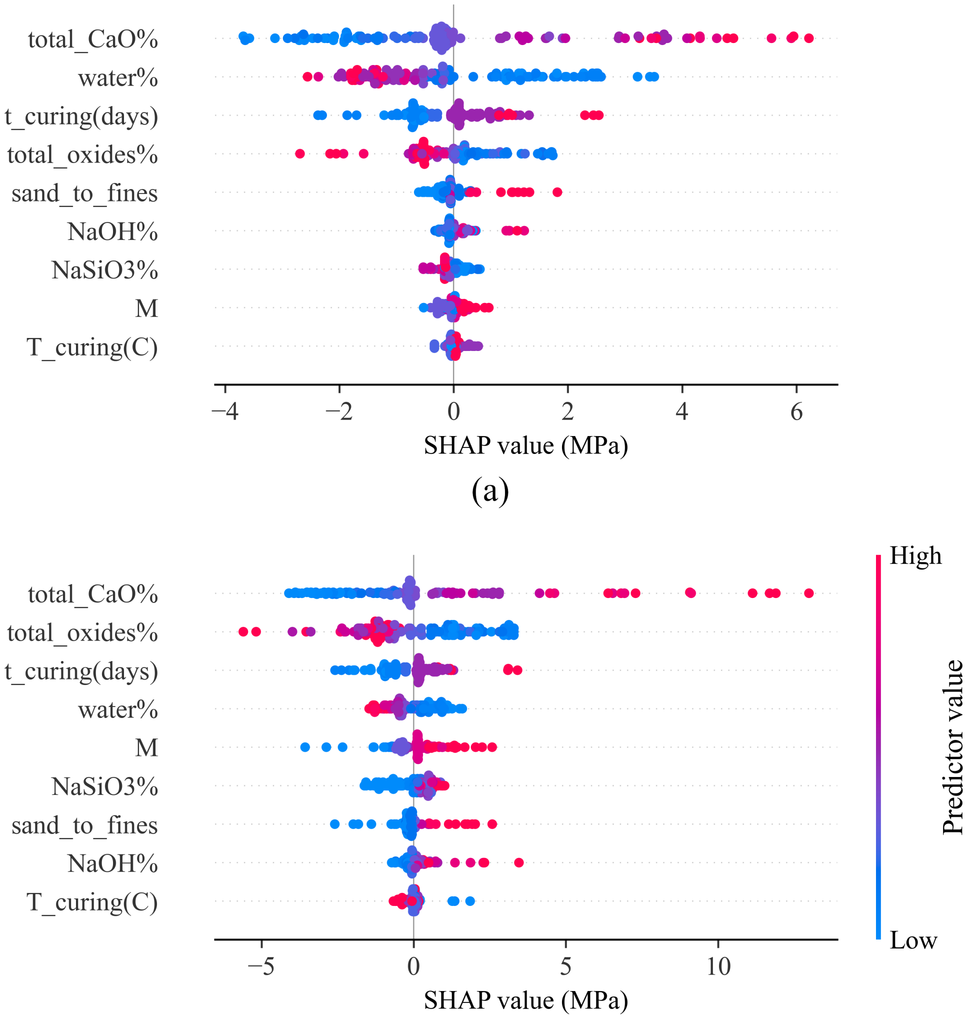
Other predictors including M, $NaOH$%, T_curing, and $Na2SiO3$% exhibit less pronounced effects. However, their contributions remain relevant and meaningful, as they influence reaction kinetics, geopolymerisation efficiency, and mix optimisation. For instance, molarity and curing temperature are critical in controlling the rate and extent of alkali activation, while $NaOH$% and $Na2SiO3$% contribute to the chemical activation process. These predictors, though lesser in importance, play supporting roles in achieving optimal soil stabilisation and strength development (Abdullah et al., 2021; Chen et al., 2022; Mypati and Saride, 2022; Odeh and Al-Rkaby, 2022; Suksiripattanapong et al., 2021).

The above findings indicate that, while predictors exhibit varying degrees of influence across the two models (as reflected in SHAP values on dependency plots), the overall nature of their influence remains consistent across the two models. This suggests that both models capture similar underlying relationships between predictors and UCS, despite differences in their internal architecture.

In the meta-analysis of previously published models, presented in Chapter 4), most predictor–response relationships were found to align with empirical evidence. However, the RM-Moz model showed a negative relationship between ash% and UCS, an outcome inconsistent with experimental findings and likely driven by the dominant effect of high-calcium slag in its training dataset. By contrast, the SHAP analysis of the newly developed models (ET-23 and GP-23) shows that the ash-derived predictor $total\_oxides$% has a positive influence on UCS up to approximately 25% (Figure 5.14c, h). While ash% and $total\_oxides$% are not identical predictors, they are closely related, as ash typically contains higher proportions of Fe_2O_3, SiO_2, and Al_2O_3 compared with clay or slag. This result is consistent with empirical findings that demonstrate a positive contribution of ash content to UCS up to an optimum of around 20%, beyond which strength commonly declines (Sukprasert et al., 2021). Accordingly, the new models address an inconsistency observed in RM-Moz by capturing more realistic and empirically consistent $ash\%$–UCS relationships.

The above discussion has several practical implications. It appears from the SHAP analysis that mixtures with high CaO content (Figure 5.14a & f) such as those combining high-CaO slag with low-CaO ash, can enhance early strength development. Such mixtures are well-suited for time-constrained construction projects where more rapid strength gain is preferred. In contrast, low-CaO mixtures require longer curing periods to achieve comparable strength gains. For such cases, suitable curing practices are crucial to maintain pH and moisture levels in the stabilised zones. These practices may include: (i) applying a low-permeability capping layer or chemically resistant and impermeable geomembrane to reduce groundwater ingress into the overlying stabilised layer, (ii) reducing moisture loss during curing phase by using appropriate covers, and (iii) installing groundwater monitoring wells to observe groundwater fluctuations during the curing phase. Surface water run-off and the risk of uplift force due to groundwater must be considered in the design stage.

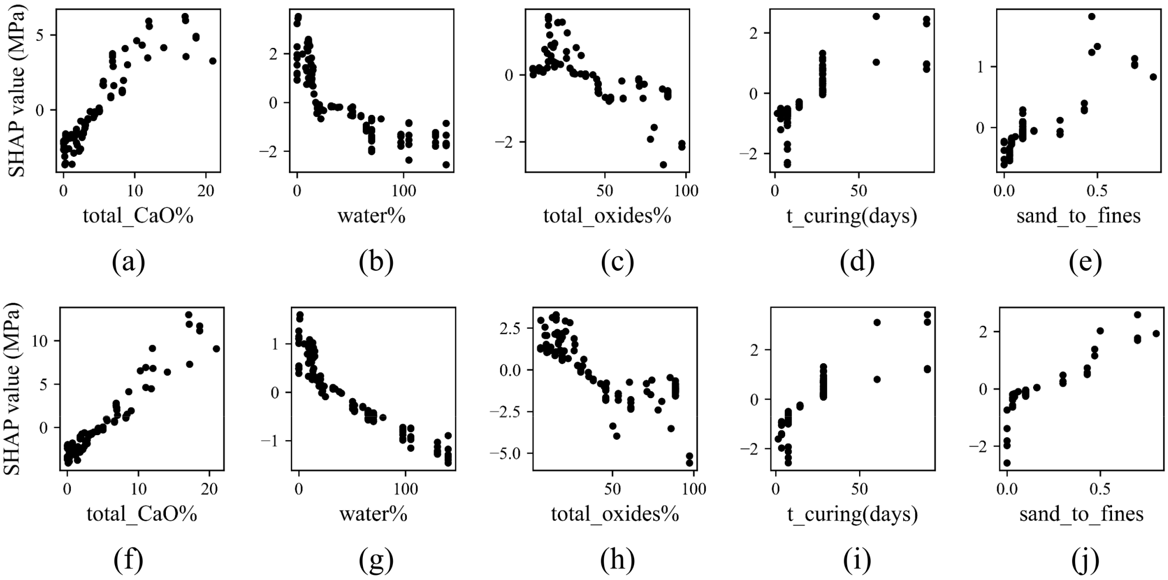
The above practices are also applicable to regions prone to high groundwater or heavy rainfall, where highly saturated soils are encountered, regardless of the type of aluminosilicate precursor used. Moisture control, during mixing and curing phases, is critical, as strength tends to decrease when water content exceeds approximately 20% (Figure 5.14b & g). This observation requires field moisture to be carefully assessed, and in high moisture conditions, pre-treatment measures such as air drying or blending with drier materials like sand may be considered prior to placement and compaction of the mixture. Adding sand increases the $sand\_to\_fines$ ratio in the prepared mixture, which enhances strength development as shown in Figure 5.14e and j.



(a)

(b)

Figure 5.13. SHAP summary plot showing the impact of predictors on the predicted UCS using the ET-23 (a) and GP-23 (b) models



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 5.14. Dependency plots showing the influence of the five important predictors on the ET-23 (a to e) and GP-23 (f to j) models based on SHAP values

Table 5.12 summarises the relative importance of each predictor in the ET-23 and GP-23 models based on their SHAP-derived rankings. Despite some variation in ranking, both models exhibit consistent qualitative trends, reinforcing that the underlying predictor–response relationships are robust across both modelling approaches. This agreement strengthens confidence in the interpretability and generalisability of the developed models.

Table 5.12. Predictor importance and consistency across ET-23 and GP-23 models, with model-specific rankings

Predictors	**Importance to *ET-23***	**Importance to *GP-23***	**Consistency Across Models**	**Interpretation**
$total\_CaO\%$	High (Rank 1)	High (Rank 1)	Strong	High CaO content promotes rapid formation of C–A–S–H and N–A–S–H gels, giving strong early strength gains.
$total\_oxides\%$	Medium–High (Rank 4)	High (Rank 2)	Strong, higher weight in GP-23	Controls aluminosilicate availability and improves geopolymerisation up to an optimum, beyond which UCS may decline.
$t\_curing(days)$	High (Rank 3)	High (Rank 3)	Strong	Longer curing enhances reaction kinetics, densification, and crystallisation of binding phases, increasing UCS.
$water\%$	High (Rank 2)	Medium (Rank 4)	Moderate (same trend, different rank)	Non-monotonic influence: total water contents of $< 20\%$ improves dissolution and workability; excess water weakens the soil structure.
$sand\_to\_fines$	Medium (Rank 5)	Medium–Low (Rank 7)	Moderate	Higher sand content improves packing and reduces water demand; low ratios indicate higher clay content and lower UCS.
M (molarity)	Low (Rank 8)	Medium (Rank 5)	Moderate, more influential in GP-23	Higher molarity enhances dissolution of aluminosilicates but excessive values may destabilise the mixture or reduce workability.
$NaSiO3\%$	Medium–Low (Rank 7)	Medium–Low (Rank 6)	Strong	Provides soluble silica that supports geopolymer gel formation but has limited effect beyond threshold levels.
$NaOH\%$	Medium–Low (Rank 6)	Low (Rank 8)	Moderate	Increases alkalinity and dissolution rate. Its influence is secondary compared with $total\_CaO\%$ and $total\_oxides\%$.
$T\_curing(C)$	Low (Rank 9)	Low (Rank 9)	Strong	Minor influence because the dataset spans a narrow temperature range. Higher curing temperatures mainly accelerate early reactions.

5.6.7 Comparison with Models in the Literature

Figure 5.15a and b compare the performance of ET-23 and GP-23 with other models reported in the literature. The comparison is based on the relationship between their R^2_{test} scores and (a) the number of experimental studies used for model development, and (b) the size of the datasets employed. The number of experimental studies is a critical factor, as models trained on limited studies capture less variability, restricting their applicability to specific soils, mix designs, sample preparation methods, and curing conditions. In contrast, datasets from multiple studies encompass greater variability, enabling broader applicability. Figure 5.15a shows that the number of studies per model ranges from 1 to 23. Excluding the new models, a general trend indicates that R²test scores tend to decline as the number of studies increases. The highest ${R^2}_{test}$ score (0.98) was achieved by RF-Zei, a Random Forest model trained on data from a single study, while the lowest score (0.80) was observed for ANN-Nis-1, an ANN model trained on 14 studies. This downward trend appears consistent across previous models, regardless of adopted modelling approach. In contrast, the models proposed in this thesis (ET-23 and GP-23) utilised data from the largest number of studies and achieved the same ${R^2}_{test}$ score of 0.93, reversing the observed trend.

Figure 5.15b illustrates the relationship between dataset size and ${R^2}_{test}$ scores. Three clusters of models can be seen. The highest performing models including RF-Zei, ANN-Leo, and GP-Leo achieve ${R^2}_{test}$ scores of greater than 0.90. In contrast, the lower performing models, i.e., ANN-Nis-1, ANN-Nis-2, and CR-Sar, which include shallow ANN-based models and a simple regression model, score between 0.8 to 0.85. Both groups were trained on fewer than 300 test records of data. Despite using larger datasets and a greater number of studies, the new models proposed in this thesis rank among the highest-performing models.

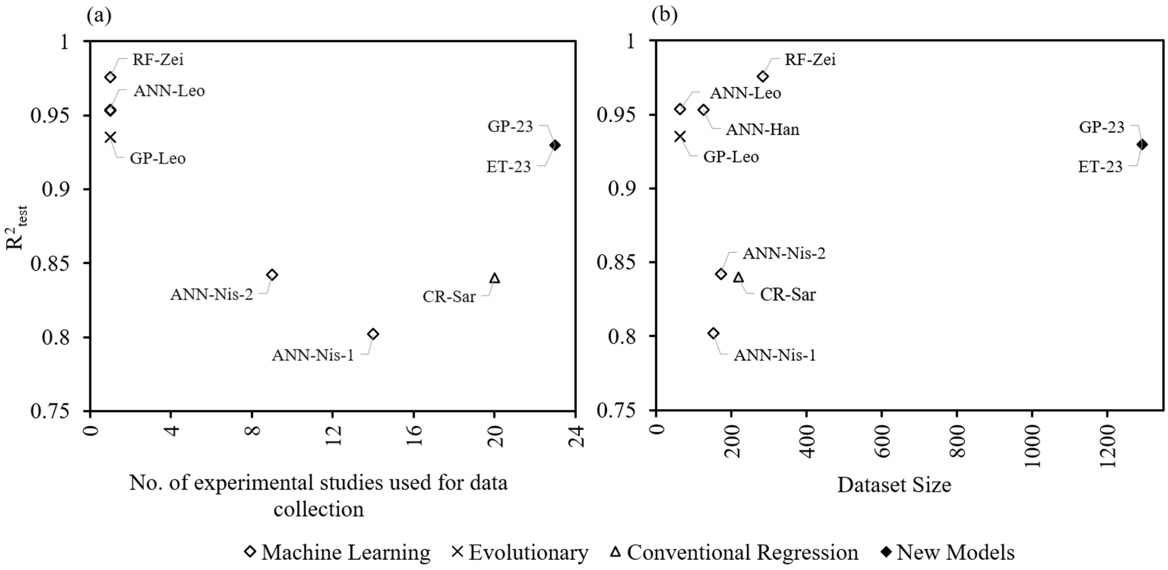


Figure 5.15. R^2_{test} of existing models and new models proposed in this thesis as a function of (a) The number of experimental studies and (b) dataset size

CHAPTER 6

EXPERIMENTAL ANALYSIS OF SWELLING STABILISATION

As illustrated in Figure 6.1, this chapter presents the experimental investigations undertaken to evaluate the swelling behaviour of soils stabilised with alkali-activated aluminosilicate by-products. The chapter begins with an assessment of key factors influencing swelling, including interaction effects and the performance of stabilisers across soils with different swelling potentials. Strength testing, microstructural analysis, durability evaluation under cyclic wetting and drying, and environmental assessment through leachability testing are then presented. Finally, the chapter synthesises these findings to propose a mechanistic understanding of swelling stabilisation. Together, these components provide a comprehensive experimental foundation for understanding and improving waste-based stabilisation techniques.

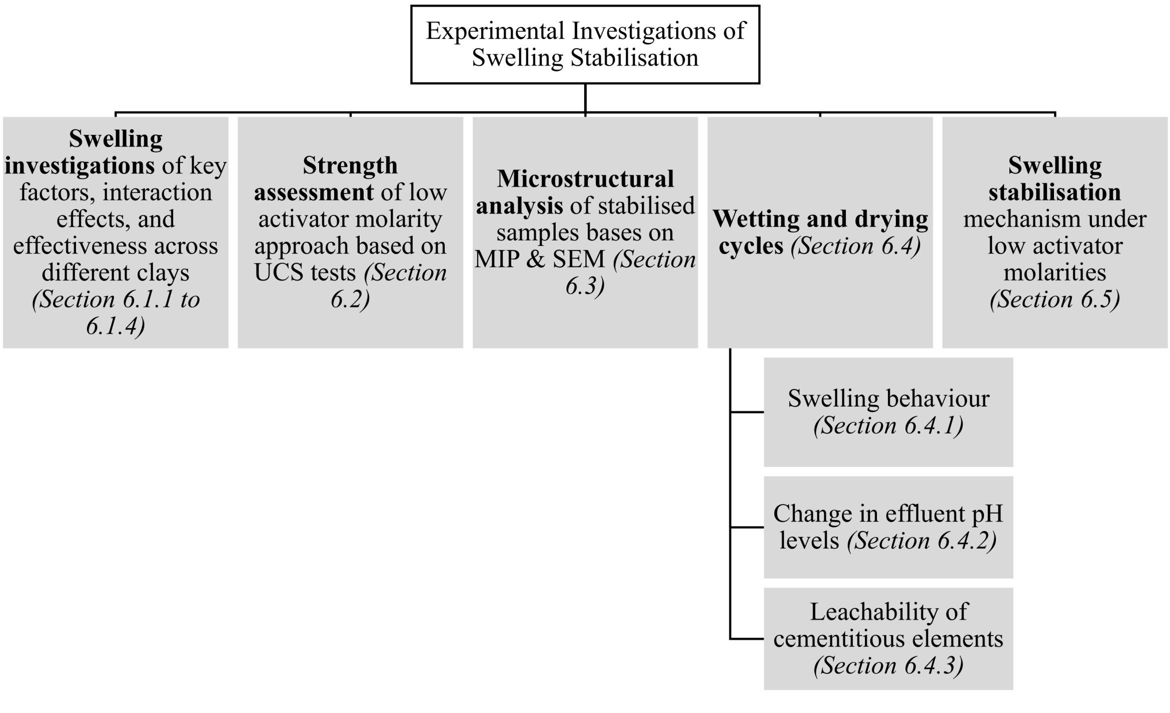


Figure 6.1. Overview of the structure and key components of Chapter 6

6.1 Swelling Tests

6.1.1 Set 1 Swell Tests: Baseline and Effect of NaOH

The swelling behaviour of three samples over a soaking period of over 80 days is presented in Figure 6.2. The untreated base soil (a) shows a rapid initial increase in swelling within the first 20 days, after which it levels off at around 25 to 26% after 30 days of soaking. The unstabilised base soil (b) also exhibits a rapid rise in swelling but stabilises at a lower swelling of around 15% after about 15 days. This decrease in swelling is likely due to the non-reactive ash and slag fractions replacing reactive base soil by the same mass. Finally, the stabilised base soil (c) demonstrates a significant reduction in swelling, stabilising at around 3% after just about 2 days of soaking, demonstrating the key role NaOH in the stabilisation of swelling.

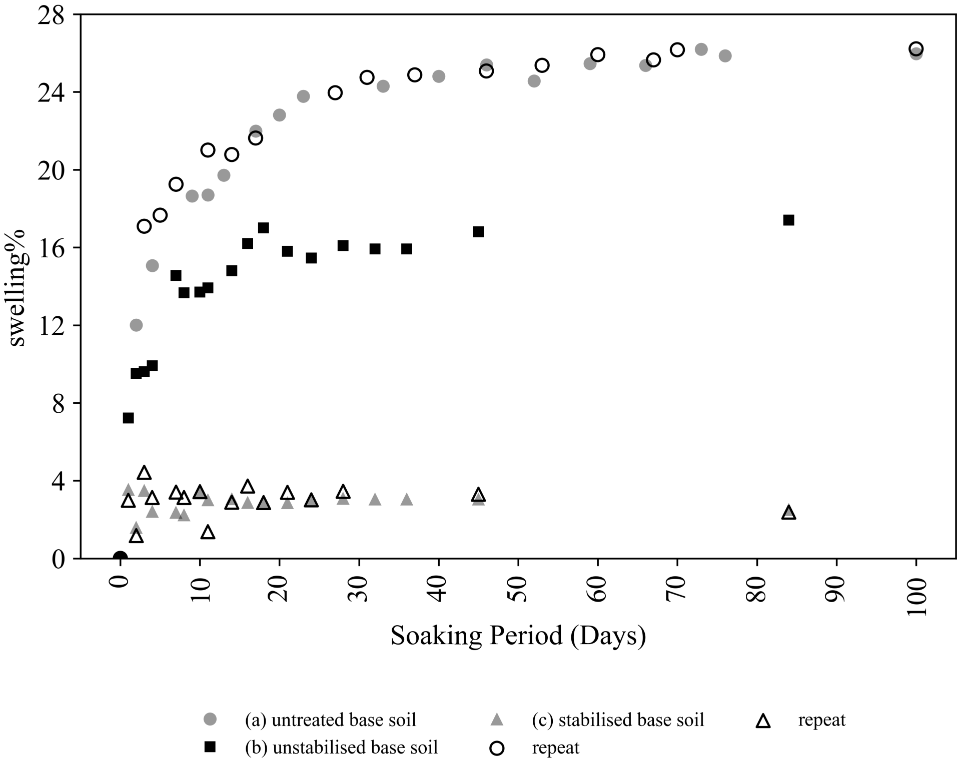


Figure 6.2. The swelling behaviour of the the untreated (a), unstabilised (b), and stabilised (c) base soil showing the important role of ALC in stabilising swelling

6.1.2 Set 2 Swell Tests: Effects of Binders, Activator Content, and NaOH Molarity

A two-level factorial experiment was designed to investigate the effects of variable contents at two specified levels: lower (L) and upper (U) limits. This approach allows for the systematic evaluation of the influence of each variable, providing insights into how changes in these levels impact the outcome. L and U limit values were determined based on a survey of relevant UCS studies as shown in Table 3.7. The content of one variable was varied per specimen while keeping the other factors constant. Specimens were cured for 30 days. Tests were repeated and the average results are presented below.

Test results presented in Figure 6.3 show the impact of increasing the contents or concentrations of ash%, slag%, ALC%, and M, from their lower to upper limits, on swelling%. Each point represents the measured swelling% when the variable is increased to its upper limit (see Table 3.7 in methods section). The horizontal dashed line represents the baseline sample where all variables were kept at their lower limits. The baseline is used as a reference for assessing the effect of each variable on swelling. Deviation of points above or below the baseline indicates a positive or negative influence on swelling, respectively. The results indicate that increasing the ash% and M to their upper limits results in swelling values near the baseline level. In contrast, ALC% exhibits the largest downward deviation from the baseline when increased, indicating it has the most significant impact in reducing swelling% compared to the other variables. Slag% also reduces swelling but to a lesser extent than ALC%. These findings stress the importance of ALC% as the most effective factor in minimising swelling.

The effect of NaOH molarity did not meet initial expectations based on UCS studies. Findings in this thesis indicate that the swelling% of samples activated with a 3 M alkaline solution are nearly identical to those activated with 14 M solutions. While such increases in molarity have been shown to significantly enhance strength development (Toufigh et al., 2022), their impact on swelling behaviour appears to be far less pronounced, suggesting that the mechanisms underlying swelling stabilisation differ from those influencing strength. This discrepancy highlights the need for further investigation into the role of molarity in swelling stabilisation and its interaction with curing time, a task undertaken in set 3 swell tests.

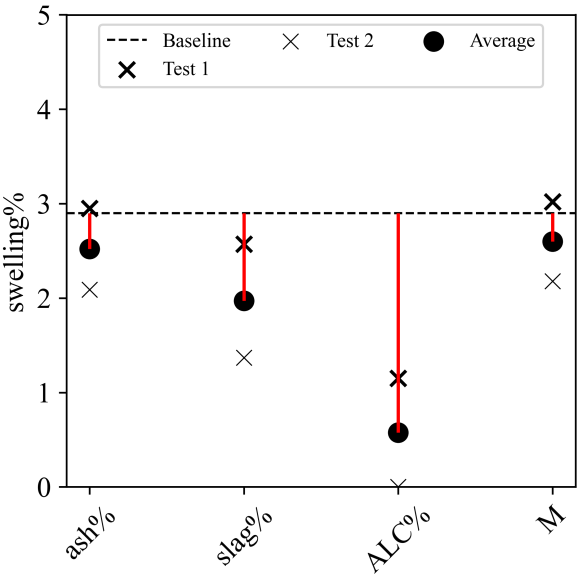


Figure 6.3. The effects of increasing variable values from lower (L) to upper (U) limit values on swelling after 30 days of curing

6.1.3 Set 3: Interaction Effects

Next, the interaction between M and t_curing was investigated. A total of 21 samples were prepared and tested. Ash%, slag%, and ALC% were kept constant at their lower limit values, while varying t_curing and M. The curing times considered were 1, 5, 8, and 11 days, and molarities considered were 0.5, 0.75, 1, 1.25, 1.5, and 3 M.

As shown in Figure 6.4, specimens cured for one day at lower molarities (0.5 M and 0.75 M) exhibited significantly higher swelling (about 14% and 7%, respectively) while those at higher molarities showed less than 3% swelling. Swelling decreased with curing time across all specimens, though the rate of reduction varied by molarity. Notably, specimens stabilised with 0.5 M still exceeded 10% swelling after 11 days, whereas those with above 0.5 M converged to approximately 1% or less beyond eight days. These results highlight that early-stage swelling is strongly influenced by molarity, but prolonged curing reduces variability. The best outcome was observed in 1 M samples cured for one day, which is substantially lower than the molarity and curing time typically required for strength development (Cristelo et al., 2013; Odeh and Al-Rkaby, 2022). Given the high cost of NaOH, this presents considerable financial advantages, as well as a reduced environmental footprint. Additionally, the shorter curing time is particularly valuable for projects operating under tight construction schedules.

The results shown in Figure 6.4 suggests that 1 M molarity is a reasonable NaOH concentration for the stabilisation of swelling. Therefore, molarity was kept constant at 1 M hereafter in this study. Also, slag% was kept constant at 10%, which is what the optimum slag% has been reported when it is used in combination with

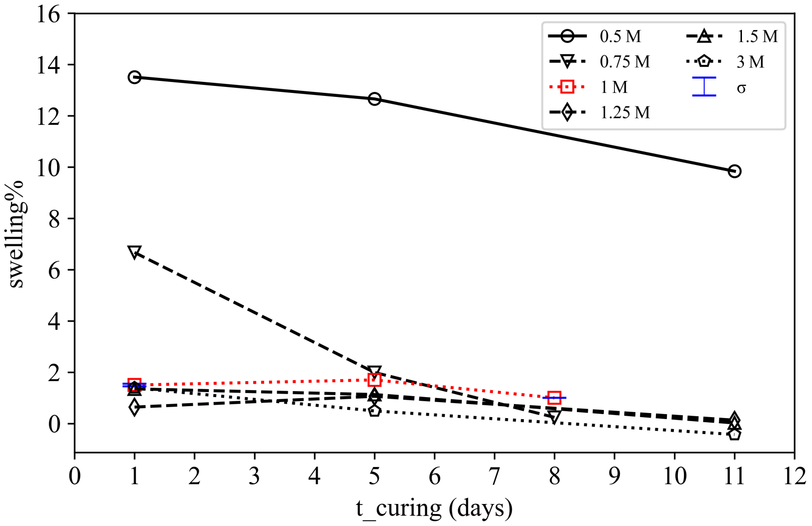


Figure 6.4. Interactions between molarity and curing time and their effects on swelling

ash (Sukprasert et al., 2021). Next, a two-level factorial experiment was designed to study the interaction effects between ash%, ALC%, and t_curing. The upper and lower levels were set as follows: for ash%, 10% and 30%; for ALC%, 15% and 25%; and for t_curing, 0.25 days and 7 days. The baseline specimen was prepared with additives set at their lower limit values.

Results shown in Figure 6.5 demonstrate the effects of variable pairs on swelling%, including: (a) ash% vs ALC%, (b) ALC% vs curing time, and (c) ash% vs curing time. Each plot compares swelling% outcomes across combinations of upper and lower variable limits against a baseline (dashed line). In plot (a), swelling is primarily influenced by ALC%. At 25% ALC, swelling drops well below the baseline regardless of ash%, as seen in the first two samples. However, at 15% ALC, swelling remains near the baseline and varies slightly with ash%, with higher ash% increasing swelling by about 2% (e.g., first vs third sample). Plot (b) reinforces ALC% as the dominant factor with higher ALC% consistently leading to significant swelling reduction, regardless of curing time (first vs third sample). Plot (c) further highlights the influence of ash%, with higher ash% slightly increasing swelling above baseline (third sample), though longer curing time can mitigate this effect as shown by the first sample in plot (c). Overall, the results confirm the dominant role of ALC% in controlling swelling, while ash% and curing time play secondary, but still influential, roles.

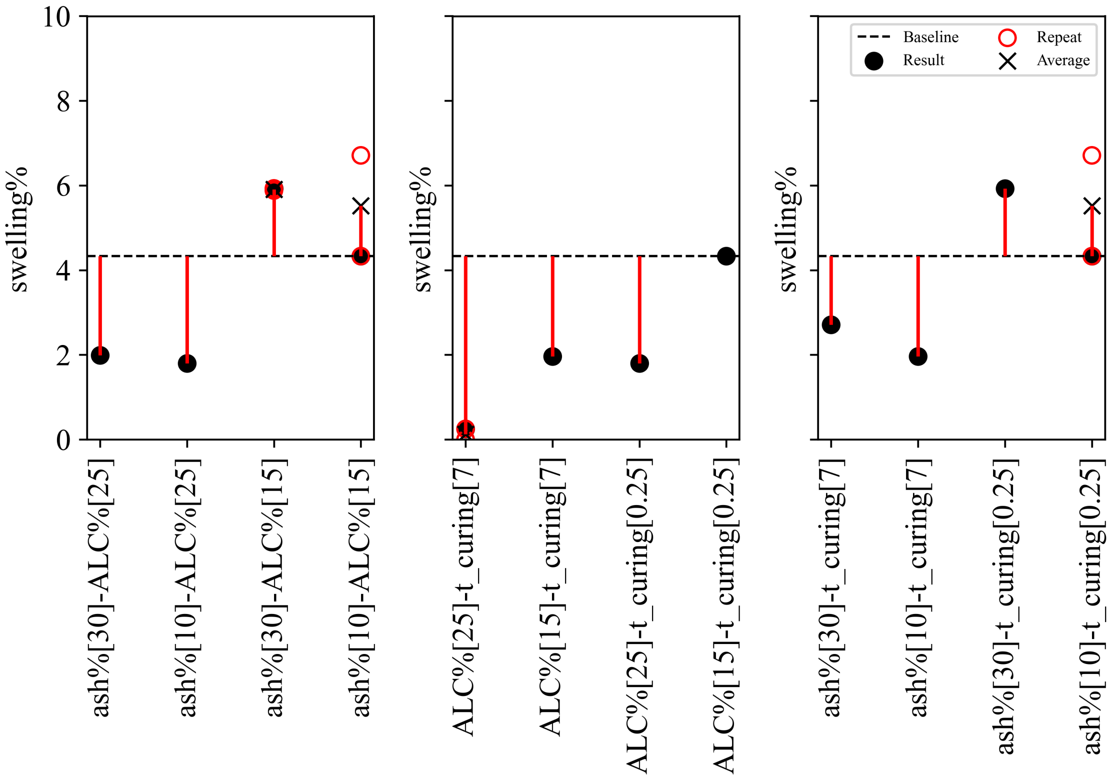


Figure 6.5. The interaction effects between variable pairs: ash% and ALC%, ALC% and t_curing, and ash% and t_curing

6.1.4 Set 4: Effect of Stabilisation on Other Soils

The experiments up to this point were essentially screening assessments focusing exclusively on the base soil to identify key variables and their interactions. Building on the above findings, further laboratory investigations were conducted to explore the effectiveness of stabilisation on soils with five different soil types as presented in Table 3.5. First, the swelling profiles of the untreated soils were assessed. As shown in Figure 6.6, the samples base_soil, amended_1, and amended_2 reached their maximum swelling after approximately 30 days. In contrast, the swelling of amended_3 and amended_4 required about 70 days. The results confirm the presence of a correlation between swelling potential and increasing plasticity index (PI) values. As it was expected, amended_1 exhibited the lowest swelling, approximately 24%, followed by base soil, amended_2, and amended_3. The highest swelling, around 55%, was observed in amended_4.

To assess the effect of stabilisation on soils with varying swelling potentials, the swelling behaviour of the base_soil and amended_1 to amended_4 soils was examined using different ash contents (10%, 20%, 30%) and ALC levels (10%, 15%, 20%, 25%), while keeping slag content (10%), NaOH molarity (1 M), and curing time (7 days) constant. Results are presented in Figure 6.7. The dashed horizontal line in each subplot indicates the swelling% of the untreated soils, serving as a baseline for comparison. Across all soils, a consistent trend is observed: increasing ALC% leads

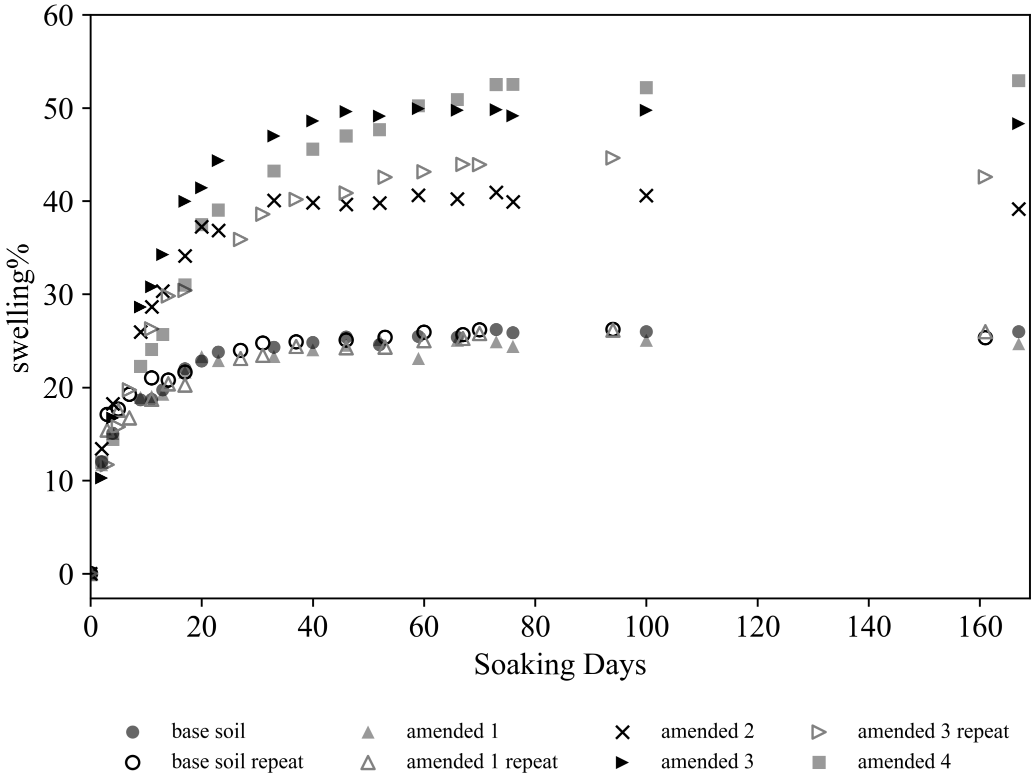


Figure 6.6. Swelling of the five untreated clay types used in this experiment

to a reduction in swelling. This effect is most significant at ALC levels of 20% and 25%, which reduce swelling to near-zero across all ash contents analysed. In contrast, an ALC of 10% yields the smallest improvement. Among the combinations tested, 20% ash paired with 20% ALC provides the most consistently high level of swelling reduction.

Another interesting observation is that stabilisation appears more effective in soils with higher initial swelling potential. This is evident from the widening gap between the dashed baseline and the stabilised lines. The greater effectiveness may be attributed to the higher clay content in these soils, which likely increases the availability of reactive elements such as Al, Si, and Na, hence improving the extent of geopolymerisation and resulting in reduced swelling in the stabilised soils.

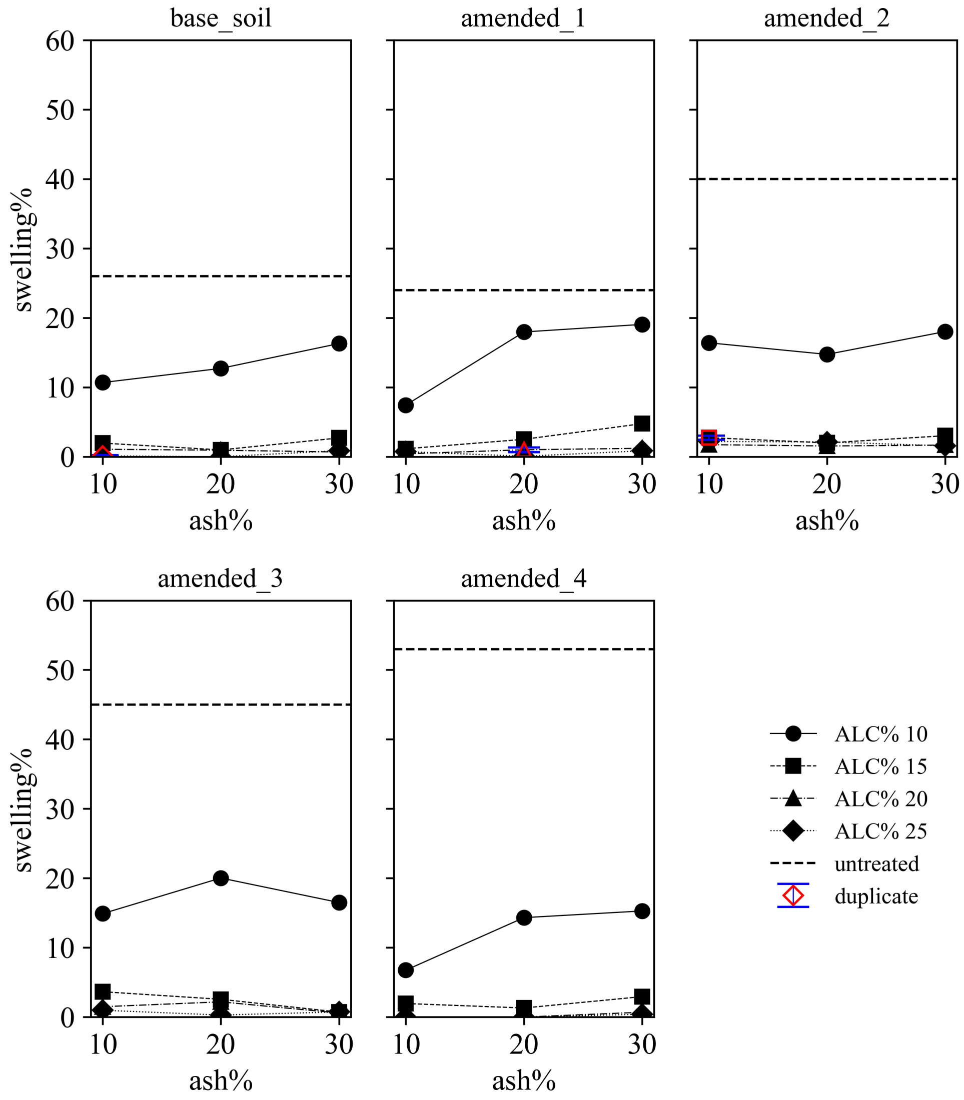


Figure 6.7. The effect of stabilisation on different soil types

6.2 Strength Development

Two batches of samples were prepared for this analysis, each focusing on a different objective. As shown in Table 6.1, batch 1 aimed to assess the impact of initDD, ash%, slag%, ALC%, M, and t_curing on UCS. Batch 2 samples were soaked in water after curing to investigate the effect of moisture exposure on the strength of stabilised samples. This was to mimic the effect of wet seasons on the stabilised soils.

Table 6.1. Experimental factors, factor types, factor levels, and values

Batch No.	initDD (g/mm^3)	ash%	slag%	ALC%[1]	M	t_curing (days)	Soaking period after curing
1	0.00128 to 0.00140	0, 10, 20	0, 10	15, 20	0, 1	1, 7, 14, 28	NA
2	0.0014	0, 20	0, 10	15, 20, 25	0, 1	7	10 days

[1] For unstabilised samples, distilled water was added instead of ALC.

Batch 1 results are shown in Figure 6.8. Most variables, including initDD, ash%, slag%, ALC%, and M have clearly positive influence on UCS except for t_curing (days), which does not exhibit a clear relationship possibly due to interaction effects with other variables, particularly M and initDD. A higher M would make the relationship more pronounced.

The highest strength gain was achieved by samples containing 10% slag, 10% ash, and activated with 20% ALC, showing a substantial improvement over both the untreated clay and the mixture activated with 15% ALC. The stress–strain responses in Figure 6.9 illustrate the significant enhancement in stiffness and peak strength achieved with increased ALC%, using a constant mix of 10% ash and 10% slag. Stabilised samples exhibit a ductile stress–strain behaviour.

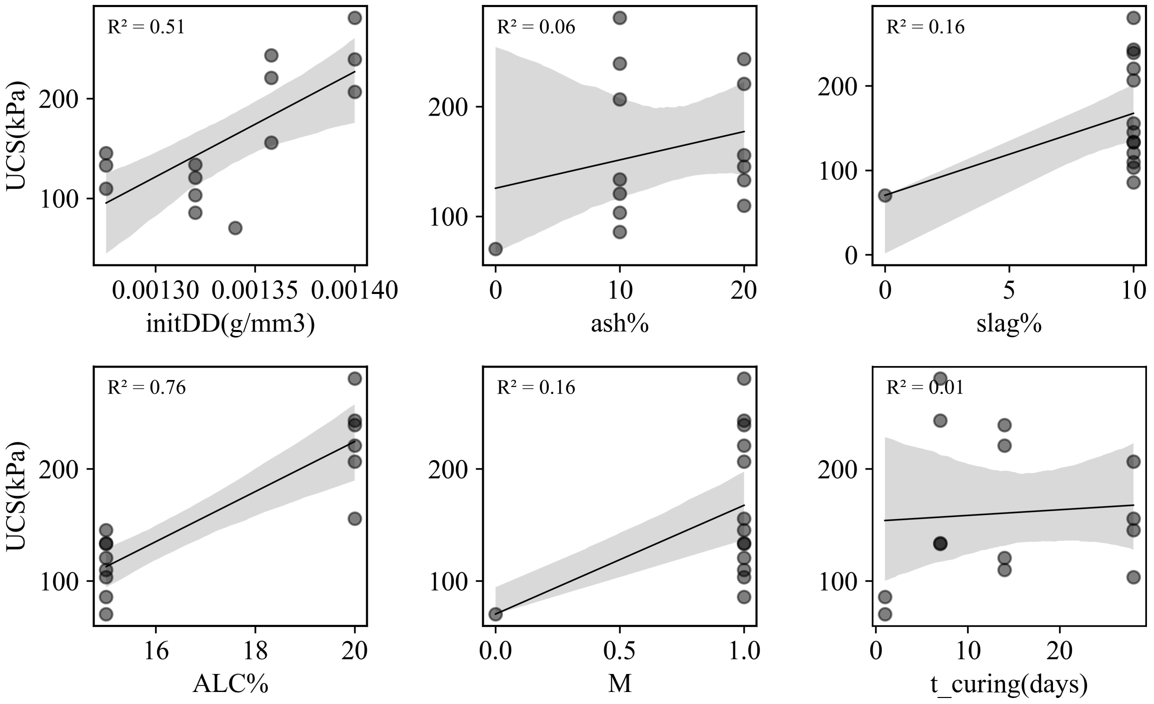


Figure 6.8. Results of batch 1 tests showing scatter plots of variables versus UCS

Batch 2 test results are presented in Figure 6.10. The final moisture content values along the x-axis represents the moisture content at the end of 10 days of soaking after curing. Results reveal a significant difference in Su values between stabilised and unstabilised samples. Stabilised samples exhibited a significant increase in strength from 22 kPa to approximately 160 kPa as ALC% increased from 15% to 25%. On the other hand, the unstabilised samples show near-identical strength values when its initial moisture content (MC_0) is increased from 15 to 25%.

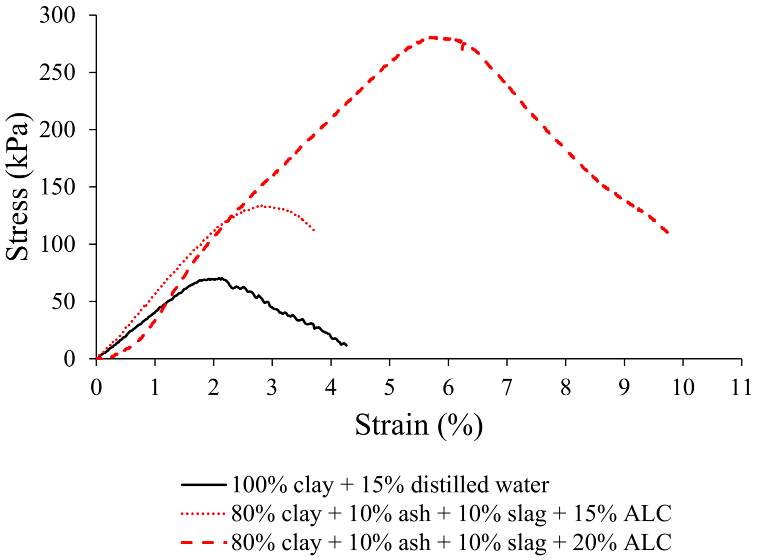


Figure 6.9. Stress-strain curves for untreated clay and stabilised mixtures containing 10% ash and 10% slag, activated with 15% and 20% ALC

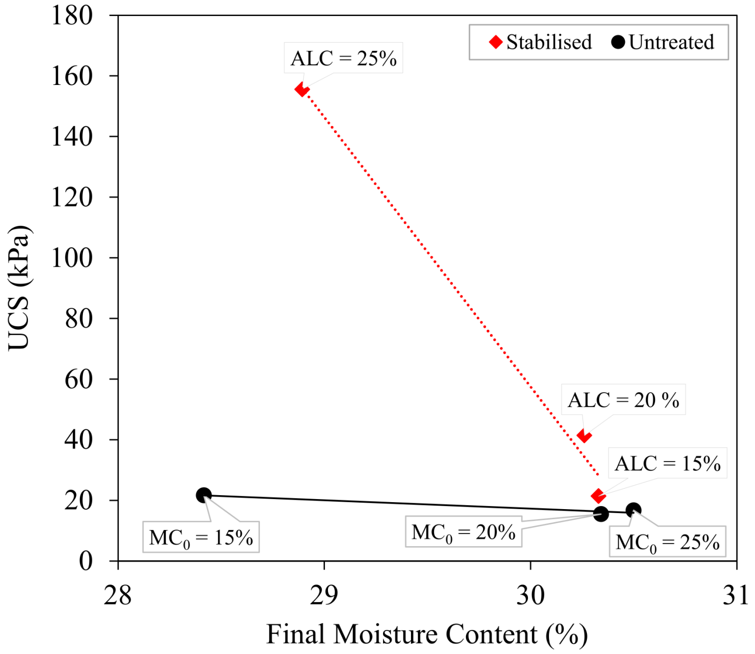


Figure 6.10. Results of the second batch of strength tests showing a significant increase in strength in the stabilised samples as ALC% increases from 15% to 25%, while the unstabilised samples exhibit near-identical strength values

6.3 Microstructural Analysis

The changes in the microstructural porosity of the stabilised samples were analysed through MIP tests. Figure 6.11 presents the Cumulative Intrusion (CI) in sub-plots (a) and (c) and Log Differential Intrusion (LDI) in sub-plots (b) & (d). Repeat tests showed some variation in CI, with maximum values ranging from 0.085 to 0.107 mL/g for the unstabilised samples (a), and 0.019 to 0.035 mL/g for the stabilised samples (c). Both stabilised and unstabilised samples exhibited minimal intrusion between pore sizes of about 300 to 20 µm, as indicated by the gentle CI slopes. In contrast, a sharper increase in intrusion was observed between 20 and 6 µm, after which the CI plateaued.

The LDI plots in Figure 6.11 (b) and (d) show that the pore size distribution in both stabilised and unstabilised samples spans approximately 6 to 100 µm. However, notable differences exist in the location and intensity of dominant pores among the repeat samples. Most unstabilised samples exhibit dominant peaks in the 6 to 10 µm range, with occasional peaks between 20 to 30 µm (Figure 6.11 (b)). In contrast, the stabilised samples display more scattered peaks across 6 to 30 µm as shown in Figure 6.11 (d). Both sample types show minor peaks near 100 µm. A key distinction is the significant reduction in peak LDI values for the stabilised samples, indicating a decrease in dominant pore volume and overall porosity following stabilisation.

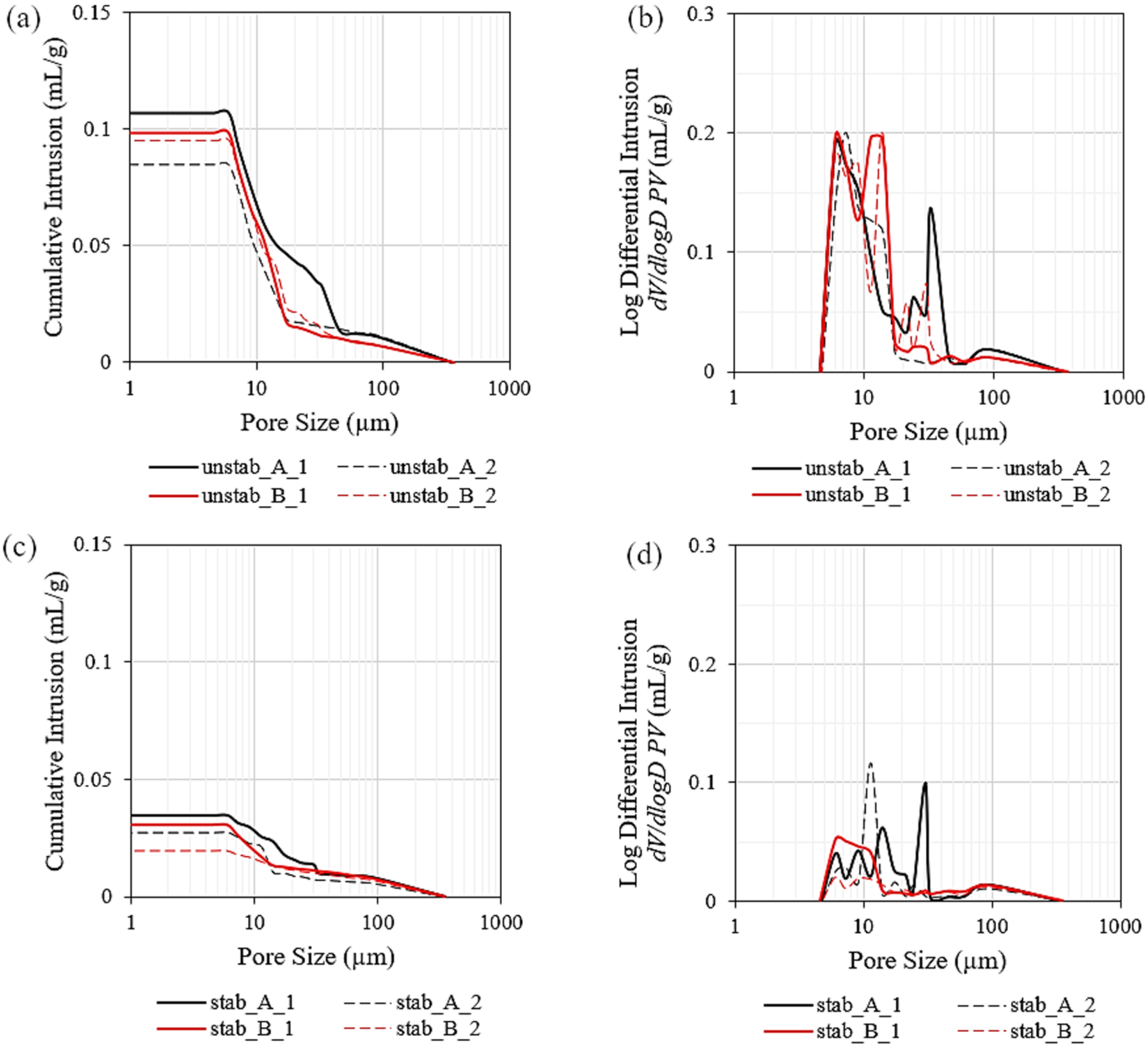


Figure 6.11. MIP results for the unstabilised (a & b) and stabilised samples (c & d); two specimens (original specimen A and repeat specimen B) were taken from each of the two samples (original sample indicated by 1 and repeat indicated by 2)

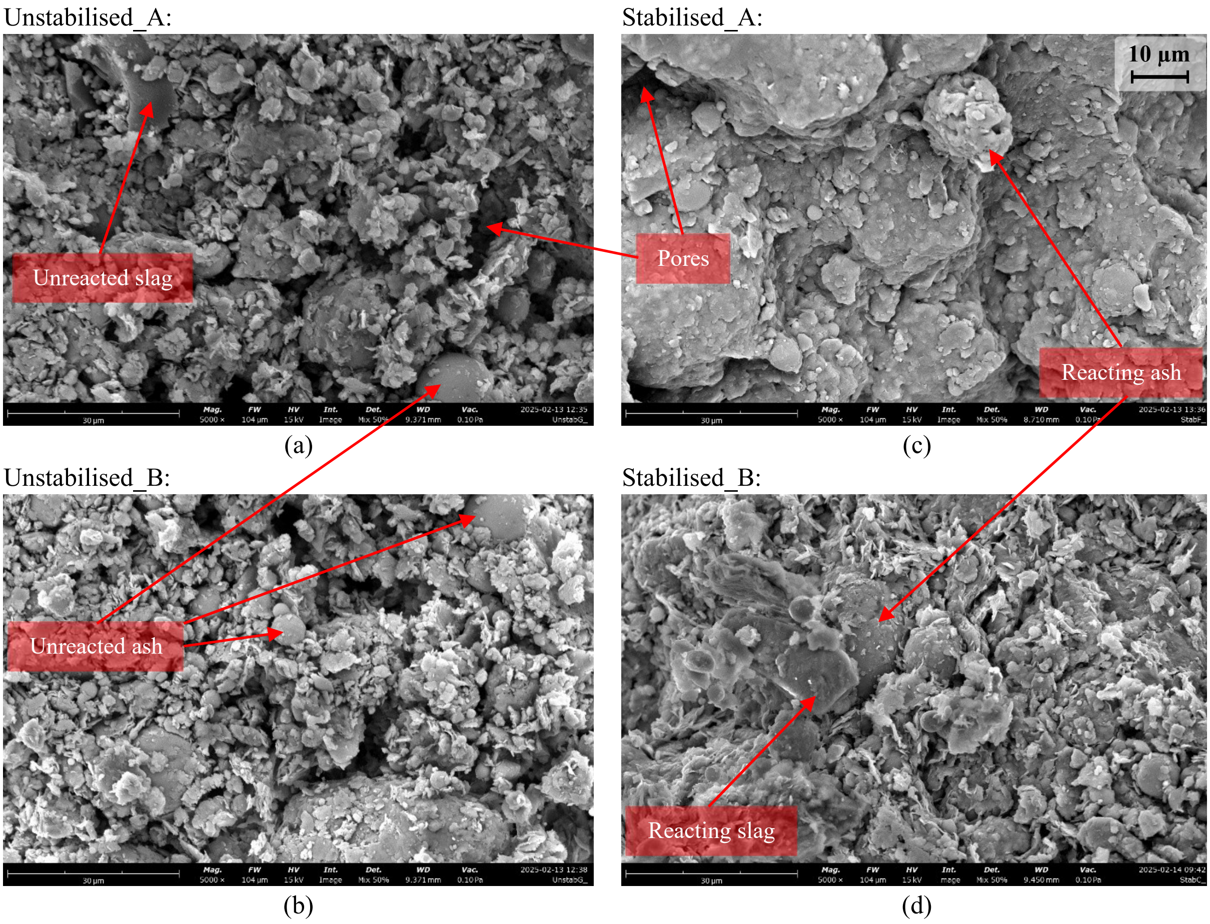
SEM analysis of the same samples was conducted to get a visual perspective of the morphological changes and pore sizes. The SEM images presented in Figure 6.12 provide further insights into microstructural changes caused by stabilisation. The unstabilised samples (sub-plots (a) & (b)) show loosely packed particles with visible pores aligning with sizes estimated by MIP. On the other hand, the stabilised samples (sub-plots (c) & (d)) exhibit a more compact and uniform microstructure. The formation of cementitious products, i.e., reacting ash and slag, is evident in the stabilised samples, likely contributing to the observed increase in strength and reduction in swelling. The reduction in porosity is likely due to the improved mixing and dispersion of mix constituents, facilitated by the initial dispersion and dissolution of aluminosilicate precursors such as clay, ash, and slag immediately after coming into contact with the alkaline solution. In addition, early-stage reactions lead to the formation of geopolymer gel phases such as amorphous NASH or CASH gels, which begin to occupy pore spaces, resulting in a denser microstructure. These findings collectively highlight the effectiveness of low activator systems in altering the pore structure. As an aluminosilicate precursor, clay itself undergoes activation in an alkaline environment, as seen in the denser matrices of the stabilised samples below. The images verify MIP findings by showing that the dominant pore sizes remain unchanged, however, the number of pores is significantly reduced in the stabilised samples.

6.4 Wetting and Drying Cycles

Changes in the swelling behaviour of stabilised soils, pH levels in effluent samples, and leachability of elements detected in effluent samples were studied under monotonic and cyclic wetting-drying conditions. Effluent samples were taken from stabilised and untreated base soil samples. As an initial step, properties such as saturated hydraulic conductivity (K_{sat}) and pore volumes (PV) were calculated to assist in test planning. The stabilised samples were mixed with 20% ash and 10% slag then activated with 20% ALC with 1 M NaOH molar concentration. The unstabilised samples were mixed with distilled water only. All samples were cured for 7 days. Test results are presented in Table 6.2. The increase in porosity from untreated to stabilised samples is attributed to reduced compressibility. In untreated clay, higher plasticity allows greater volume contraction under load, collapsing large pores. Stabilisation forms cementitious bonds that limit deformation, preserving more of the pore structure during consolidation and resulting in higher porosity.

Table 6.2. Hydraulic conductivity and pore volumes of the stabilised and unstabilised base soil

Sample Type	K_{sat} **(m/sec)**	PV **(mL)**
Untreated base soil	4.94×10^{-11}	3.9
Stabilised samples	4.55×10^{-9}	10.3



(a) (c)

(b) (d)

Figure 6.12. SEM images showing the unstabilised (a & b) and stabilised samples (c & d)

Additional samples were prepared, as shown in Table 6.3 to assess leachability and swelling behaviour under wetting and drying cycles. Similar to the above, the stabilised samples were mixed with 20% ash and 10% slag then activated with 20% ALC with 1 M NaOH molar concentration whereas the untreated samples only contained base soil mixed with distilled water. Due to limited time and low permeability of the untreated samples, only one cycle was feasible for the leaching test, which was used as the baseline for this evaluation. Sample Untreated_A was set up in a falling head permeameter apparatus with a pressure-volume controller applying a pressure of 5 kPa, which was the minimum amount of pressure required to trigger water infiltration into the soil. For the stabilised samples, no pressure was required due to their higher permeability. A total of three cycles were carried out for the stabilised samples.

6.4.1 Swelling Behaviour

The effect of wet-dry cycles on the swelling% of untreated base soil and stabilised samples are shown in Figure 6.13. Although wetting cycles leading to full saturation, followed by complete drying, may represent overly extreme conditions, they

Table 6.3. Samples prepared for assessing swelling and leachability under cyclic wet-dry conditions

Sample ID	Tests	Soil used	ash%	slag%	ALC%[1]	M	t_curing
Untreated_A	Infiltration test under 5 kPa GDS pressure	Base soil	0	0	0	0	7
Untreated_B	1DFS	Base soil	0	0	0	0	7
Stabilised_A	1DFS and infiltration tests	Base soil	20	10	20	1	7
Stabilised_B	1DFS and infiltration tests	Base soil	20	10	20	1	7

[1] For unstabilised samples, distilled water was used instead of ALC.

provide a comparative analysis of the long-term behaviour of stabilised and untreated soils samples. The untreated sample shows significant swelling variations while undergoing wetting and drying. In contrast, the duplicate stabilised samples exhibit identical and consistent behaviour across all cycles, with relatively minor swelling behaviour compared to the untreated sample. An important observation is that while stabilisation significantly reduces the volumetric reactivity, there appears to be a progressive but slow increase in swelling with each subsequent cycle. However, it remains unclear whether this trend stabilises or continues beyond the third cycle. Due to the limited amount of time, the experiments were terminated at the end of the third cycle. Further investigations can explore this trend to provide insights into the long-term behaviour of the stabilised soil.

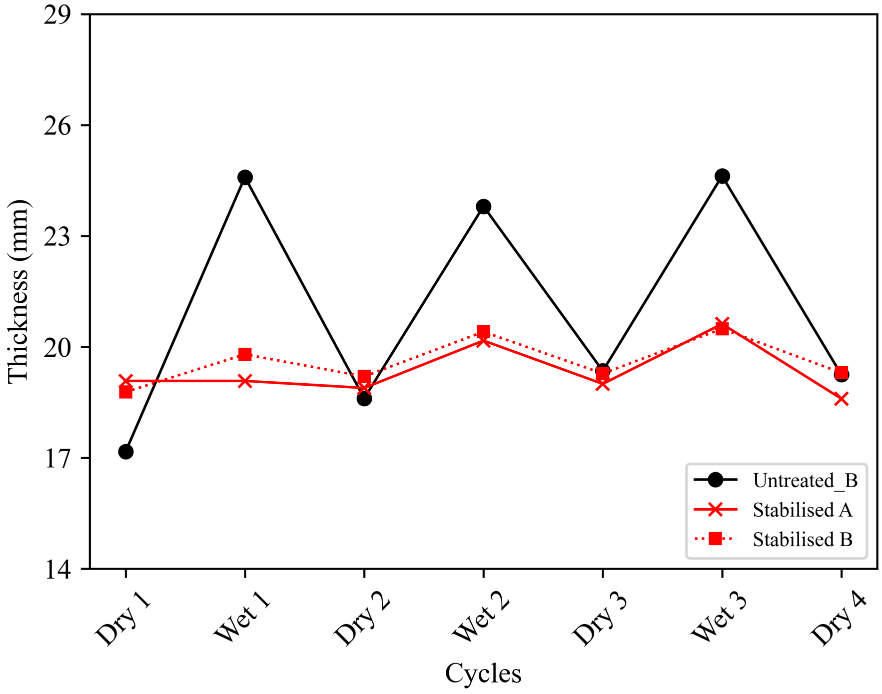


Figure 6.13. The effect of wetting and drying cycles on swelling%

Figure 6.14 compares the water absorption capacity (WAC%), calculated as the difference between soaked and dry sample weights divided by the dry weight, of the stabilised and stabilised soils. Results show that the stabilised samples exhibited consistently lower WAC% compared to the unstabilised sample, indicating lower susceptibility to water absorption across the three cycles. This suggests that stabilisation reduces water absorption, potentially enhancing the durability and performance of the soil under cyclic conditions. This is consistent with the Figure 6.15 showing reduced volumetric change in the stabilised sample.

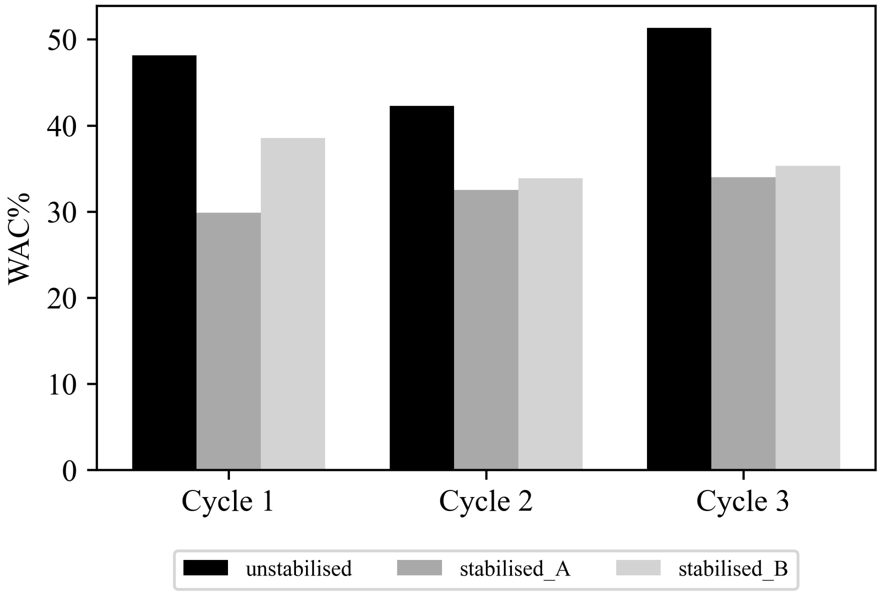


Figure 6.14. The effect of geopolymerisation on WAC of stabilised and unstabilised samples subjected to cyclic wetting and drying

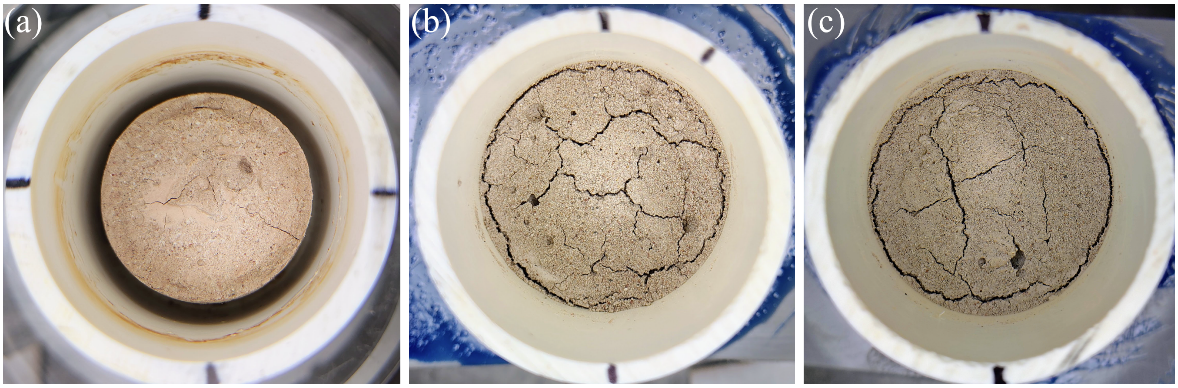


Figure 6.15. Unstabilised (a), stabilised_A (b), and stabilised_B (c) samples after the third cycle of drying

6.4.2 Change in Effluent pH Levels

Elevated pH levels due to dissolved alkaline ions such as OH^-, Na^+, K^+, and Ca^{2+}, can potentially harm sensitive natural ecosystems, and contaminate groundwater and the built environment. To assess this risk, pH levels in effluent from both untreated base soil and stabilised samples were measured and compared. Figure 6.16 presents the pH variations across multiple effluent samples, each representing one PV. For Stabilised A, reduced permeability following each drying cycle prevented collection of the full 10 PVs in cycles 2 and 3.

As shown in Figure 6.16 (a), the pH of effluent from the untreated base soil remained stable between 7 and 8, indicating a neutral to slightly alkaline environment. In contrast, effluents from stabilised samples A and B (Figure 6.16 (b) & (c)) consistently exhibited high pH levels (approximately 12), with a gradual decline over the 30-day wetting cycle. This high pH indicates the potential risks of soil stabilisation if not adequately designed and mitigated.

Noticeable drops in pH were observed after each drying cycle. This may be due to the elevated drying temperature of 50°C, which may have accelerated the use of dissolved ions in the formation of geopolymer gels, hence reducing their availability in subsequent wetting cycles. While this accelerated reaction may simulate long-term changes, it could also limit the generalisability of the results to field conditions where such temperatures are not consistently present. However, the use of 50°C in the drying phase are relevant to arid to semi-arid regions where expansive soils are typically encountered and soil temperatures during the summer season can approach 50°C (Song et al., 2020). From this perspective, the drying conditions applied in the laboratory can be seen as representative of extreme but realistic field scenarios.

The observed pH of approximately 12 is comparable to the elevated pH levels reported due to concrete, which can reach up to 13 (Li et al., 2020). Although it has been reported that such high pH zones typically neutralise over time through natural processes such as dilution and carbonation (Chen et al., 2020; Wan et al., 2013), its potential environmental risks cannot be ignored. To mitigate these risks, several site-specific factors may be considered in the planning and design stages. They include: (i) the topographic setting of the site, (ii) the proximity of stabilised zone to groundwater, sensitive natural ecosystems, and underground utilities and infrastructure; (iii) the depth and extent of the stabilised zone; and (iv) the stratigraphy and properties of the surrounding soils (e.g., permeability). Additional design and operational measures may include scheduling stabilisation works during the summer season, using a low permeability bridging layer or chemically resistant geomembrane to limit water infiltration in highly sensitive areas, minimising moisture evaporation, and installing groundwater monitoring wells.

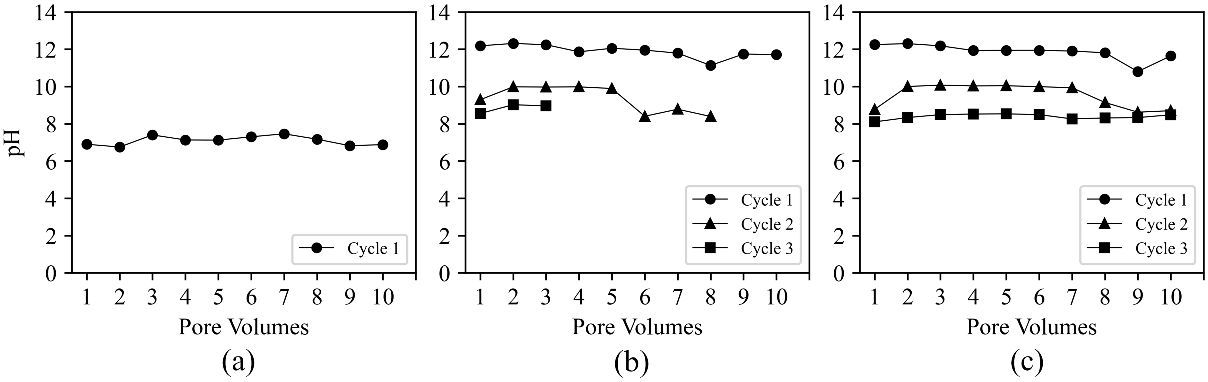


Figure 6.16. Change of pH in effluent samples collected from samples Untreated_A (a), Stabilised_A (b), and Stabilised_B (c)

6.4.3 Leachability of Cementitious Elements

The effluent samples were chemically analysed using the XRF method to detect and measure the intensities of elements. The 2-theta values were used to detect these elements, while the intensity measured in counts per second (cps) provided an estimate of their concentrations, with higher cps indicating higher concentrations. The amount of effluent per test was kept constant at 4 mL to ensure consistency in sample analysis and comparability of XRF results. Maintaining a uniform effluent volume reduces variability in measurements and ensures that cps values for each element accurately reflect differences in element concentrations between samples.

The detected cementitious elements include aluminium (Al), silicon (Si), calcium (Ca), potassium (K), magnesium (Mg), and sodium (Na). The maximum intensity values of the untreated base soil, as represented by horizontal dashed line in the plots of XRF results, are used as the baseline for comparison. Plots are presented in Appendix D. Ideally, intensity peaks for the stabilised samples should lie below the baseline, indicating reduced leaching of elements. A summary of the leaching analysis results is provided in Table 6.4. In the untreated baseline, minor leaching of Al and Si was observed, with leaching diminishing as wetting continued. In contrast, Ca and K showed continuous leaching throughout the wetting cycle, indicating a higher mobility of these ions. No leaching of Mg and Na was detected in the untreated baseline sample.

In the stabilised samples, Al and Si exhibited similar leaching behaviour to the untreated sample. In Cycle 1, both elements were released at low concentrations in Stabilised_A and Stabilised_B. In Cycles 2 and 3, leaching of Al and Si occasionally exceeded levels observed in the untreated sample, but generally showed a decreasing trend with continued wetting. Ca leaching remained evident in both stabilised samples. In Stabilised_A, Ca concentrations initially exceeded those of the untreated sample in Cycle 1 but declined in subsequent pore volumes. However, in Cycles 2 and 3, Ca leaching returned to levels similar to the untreated sample, but with no clear decreasing trend. In contrast, Stabilised_B displayed more sporadic Ca release across pore volumes, without a consistent pattern of reduction. This variability may reflect differences in microstructural development or heterogeneity in the binding of Ca within the stabilised matrix.

K showed persistent leaching across all three cycles in both stabilised samples, with no clear trend of reduction. This continued release is likely due to the high solubility of K and its limited role in gel formation, particularly in systems with elevated N^+ concentrations, where K^+ may remain largely in the pore solution rather than becoming chemically incorporated into the geopolymer matrix. Despite being present in the initial mix, Mg and Na were not detected in the effluent from either stabilised sample, consistently with the untreated case. Their absence may be attributed to their low mobility and/or effective incorporation into stable cementitious phases during geopolymerisation.

Mahedi et al (Mahedi et al., 2019) evaluated the environmental suitability of cement-activated fly ash and slag mixtures by measuring trace element concentrations in effluents collected from stabilised and untreated soils. Their analysis, benchmarked against U.S. EPA Secondary Drinking Water Regulations (U.S. EPA, 2018), revealed that Al^{3+} frequently exceeded the EPA's maximum contaminant level (MCL) of 200 µg/L across all ash and slag blends. Ca also exhibited sustained leaching, a trend similarly observed in the present study. While trace elements such as Cu and Zn were not detected in our samples, Mahedi et al. reported their concentrations remained below the EPA's MCLs of 1000 µg/L and 5000 µg/L, respectively.

Although leachate concentrations in this study are expressed in cps rather than absolute mass units, the relative differences in aluminium levels between effluent from stabilised and untreated soils are still informative. Such comparisons provide insights into the influence of stabilisation on leaching behaviour, even without direct conversion of cps values to quantitative concentrations. We found that the concentrations of Al in effluent samples collected from stabilised samples were lower than the untreated sample across all pore volumes in the first wetting cycles. In the second and third wetting cycles, the concentrations occasionally exceeded that of the untreated soil, particularly in the first two or three pore volumes. By contrast, Mehdi et al. found the opposite, i.e., Al concentrations in the stabilised samples were higher than those of the untreated samples.

Overall, our results suggest that leaching of Al, Si, Mg, and Na is limited and relatively stable across wetting cycles. However, the persistent leaching of K and especially Ca indicates that not all elements are fully immobilised during the relatively short 7-day curing period used in this study. This underscores the importance of extended curing and mix optimisation to reduce leachability and improve environmental performance. While the current study does not assess leaching against absolute regulatory thresholds, comparisons with Mehdi et al. and EPA guidelines can help in understanding potential environmental risks and inform future studies aiming to quantitatively assess leachability in stabilised soils.

Table 6.4. Leachability of cementitious elements

Cementitious Elements	**Untreated_A**[1]	**Stabilised_A**[1]			**Stabilised_B**[1]		
	Cycle 1	**Cycle 1**	**Cycle 2**	**Cycle 3**	**Cycle 1**	**Cycle 2**	**Cycle 3**
Al	ML-Stop	BB-Stop	OE-Stop	OE-Stop	BB-Stop	OE-Stop	OE-Stop
Si	ML-Stop	BB-Stop	OE-Stop	OE-Stop	OE-Stop	OE-Stop	BB-Stop
Ca	CONT	OE-Stop	OE-Cont	BB-Cont	OE-Cont	BB-Cont	BB-Stop
K	CONT	OE-Cont	OE-Cont	BB-Cont	OE-Cont	OE-Cont	BB-Cont
Mg	NL	NL	NL	NL	NL	NL	NL
Na	NL	NL	NL	NL	NL	NL	NL

[1] ML-Stop: minor leaching with a tendency to stop as wetting continued; CONT: continuous leaching without any sign of declining; NL: no leaching detected; BB-Stop: leaching remains below baseline with a tendency to stop as wetting continued; BB-Cont: leaching continued while remaining below baseline; OE-Stop: leaching occasionally exceeded the baseline but with a tendency to stop as wetting continued; OE-Cont: leaching occasionally exceeded the baseline but remained mostly below it and continued throughout wetting.

6.5 Mechanisms Behind Soil Response to Low Activator Dosages

The above findings indicate that in low activator molarity systems, swelling stabilisation and strength enhancement are governed by two distinct but complementary mechanisms. As discussed in Section 6.1.3, swelling stabilisation appears to occur rapidly in the geopolymerisation process, likely during the initial dissolution phase, where the dispersion of solid aluminosilicate structures in clay, ash, and slag, and the release of Al and Si species into the mix under alkaline conditions, reduce the soil's physicochemical reactivity and stabilise swelling. In contrast, strength development under such low activator conditions appears to be driven by: (i) the initial dry density of the mixture, as compaction induced densification enhances mechanical interlocking of particles, and (ii) the continued geopolymer gel evolution producing a low plasticity material suitable to be compacted into a stable subgrade. This contrasts with studies using higher NaOH concentrations, where strength improvements are more strongly associated with rapidly forming three-dimensional gels capable of achieving significantly higher UCS values and forming brittle concrete-like structure (Miranda et al., 2019).

The observed reduction in swelling (Figure 6.4) and increase in strength (Figure 6.8 & Figure 6.10) can be explained by the two low-activator molarity mechanisms explained above. The effectiveness of such systems depends on factors such as stabilisation objective (e.g., modified unbound vs. lightly bound vs. bound), NaOH molarity, ALC%, initial dry density, curing time, and the availability of reactive aluminosilicates. While short curing periods may be sufficient for swelling reduction due to early physicochemical effects, extended curing may be needed to achieve greater strength gains, through continued gel evolution, depending on design requirements and compaction energy. This distinction is crucial in designing stabilisation strategies, as mitigating swelling does not guarantee satisfactory strength gain.

To address both swelling and strength requirements for shallow stabilisation, this thesis proposes a potentially more efficient and sustainable approach by combining geopolymerisation to stabilise swelling with mechanical compaction to meet strength requirements, thereby eliminating the need for costly high-molarity activator solutions and long curing times. This combined strategy offers a practical, cost-effective, and likely more environmentally friendly solution, particularly well-suited for subgrade stabilisation for road and railway infrastructure. It enables the adoption of standard subgrade compaction commonly used for low-expansive materials. For example, Australian guidelines such as TfNSW-R44 and Austroads recommend achieving $\geq$ 95 to 100% relative compaction with respect to the standard Proctor maximum dry density (Austroads, 2021; Transport for NSW, 2022), depending on project-specific performance requirements. Soil mixtures activated with low activator dosages satisfy the requirements of the 'unbound granular materials' as described by Austroad (Austroads, 2021). These materials are required to have low plasticity and a maximum 28-day UCS of 1MPa when cured under moist curing. Stabilised materials with higher strength and brittleness are classified by Austroad as 'bound' materials, which are more suitable to be used as sub-base and base materials rather than subgrade materials. For example, the Department of Transportation in the state of California requires the stabilised subgrades to have a UCS (unsoaked) of at least 115 kPa, depending on the design requirements (Jones et al., 2010). As shown in Figure 6.8, UCS values for samples with higher initial dry densities range from approximately 200 to 280 kPa, meeting the requirements of both standards.

CHAPTER 7

PREDICTING SWELLING OF STABILISED SOILS

As illustrated in Figure 7.1, this chapter follows a structure similar to that of Chapter 5. It begins with data preparation and refinement, before developing new swelling prediction models using Extra Trees (ET), Genetic Programming (GP), and Artificial Neural Networks (ANN). The chapter then presents a comprehensive evaluation of model performance and uncertainty, followed by statistical comparisons between the two best-performing models (ET and GP). Finally, SHAP-based interpretability analyses are used to examine how individual predictors influence the swelling behaviour of stabilised soils.

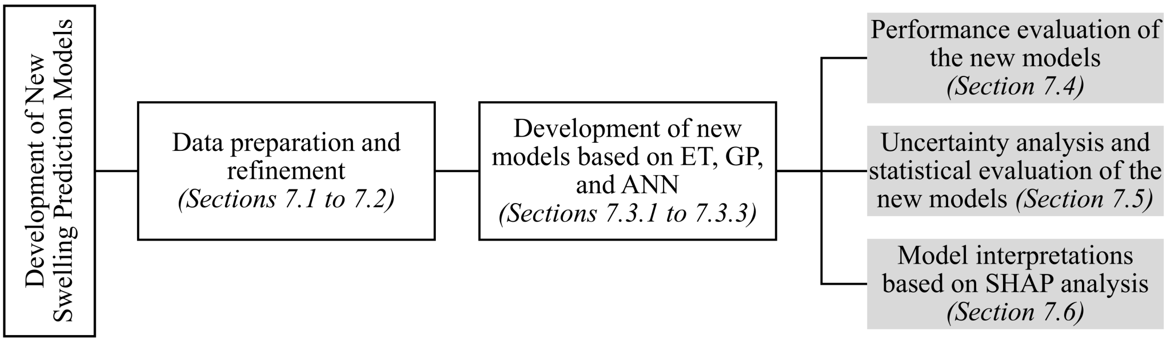


Figure 7.1. Mind-map summarising the structure and major analytical components of Chapter 7

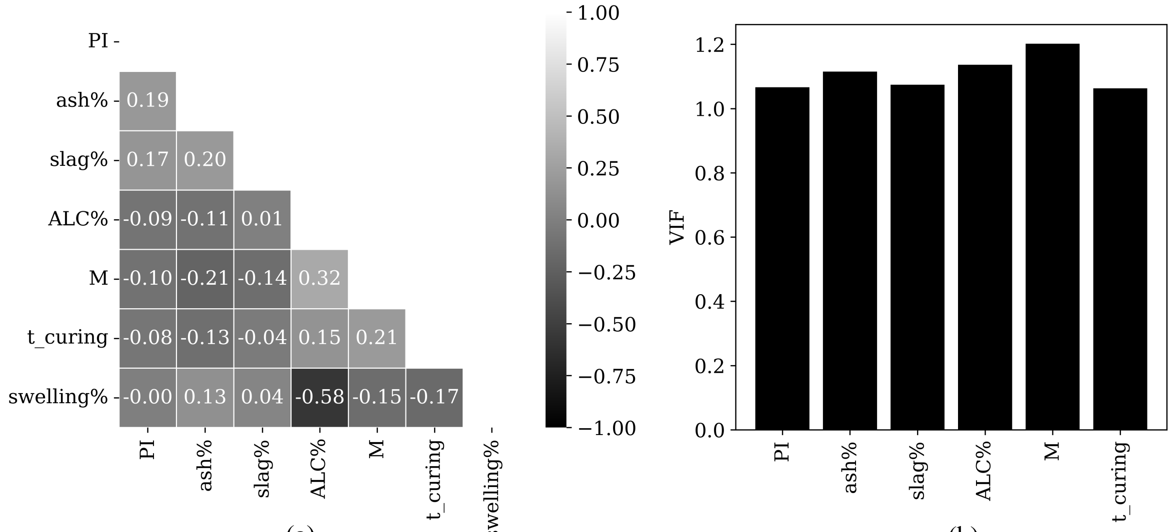
7.1 Data Preparation

The swelling experiments conducted as part of this thesis, and reported in Chapter 6, produced 260 test records. These records form the dataset used for the machine learning prediction of swelling in stabilised soils presented in this chapter. This dataset comprised the following variables: *soil%*, *PL*, *LL*, *PI*, *ash%*, *slag%*, *M*, *ALC%*, *t_curing*, and the target parameter, *swelling%*, which was calculated using Equation 3.8. The dataset included results from experiments conducted on the base soil and four amended soils (amended_1 to amended_4), all of which had different smectite contents, Atterberg limits, and swelling potentials. Laboratory test results were compiled into a single dataset (dataset 1) consisting of over 260 test records, including results of duplicate tests. To ensure that it was suitable for modelling, the dataset underwent a series of pre-processing and analysis steps to refine the data for modelling. Dataset 1 was further refined yielding dataset 2 and then dataset 3, in two successive steps, as described below, with dataset 3 used for developing the machine learning models.

7.2 Assessing Covariance and Multicollinearity

To prepare the dataset for assessing covariance and multicollinearity, duplicate test records were first combined into one, containing the average swelling value of the entire duplicate set. This modification yielded dataset 2 with 221 test records. Correlations were then calculated, which indicated strong covariance among *soil%*, *LL*, *PL*, and *PI*. As a result, only the *PI* was retained, while the other highly correlated predictors were removed from the analysis, resulting in dataset 3, which also contained 221 test records.

Figure 5.4 presents r_s and VIF results showing low correlations among predictors with VIF scores of less than 1.2 indicating the absence of high covariance and multicollinearity in dataset 3. The final set of refined predictors in dataset 3, along with their statistical summaries, are presented in Table 7.1. Figure 7.3 presents the distribution of dataset 3 showing both irregularities and discontinuities in the refined data.



(a)

(b)

Figure 7.2. Results after refinement (dataset 3) showing reasonably low r_s values (a) and VIF scores (b)

Table 7.1. The final set of predictors and their statistical descriptions in dataset 3

Variables	Type	Unit	Min	Mean	Max	Predictor Description
PI	Predictor	-	29.37	38.56	57.15	PI of the unstabilised soil
ash%	Predictor	%	5	16.82	40	Ash fraction in the sample mixture
slag%	Predictor	%	5	9.02	30	Slag fraction in the sample mixture
ALC%	Predictor	%	10	19.32	60	Alkali liquid content
M	Predictor	M	0.5	1.48	14	NaOH molarity
t_curing	Predictor	Days	0.25	6.71	34	Curing time
swelling%	Target	%	0	5.26	20.08	One-dimensional free swelling

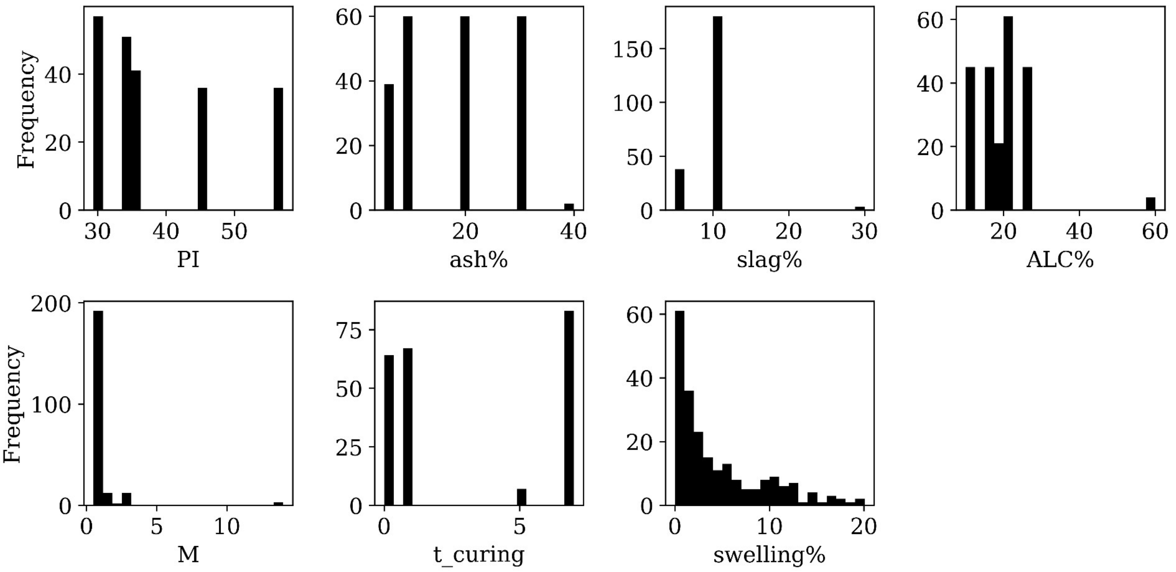


Figure 7.3. Histograms of predictors in refined dataset 3

7.3 Model Optimisation Results

The findings presented in Chapter 5 for predicting UCS indicate that all tree-based ensembles have comparable performance, with the exception of AdaBoost, which lagged behind, and ET, which achieved marginally better overall results. On this basis, this chapter adopts the ET method to be compared with GP and ANN in the context of swelling stabilisation.

7.3.1 Extra-trees Regressor

A clear advantage of the ET approach was that the process of its hyperparameter optimisation was significantly more computationally efficient and less labour-intensive than that of the GP and ANN methods. Out of the five hyperparameters optimised, the interaction between `max_features` and `max_depth` was critical in mitigating learning issues such as overfitting and underfitting. Table 7.2 summarises the search spaces and the final optimised hyperparameter values for the ET model.

Table 7.2. Hyperparameter search spaces and their optimised values for the ET model

Hyperparameters	**Hyperparameter Search Space**	**Optimal Values**
`n_estimators`	10 to 25	13
`max_features`	4, 6, 7, 8, 10	6
`max_depth`	4, 6, 8, 10	8
`min_samples_leaf`	2, 3, 5	2
`min_impurity_decrease`	0.001 to 0.5	0.02

7.3.2 Artificial Neural Network Model

Similar to the optimisation of the UCS model, ANN-23, the interactions between (i) the number of hidden layers and the number of nodes per hidden layer, and (ii) the learning rate and batch size were critical in optimising the swelling model. Performance improved by increasing the number of hidden layers to two, containing 8 and 6 nodes, respectively. The final tuned values for the ANN model are presented in Table 7.3.

Table 7.3. Hyperparameter search space and optimised hyperparamter values for the ANN model

Hyperparameters	**Search Space**	**Optimal Values**
No. of input nodes	6	6
No. of hidden layers	1, 2, 3	2
No. of nodes in hidden layer 1 (connection type = `dense`)	4 to 15	8
Activation function for layer 1 nodes	ReLU	ReLU
No. of nodes in hidden layer 2 (connection type = `dense`)	4 to 15	6
Activation function for layer 2 nodes	ReLU	ReLU
Output Layer with a single node (connection type = `dense`)	1	1
Activation function for the output node	`Linear`	`Linear`
Optimisation method (`solver`)	`Adagrad`	`Adagrad`
Metric used to calculate error (`loss`)	MAE, MSE	MSE
A specified rate (`learning_rate`) to allow weights and biases to be updated based on the calculated `loss`	0.01 to 0.3	0.22
No. of examples processed together before updating model parameters (`batch_size`)	5, 7, 10, 15, 30, 60	7
No. of times the entire training set was used to tune weights and biases (`epoch`)	0 to 1000	135

Figure 7.4 shows the learning curve for the ANN model, indicating a progressive decrease in MSE for both training and testing datasets over 135 epochs. Initially, the training error decreased sharply, indicating rapid improvement in the model's ability to fit the training data. The testing set followed a similar trend, gradually converging with the training line by approximately 110 epochs. After approximately 115 epochs, both the training and testing lines began to completely *converge*

and *flatten off*. These observations are acceptable indicators of a reasonably effective training process. The optimal architectural set-up of the ANN model is illustrated in Figure 7.5. The full mathematical formulation of the ANN model is given in Appendix E.

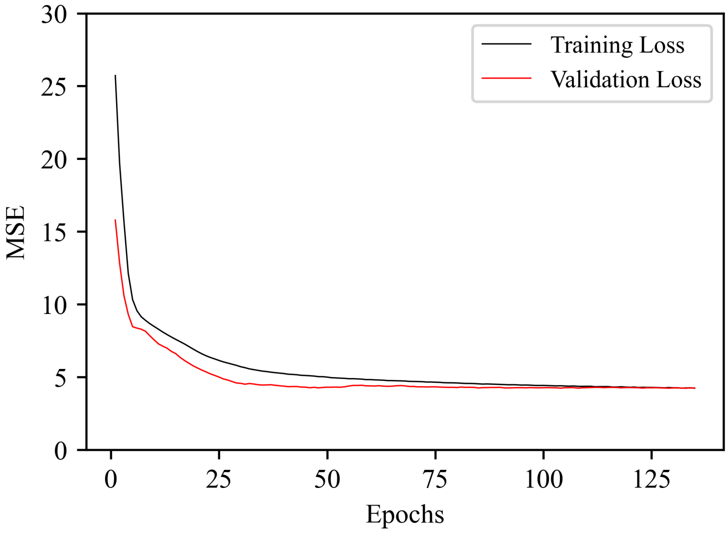


Figure 7.4. The learning curve for the ANN model showing a sharp decrease in MSE up to about 30 epochs during both training and testing stages, followed by convergence with further reductions in MSE, and stabilisation

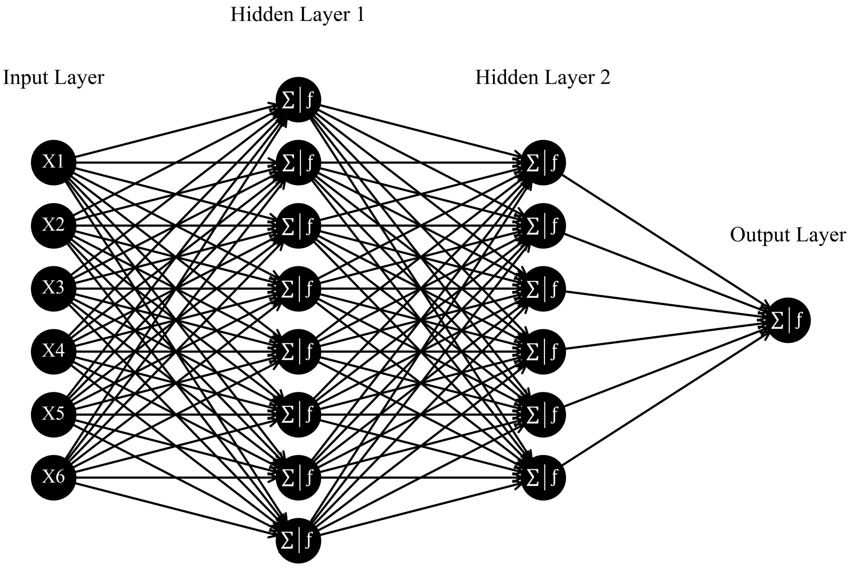


Figure 7.5. The optimal architectural setup of the ANN model consisting of 6 input nodes connected to the first hidden layer with 8 nodes, a second hidden layer with 6 nodes, with a single-node output

7.3.3 Genetic Programming Model

The hyperparameters `population_size`, `generation`, and the set of mathematical `functions` emerged as the most important factors in the optimisation of the UCS model, GP-23. Based on thisa targeted search was first conducted to identify the optimal set of mathematical functions that best captured the underlying relationships in the swelling data. With this function set fixed, a focused tuning of `population_size` and `generation` was carried out next, carefully exploring their interaction effects. The hyperparameter search space and the optimal values for the GP model are presented in Table 7.4.

Table 7.4. Hyperparameter search space and optimised hyperparameter values for the GP model

Hyperparameters	**Search Space**	**Optimal Values**
`population`: Number of individuals in each generation	100 to 1000	400
`generation`: The number of cycles the algorithm runs	20 to 500	280
`tournament`: The number of models (individuals) randomly selected from the population to compete in a "tournament" for survival and reproduction	9, 10, 11, 12, 15	12
`stopping_criteria`: Conditions that terminate the evolutionary process before reaching the maximum number of generations	0.01	0.1
`fitness`: The performance measure used to evaluate and compare candidate models during the evolutionary process	MSE, RMSE	RMSE
`max_samples`: The maximum number of samples (data points) used for training the models in each generation	0.3 to 1	0.92
`p_crossover`: The probability that two parents will pass parts of their structure to their offspring to explore new solutions	0.7, 0.8, 0.9, 0.95	0.9
`p_subtree_mutation`: The probability that a randomly selected subtree will be replaced with a newly generated subtree, introducing variation and helping avoid local optima	0.01, 0.03	0.01
`functions`: The set of metathetical functions (building blocks) used to construct polynomial terms in candidate models	$-$, $+$, $\times$, $\surd$, abs(), $-ve$, inverse	$+$, $\times$, inverse
Others	Default	Default

The optimised GP expression tree presented in Appendix F yields the closed-form model shown in Equation F.1, where $ash\%$, $slag\%$, $ALC\%$, M and t_curing are denoted by a, s, A, M and t, respectively. The optimised tree excludes PI, due to its negligible impact in the evolutionary process. Before using this formula, all inputs must be normalised according to Equation 3.7 using the summary statistics presented in Table 7.1.

7.4 Performance Evaluation & Comparison

Distributions of cross-validation R^2 scores for the ET, GP, and ANN models are presented in Figure 7.6. The mean R^2 scores of the three models clustered between

0.78 and 0.80. The ET model has the highest mean (0.80) R^2 and a relatively tight interquartile range (± 0.055). Model GP demonstrated comparable performance with an average score of 0.79 and slightly smaller variability (± 0.049). In contrast, ANN, while capable of achieving high R^2 scores, demonstrated slightly greater variability (± 0.0) with an average CV score of 0.78.

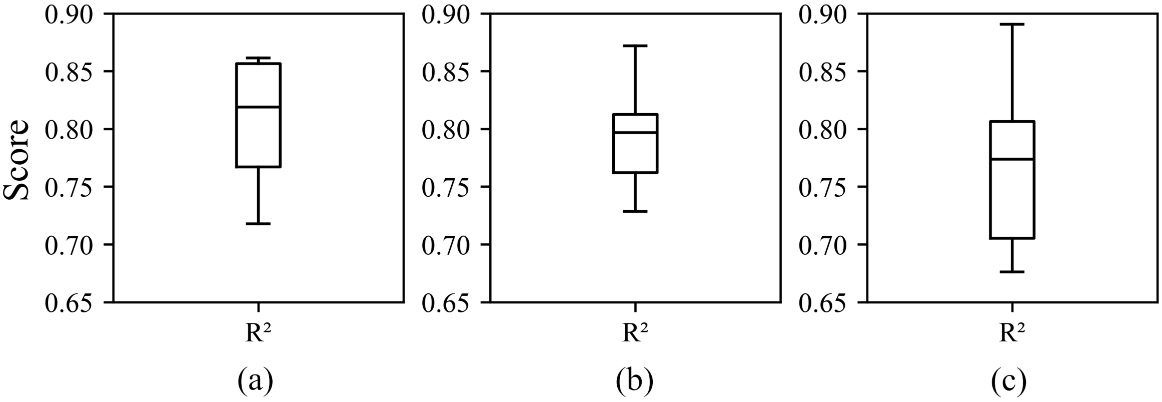
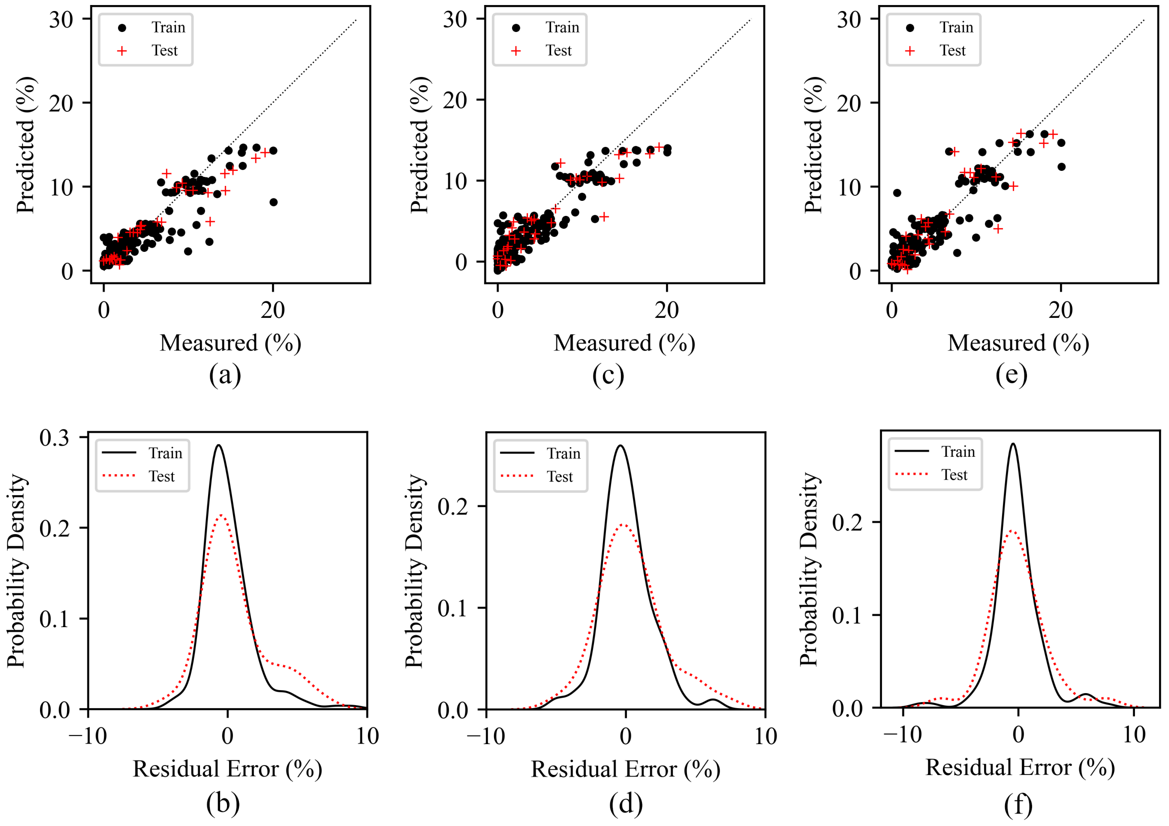


Figure 7.6. Distribution of cross-validation scores based on R^2 for the ET (a), GP (b), and (c) ANN models

Scatter plots comparing the predicted versus laboratory-measured *swelling%* values were used as a visual assessment of model accuracy. Ideally, the data points would closely follow the 45-degree line, indicating high accuracy in predicted output. Any noticeable deviation from the 45-degree line would indicate training issues such as overfitting or underfitting, depending on the distribution and pattern of the errors. Figure 7.7 shows the performance accuracy of the models in predicting *swelling%*. Figure 7.7a, c, and e present scatter plots comparing the predicted with the measured *swelling%* for the training and testing sets for models ET, GP, and ANN, respectively. In all models, the training and testing data points show similar distributions, with better alignment with the 45-degree line at lower target values but diverge more at higher values, where points become increasingly scattered.

Figure 7.7b, d, and f present the residual error distributions for both training and testing sets for ET, GP, and ANN, respectively. The residual errors for all models are centred (peaked) approximately around zero and follow approximately a symmetrical pattern with some irregularities at the two ends. All three models show sharp probability density curves for both training and testing sets, with ET showing the highest testing peak (Figure 7.7b), followed by GP (d) and ANN (f). These results indicate the effectiveness of all models in extracting meaningful insights from the training data to estimate *swelling%* while maintaining a similar level of generalisation performance across the unseen test data.



(a) (c) (e)

(b) (d) (f)

Figure 7.7. Predicted versus measured swelling% values and their residual errors for ET (a & b), GP (c & d), and ANN (e & f) models

The performance metrics, including R^2, CV scores, MAE, MSE and RMSE, are summarised in Table 7.5. Across all models, training and testing results are closely aligned, indicating that training issues such as overfitting and underfitting have been effectively mitigated. ET consistently achieved the highest R^2 and lowest errors in both training and testing phases. GP and ANN exhibited almost identical performance with slight variations in scores. ANN slightly outperformed GP on most metrics, whereas GP performed better in CV R^2, MSE_{train}, and $RMSE_{train}$.

Table 7.5. Training and testing performance for models ET, GP, and ANN

Model	**R²**		**CV R²**	**MAE**		**MSE**		**RMSE**	
	Train	**Test**	**Score**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
ET	0.82	0.82	0.80 ±0.055	1.27	1.62	3.82	5.39	1.95	2.32
GP	0.81	0.79	0.79 ±0.049	1.42	1.75	3.93	6.55	1.98	2.56
ANN	0.81	0.82	0.78 ±0.067	1.33	1.69	4.00	5.65	2.00	2.38

7.5 Model Uncertainty Analysis

The performance uncertainty of the three models was analysed based on the R^2 scores obtained from bootstrap resampling. For the ET model, the training phase, Figure 7.8a, achieved a mean R^2 of 0.88, with a 95% confidence interval (CI) ranging from 0.86 to 0.92, while the testing phase (Figure 7.8b) yielded a slightly lower mean R^2 of 0.82, with a wider 95% CI from 0.75 to 0.87. The GP model achieved the same mean R^2 scores as the ET model but with wider CI's (0.78 to 0.91 for training and 0.72 to 0.82 for testing). The lowest scores and widest CI's were recorded for the ANN models across both training and testing sets. Overall, the ET and GP models outperformed the ANN in terms of both predictive accuracy and stability, as evidenced by higher mean scores and tighter CI's. These findings are consistent with the standard deviations of R^2 scores calculated through cross-validation. See Table 7.5.

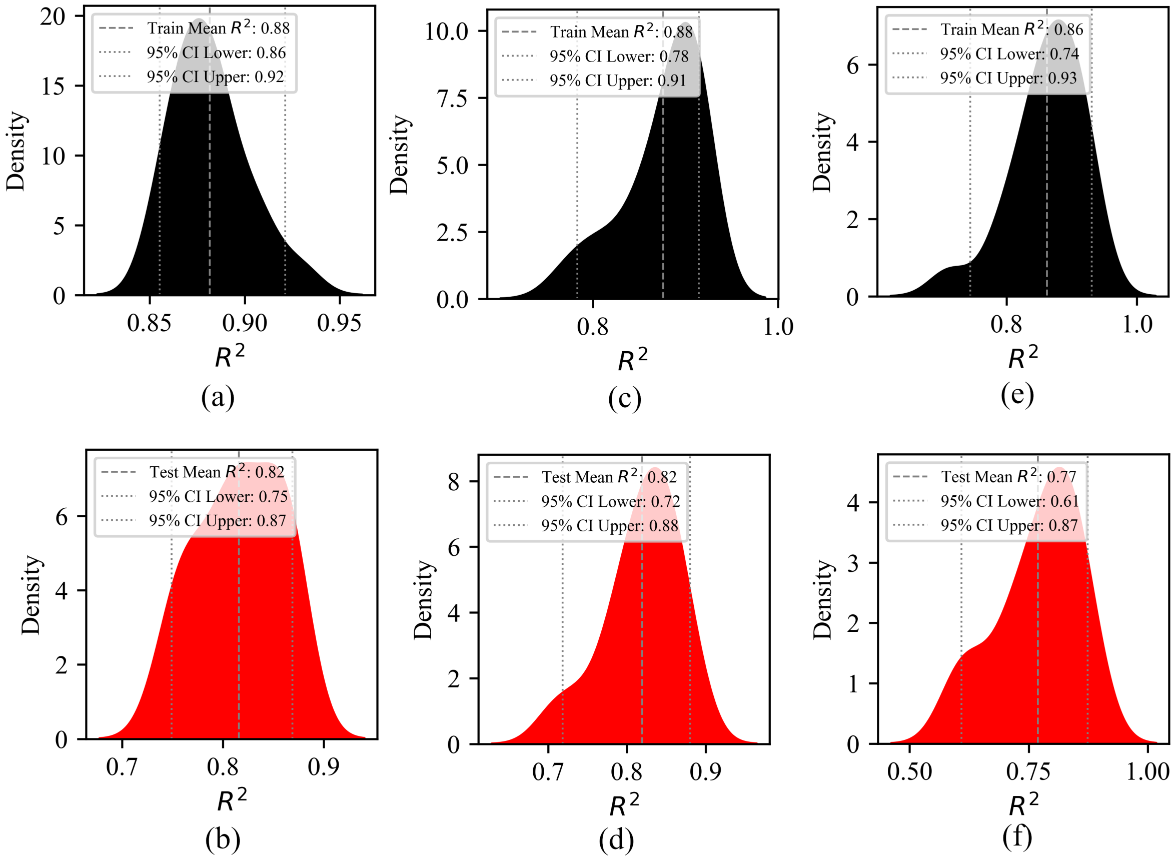


Figure 7.8. Analysis of performance uncertainty for the training and testing phases for the ET (a & c), GP (b & d), and ANN (e & f) models

The Wilcoxon signed-rank test on the paired R^2_{test} scores yielded 13 positive and 7 negative differences ($\Delta = R^2_{\text{ET}} - R^2_{\text{GP}}$), indicating a slight preference for the ET model. However, with a p-value of 0.0759, above the conventional 0.05 threshold, the null hypothesis, which states that the median difference between ET and GP R^2_{test} scores is zero, cannot be rejected. Although ET exhibits a slight advantage, the lack of statistical significance suggests that the observed performance gap is likely due to random variation rather than significant differences in generalisation performance.

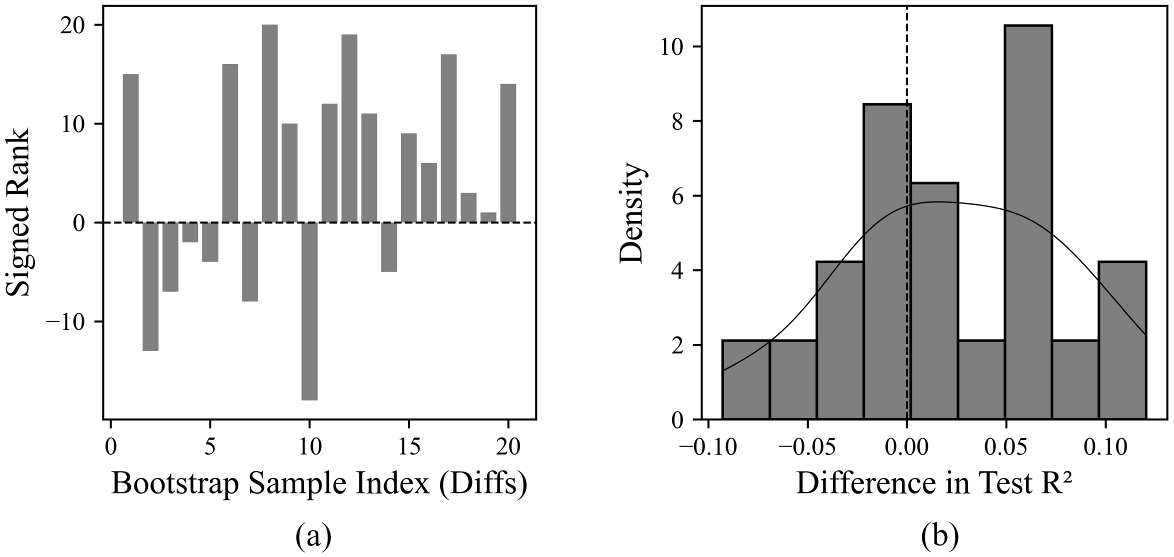
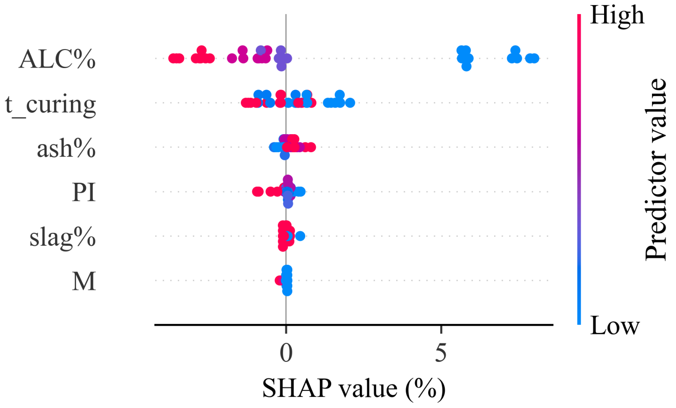


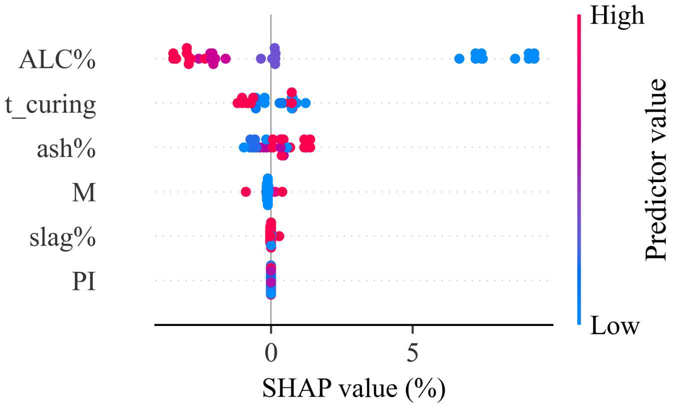
Figure 7.9. Wilcoxon signed-rank analysis of the paired $R^2{}_{test}$ differences ($\Delta = R^2_{ET} - R^2_{GP}$), (a) Signed ranks for all non-zero differences showing 13 positive and 7 negative values, and (b) Histogram of the paired $R^2{}_{test}$ differences, displaying an approximately symmetric distribution around zero and a non-significant p-value (0.0759)

7.6 Model Performance Interpretation

SHAP analysis results for the ET and GP models are shown in Figure 7.10 (a) and (b), respectively. These results illustrate the importance and influence of each predictor based on their SHAP values along the x-axis. In both models, *ALC%* emerged as the most influential predictor, indicating its strong capacity to shift predicted values away from the mean. Similarly both models ranked *t_curing* and *ash%* as the second and third most important predictors. Other predictors showed moderate to low levels of importance, with some variation in their relative rankings across the two models. Predictor *M*, ranked last in the ET model, came up as the fourth most important predictor in the GP model. *PI*, the fourth-ranked predictor in ET, was completely dropped from the GP expression tree due to its negligible contribution. Predictor *slag%* remained as the second-last important predictor in both models.



(a)



(b)

Figure 7.10. Predictor importance based on SHAP values for the ET (a) and GP (b) models

SHAP dependency plots presented in Figure 7.11 show further details on how individual predictors influence model outputs based on their SHAP values. Each subplot represents the relationship between a specific predictor and its corresponding SHAP value. For both models, the plots reveal mostly similar trends in how predictors affect the outputs with some difference, particularly in the case of *slag*% as discussed below.

In ET, *PI* exhibits a negative linear relationship with *swelling*% suggesting that swelling stabilisation is more effective as clay content increases likely due to greater amount of aluminosilicate clay fraction in the mix. The model yields negative SHAP values for *PI* values of approximately greater than 45% suggesting reduction of predicted values from their mean outputs.

In both models, *ash*% appears to have an overall positive relationship with swelling. *ash*% yields lower model outputs relative to the mean up to approximately 18% ash content, beyond which it pushes model output above the mean output suggesting increased *swelling*%.

Predictors *slag*% and *M* show similar influence to model outputs in both cases. SHAP values are consistently positive when predictor values are close to 0% and negative as predictor values increase, particularly in the case of *M*. This suggests that increasing their content reduces *swelling*% yielding better stabilisation outcome. *ALC*% demonstrates a similar impact, however, the threshold value for *ALC*% at which there is a shift in SHAP values from positive to negative is consistently around 15%. This suggests that lower *ACL*% pushes model outputs above their mean values resulting in higher swelling predictions and vice versa.

The impact of *t_curing* appears to be more sporadic with a negative correlation with SHAP values. This is consistent with laboratory results showing that swelling can be significantly reduced within the first 24 hours after the commencement of curing depending on interactions with other factors such as *ALC*% and *M*. Results indicate that *swelling*% can be further reduced with longer curing time.

While experimental results on aluminosilicate-based stabilisation remain limited, the findings from the SHAP analyses broadly align with the experimental outcomes reported in Chapter 6. This is unsurprising, given that the models developed here are based on the same experimental dataset; however, the agreement provides confidence that the models have successfully captured the patterns revealed by the laboratory investigation.

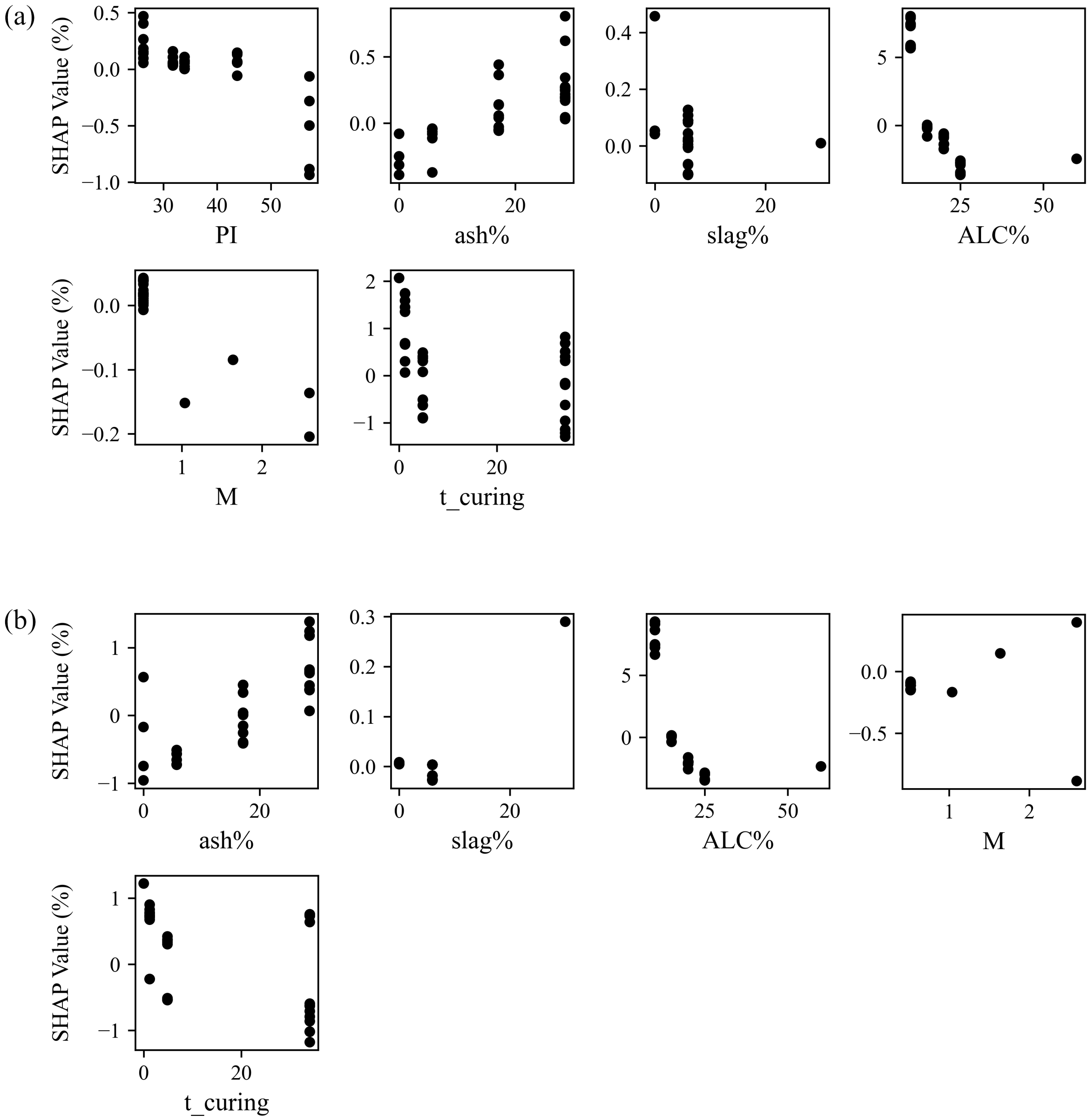


Figure 7.11. Dependency plots for the ET (a) and GP (b) models

CHAPTER 8

CONCLUSIONS

8.1 Research Summary

Industrial wastes rich in alumina and silica, such as fly ash and slag, offer promising potential for improving soil behaviour while reducing waste and reliance on energy-intensive binders with larger CO_2 footprints, such as cement. However, the conventional trial-and-error-based experimental approach to determining optimal mix designs, is resource-intensive, time-consuming, and costly, and hence acts as a significant barrier to the use of these binders. The present thesis addressed this barrier by promoting the integration of predictive modelling in soil stabilisation planning and design. This approach has the potential to overcome the above-mentioned barrier by reducing the amount of laboratory assessments and limiting them to a minimal number of validation tests. The thesis developed predictive models using proven regression techniques to estimate key stabilisation properties namely the unconfined compressive strength (UCS) and swelling potential.

The effectiveness of alkali-activated aluminosilicate by-products in increasing the UCS of low-strength soils has been extensively investigated in the literature, with empirical studies reporting positive outcomes. The present thesis tapped into the rich body of existing empirical data to develop more versatile models to predict UCS. However, for swelling stabilisation, while there is empirical evidence suggesting that these binders are effective in reducing swelling, critical research gaps persist in the literature, including the influence of key factors, interaction effects, underlying stabilisation mechanisms, durability and potential environmental impacts due to high pH of stabilised soil mixtures, and the optimised use of binders and NaOH activator for swelling stabilisation. The present thesis systematically addressed these gaps through an experimental testing program, which was subsequently expanded to cover five soil types, ranging from low to very high plasticities and swelling potentials. This experimental dataset was then used to produce training data to develop new predictive models for swelling.

8.2 Strength Development

Previous aluminosilicate waste-based predictive models, specifically those that do not use cement and lime, are often based on limited number of experimental data derived from one or few soil types. This study utilised an expanded experimental dataset collected from 23 experimental studies, to develop three exclusive waste-based UCS models based on ET, GP, and ANN, referred to ET-23, GP-23, and ANN-23, respectively. The use of a richer dataset allowed for deeper insights into the key predictors and their interactions. Including cementitious chemical compound contents as predictors resulted in reduced MAE, MSE, and RMSE, whereas the addition of sample preparation methods as predictors provided no measurable benefit. Models ET-23 and GP-23 demonstrated better performance by achieving higher R^2_{test} scores (0.93) and lower errors compared to ANN-23. Further evaluation of ET-23 and GP-23 based on bootstrap resampling and Wilcoxon test indicated similar performance stability, i.e., low uncertainty, and statistically insignificant difference between the performance of the two models.

The behaviours of ET-23 and GP-23 were analysed and compared using SHAP analysis, which provided insights into the influence of individual predictors. The two models demonstrated consistent predictive behaviour, with predictors contributing in a similar manner. However, the degree of their influence varied slightly, reflecting differences in how each model processes inputs while capturing underlying relationships in the dataset. A newly introduced predictor, *total_CaO*%, demonstrated a clear positive quasi-linear correlation with UCS, supporting the informed selection of CaO-rich binders for effective soil stabilisation. In contrast, another new predictor, *total_oxides*%, influenced UCS in a non-linear manner enhancing strength up to around 30% before declining, hence suggesting an optimal dosage levels for aluminosilicate-based additives in soil stabilisation practice. Water content was found to follow an exponentially decreasing relationship with UCS, emphasising the need for precise moisture control during the treatment process. The analysis also highlighted the potential impact of the data gap in long-term curing spanning from 90 to 365 days, which may introduce a slight bias toward short-term strength prediction. The limited representation of longer curing periods in the dataset could lead to models bias towards early-strength estimation. Addressing this gap with additional long-term data could further improve the model's reliability for predicting UCS over extended time frames.

8.3 Swelling Stabilisation

Experimental findings showed that alkaline liquid content is a dominant factor in stabilising swelling. Stabilisation was effective across a wide range of soils with different initial swelling potentials. Stabilisation aimed at reducing swelling was also found to enhance soil strength, with clear dependency on activator content and compaction energy. The findings reflect previously reported interactions between binder and activator contents (Mypati and Saride, 2022) showing that ash-to-activator ratios of greater than 0.75 yields the best stabilisation outcome. Microstructural analyses showed a significant reduction in micro-pores in the stabilised soils suggesting denser and more compact matrices.

The leachability assessment revealed that stabilisation reduced the mobility of some cementitious elements, particularly aluminium and silicon, which showed a general downward trend across wetting cycles. However, calcium and potassium continued to leach at low levels, indicating incomplete immobilisation. Sodium and magnesium were not detected in the effluent, suggesting effective retention within the stabilised matrix. Overall, while stabilisation improved chemical containment, the persistent leaching of certain elements highlights potential durability concerns under cyclic wetting conditions.

Effluent pH remained elevated throughout the fist wetting cycle, potentially posing risks to nearby ecosystems and groundwater if not adequately designed and managed. Although long-term pH levels may gradually decrease due to dilution and carbonation, the study highlights the need for appropriate design, operational, and post-construction maintenance strategies to mitigate potential environmental impacts, particularly in sensitive or high-risk areas.

A key empirical finding of the study addresses another critical barrier to the widespread adoption of such alkali-activated aluminosilicate binders, i.e., the high cost and CO_2 footprint of NaOH production, which is comparable to that of cement when used in high concentrations. The study found that swelling stabilisation requires significantly lower NaOH molarity and shorter curing time than those typically required for strength development. Building on this finding, the study proposed a more sustainable and integrated stabilisation approach to address both swelling and strength requirements at the same time. This approach combines geopolymerisation for swelling control with mechanical compaction for strength, eliminating the need for high-molarity activators and extended curing. The resulting hybrid method offers a practical, cost-effective, and environmentally favourable solution, particularly suited to shallow stabilisation applications such as road and runway subgrades, slab-on-ground, and slab-and-pier systems.

8.4 Swelling Prediction

Three new predictive models were developed to predict swelling using laboratory-generated data from five soil types with varying swelling potentials. The soils consisted of low to very high plasticity classifications, i.e., CL to CH, in the Unified Soil Classification System, covering the range most susceptible to swelling. Over 260 test records of experimental data was used to produce 221 test records of refined data to train three models based of ET, GP, and ANN methods. Once again, the

ET and GP models achieved better predictive accuracy and stability based on bootstrap resampling. The Wilcoxon signed ranks showed no significant difference in performance between ET and GP models. SHAP analysis supported experimental observations by highlighting *ALC%* as the most influential predictor in reducing swelling. Both models showed consistent predictor influences, with *ALC%* emerging as the most influential predictor, followed by *t_curing* and *ash%*. Predictors *M%*, *PI*, and *slag%* exhibited moderate to low importance, with minor differences in their ranking between models. the GP model excluded *PI* due to its insignificant contribution. Overall, the findings demonstrate that both EP and GP can effectively learn and model the complex relationships governing soil swelling behaviour.

8.5 Research Limitations

The experimental and analytical findings presented in this study should be interpreted in light of several research limitations. Three main limitations apply to the experimental component of the study. First, due to constraints in available time and resources, the assessment of swelling behaviour in stabilised soils was conducted using the free swell test. This method was selected for its speed and simple sample-preparation requirements, which made it possible to address the research questions and generate sufficient training data within the time-frame of a PhD project. However, the free swell test does not account for the influence of confining stress, which may impact swelling behaviour in the field. Therefore, further research using more comprehensive tests, such as odometer-based swell tests under controlled stress conditions, is necessary to better understand the response of stabilised soils under load.

Second, sodium bentonite was added to the base soil to produce specimens with varying swelling potentials. This approach was necessary to evaluate the effectiveness of stabilisation across a range of expansive behaviours and to expand the dataset to generate training data. However, the use of artificially modified soil mixtures may not fully capture the natural variability of expansive clays. Also, given the critical role of calcium in the stabilisation process, soils containing, or amended with, calcium bentonite are likely to exhibit different behaviour. This requires separate investigations to evaluate their suitability for stabilisation with aluminosilicate waste.

Third, the use of wetting cycles, up to full saturation, followed by complete drying may represent overly extreme conditions that may not reflect field scenarios. In natural environments, soils are subjected to partial wetting and drying, with varying moisture amplitudes depending on climate, depth, and soil cover. Therefore, future studies should consider more realistic wet-dry conditions that simulate field-relevant moisture fluctuations to better assess the performance and resilience of stabilised soils under actual service conditions. Hysteresis effects may be important in this context, with degrees of leaching and durability of gains that may be dependent of the pathway followed during the cycles and not just the degree of saturation.

Two key limitations relevant to predictive modelling are also present in this study. Although the newly developed models cover a significantly broader range of soil types, mix designs, and binder combinations compared to previous studies, their applicability remains bounded by the specific ranges of predictor values within the training datasets. Consequently, these models may not perform as accurately when applied to conditions that fall outside those bounds. In addition, since the models were developed using data obtained under controlled laboratory conditions, their predictive accuracy may not fully translate to in situ conditions, where field variables and environmental factors introduce additional complexity.

8.6 Recommendations for Further Research

To further advance the current state of the art and practice of geopolymer-based soil stabilisation, future research can focus on several critical areas:

i. Further experiments are needed to assess the influence of varying mechanical compaction energy and overburden pressures on the swelling and strength behaviour of stabilised soils.
ii. While wet-dry cycles represent extreme conditions, increasing the number of cycles and incorporating freeze-thaw cycles could provide a more comprehensive assessment of the durability and applicability of such binders. Such studies would offer valuable insights into the resilience and long-term performance of stabilised soils in real-world conditions. Additionally, future research should extend beyond laboratory studies to investigate in-situ stabilisation performance with respect to both swelling and UCS. Such investigations represent a critical step toward the practical implementation of stabilisation in engineering projects. The findings of this thesis demonstrate a strong case for in-situ evaluation, given the highly promising potential of aluminosilicate waste as a stabilising agent.
iii. Incorporating life cycle assessments to evaluate the environmental and economic impacts of using alkali-activated agro-industrial aluminosilicates in soil stabilisation could guide the development of greener and more cost-effective solutions. Such interdisciplinary approaches would significantly contribute to advancing geotechnical engineering and sustainable infrastructure development.
iv. Exploring alternative and less energy-intensive activators such as silica fume, glass wastes (Parthiban et al., 2022) and other unexplored additives has the potential to further reduce the dependence on NaOH. Reducing or replacing NaOH with a more sustainable alternative can improve the sustainability and affordability of the process, making it more viable, acceptable and accessible for a wider range of applications and geotechnical

contexts. These areas of research can collectively strengthen the foundation for the practical and sustainable use of wastes in soil stabilisation.

Similarly, further research is required to enrich future training datasets and predictive modelling:

i. Predictive models, while not a substitute for experimental testing, can play a critical role in optimising the design and decision-making processes by focusing experimental efforts on validation and critical variables and conditions. Future studies could investigate the applicability of these findings to other critical soil properties, such as permeability and resistance acid sulphate attacks, broadening the scope of stabilisation research.
ii. Expanding datasets through further experiments to address the UCS data gap involving curing time starting from 90 to 180 days, and incorporation of other additives, such as high silica agricultural residues or other cementitious industrial by-products with varying compositions, could further improve the versatility of predictive models.
iii. The existing models, including those newly developed in this study, are based on laboratory-cured data. However, such conditions may not accurately represent real-world, large-scale applications where soils are compacted and cured in situ. Field conditions are likely to introduce greater variability due to factors such as inconsistent mixing of ingredients, uneven distribution of the alkaline activator, and environmental influences like excessive rainfall or drying, groundwater presence, and non-uniform moisture content of the host soil. These factors contribute to heterogeneity in the stabilised soil matrix. Therefore, further research is needed to develop and validate practical calibration methods that can align model predictions to better reflect in-situ stabilisation outcomes.

Bibliography

Abdila, S. R., M. M. A. B. Abdullah, R. Ahmad, D. D. Burduhos Nergis, S. Z. A. Rahim, M. F. Omar, A. V. Sandu, P. Vizureanu, and Syafwandi (2022, January). Potential of Soil Stabilization Using Ground Granulated Blast Furnace Slag (GGBFS) and Fly Ash via Geopolymerization Method: A Review. *Materials 15*(1), 375.

Abdullah, H. H., M. A. Shahin, and P. Sarker (2019, April). Use of Fly-Ash Geopolymer Incorporating Ground Granulated Slag for Stabilisation of Kaolin Clay Cured at Ambient Temperature. *Geotechnical and Geological Engineering 37*(2), 721–740.

Abdullah, H. H., M. A. Shahin, and M. L. Walske (2019, December). Geo-mechanical behavior of clay soils stabilized at ambient temperature with fly-ash geopolymer-incorporated granulated slag. *Soils and Foundations 59*(6), 1906–1920.

Abdullah, H. H., M. A. Shahin, and M. L. Walske (2020, June). Review of Fly-Ash-Based Geopolymers for Soil Stabilisation with Special Reference to Clay. *Geosciences 10*(7), 249.

Abdullah, H. H., M. A. Shahin, M. L. Walske, and A. Karrech (2021, January). Cyclic behaviour of clay stabilised with fly-ash based geopolymer incorporating ground granulated slag. *Transportation Geotechnics 26*, 100430.

Abdullah, M. S. and F. Ahmad (2017). Effect of Alkaline Activator to Fly Ash Ratio for Geopolymer Stabilized Soil. *MATEC Web of Conferences 97*, 01012.

Adedokun, S. and J. Oluremi (2019). A review of the stabilization of lateritic soils with some agricultural waste products. *International Journal of Engineering 17*, 63–74.

AGS (2025). Practical applications of generative ai in geotechnical engineering. `https://geomechanics.org.au/videos/practical-applications-of-generative-ai-in-geotechnical-engineering/`. Seminar video, 10 September 2025.

Ahmad, S., M. Shah Alam Ghazi, M. Syed, and M. A. Al-Osta (2024, June). Utilization of fly ash with and without secondary additives for stabilizing expansive soils: A review. *Results in Engineering 22*, 102079.

Ahmaruzzaman, M. (2010, June). A review on the utilization of fly ash. *Progress in Energy and Combustion Science 36*(3), 327–363.

Ahmed Hisseini, B. H., A. Bennabi, R. Hamzaoui, L. Makki, and G. Blanck (2021, November). Treatment and Recovery of Clay Soils Using Geopolymerization Method. *International Journal of Geomechanics 21*(11), 04021206.

Akinwande, M. O., H. G. Dikko, and A. Samson (2015). Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis.

Open Journal of Statistics 05(07), 754–767.

Altmeyer, W. T. (1955, March). Discussion of Engineering Properties of Expansive Clays. Volume 81 of *658*.

Andavan, S. and V. K. Pagadala (2020). A study on soil stabilization by addition of fly ash and lime. *Materials Today: Proceedings 22*, 1125–1129.

ASTM C618 (2019). Specification for coal fly ash and raw or calcined natural pozzolan for use in concrete.

ASTM International (2024). ASTM d2166/d2166m-16: Standard test method for unconfined compressive strength of cohesive soil. Accessed: 2024-12-12.

ASTM International, Committee D18 (2021). ASTM D4546-21: Standard Test Methods for One-Dimensional Swell or Collapse of Soils. Accessed: 2024-12-12.

ASTM International (D32 Committee) (2017). ASTM D4284-12: Test Method for Determining Pore Volume Distribution of Catalysts and Catalyst Carriers by Mercury Intrusion Porosimetry.

Austroads (2021). Guide to Pavement Technology Part 4D: Stabilised Materials. *Austroads Ltd*.

Avirneni, D., P. R. Peddinti, and S. Saride (2016, September). Durability and long term performance of geopolymer stabilized reclaimed asphalt pavement base courses. *Construction and Building Materials 121*, 198–209.

Ayub, F. and S. A. Khan (2023, November). An overview of geopolymer composites for stabilization of soft soils. *Construction and Building Materials 404*, 133195.

Bhatt, A., S. Priyadarshini, A. Acharath Mohanakrishnan, A. Abri, M. Sattler, and S. Techapaphawit (2019, December). Physical, chemical, and geotechnical properties of coal fly ash: A global review. *Case Studies in Construction Materials 11*, e00263.

Blesson, S. and A. U. Rao (2023, April). Agro-industrial-based wastes as supplementary cementitious or alkali-activated binder material: a comprehensive review. *Innovative Infrastructure Solutions 8*(4), 125.

Chen, F. H. (1975). Nature of Expansive Soils. In *Developments in Geotechnical Engineering*, Volume 12, pp. 1–31. Elsevier.

Chen, J., M. Sanger, R. Ritchey, T. B. Edil, and M. Ginder-Vogel (2020, January). Neutralization of high pH and alkalinity effluent from recycled concrete aggregate by common subgrade soil. *Journal of Environmental Quality 49*(1), 172–183.

Chen, K., D. Wu, Z. Zhang, C. Pan, X. Shen, L. Xia, and J. Zang (2022, January). Modeling and optimization of fly ash–slag-based geopolymer using response surface method and its application in soft soil stabilization. *Construction and Building Materials 315*, 125723.

Chen, X., M. K. Zhou, P. L. Cong, and X. Li (2009). Influence of SO3 Content on the Strength of Cement-Fly Ash Stabilized Crushed-Stones. In G. Li, Y. Chen, and X. Tang (Eds.), *Geosynthetics in Civil and Environmental Engineering*, pp. 398–402. Berlin, Heidelberg: Springer Berlin Heidelberg.

Consoli, N. C., C. G. da Rocha, and S. Maghous (2016, November). Strategies for Developing More Sustainable Dosages for Soil–Coal Fly Ash–Lime Blends. *Journal of Materials in Civil Engineering 28*(11), 04016130.

Cozzolino, C., A. Buja, M. Rugge, A. Miatton, M. Zorzi, A. Vecchiato, P. Del Fiore, S. Tropea, A. Brazzale, G. Damiani, L. dall'Olmo, C. R. Rossi, and S. Mocellin (2023, January). Machine learning to predict overall short-term mortality in cutaneous melanoma. *Discover Oncology 14*(1), 13.

CRC for Low Carbon Living (2015, December). State of Practice: High Volume Applications of Fly Ash and Barriers to Commercialisation. Research Project Report RP1004-II, CRC Australia, Sydney.

Cristelo, N., S. Glendinning, L. Fernandes, and A. T. Pinto (2012, April). Effect of calcium content on soil stabilisation with alkaline activation. *Construction and Building Materials 29*, 167–174.

Cristelo, N., S. Glendinning, L. Fernandes, and A. T. Pinto (2013, August). Effects of alkaline-activated fly ash and Portland cement on soft soil stabilisation. *Acta Geotechnica 8*(4), 395–405.

Cristelo, N., S. Glendinning, T. Miranda, D. Oliveira, and R. Silva (2012, November). Soil stabilisation using alkaline activation of fly ash for self compacting rammed earth construction. *Construction and Building Materials 36*, 727–735.

Cristelo, N., S. Glendinning, and A. Teixeira Pinto (2011, May). Deep soft soil improvement by alkaline activation. *Proceedings of the Institution of Civil Engineers - Ground Improvement 164*(2), 73–82.

Davidovits, J. (2020). *Geopolymer: chemistry & applications* (5th ed ed.). Saint-Quentin: Institut Géopolymère.

de Oliveira, L. B., A. R. de Azevedo, M. T. Marvila, E. C. Pereira, R. Fediuk, and C. M. F. Vieira (2022, June). Durability of geopolymers with industrial waste. *Case Studies in Construction Materials 16*, e00839.

Detphan, C., P. Kaeorawang, B. Injorhor, K. Chompoovong, S. Hanjitsuwan, T. Phoo-Ngernkham, and P. Chindaprasirt (2022). Improving drying shrinkage and strength development of alkali-activated high-calcium fly ash using commercial-grade calcium sulfate as expansive additive. *Engineering and Applied Science Research 49*, 5864. Publisher: Faculty of Engineering, Khon Kaen University.

Devarangadi, M. and U. S. M (2020, July). Correlation studies on geotechnical properties of various industrial byproducts generated from thermal power plants, iron and steel industries as liners in a landfill- a detailed review. *Journal of Cleaner Production 261*, 121207.

Dissanayake, T. B. C. H., S. M. C. U. Senanayake, and M. C. M. Nasvi (2017, February). Comparison of the Stabilization Behavior of Fly Ash and Bottom Ash Treated Expansive Soil. *Engineer: Journal of the Institution of Engineers, Sri Lanka 50*(1), 11.

D'Alonzo, M. and C. Cipriani (2012). Vibrotactile Sensory Substitution Elicits Feeling of Ownership of an Alien Hand. *PLoS ONE 7*(11), e50756. Publisher: Public

Library of Science (PLoS).

Eades, J. L. and R. E. GRTh'I (1966). A Quick Test to Determine Lime Requirements For Lime Stabilization. *Highway Research Board*, 12.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Bootstrap Methods*.

Faizi, H., D. Airey, and A. El-Zein (2025a, November). Balancing swelling control and strength in clayey subgrades using waste-based geopolymers. In *Proceedings of the 2025 Australian Geomechanics Society (Sydney) Symposium*, Sydney, Australia. Accepted for presentation and publication.

Faizi, H., D. Airey, and A. El-Zein (2025b). Stabilising expansive subgrades with agro-industrial waste: An experimental investigation. *Transportation Geotechnics*, 101729.

Faizi, H. and A. El-Zein (2025a). Evaluating swelling behaviour and durability of geopolymerised expansive soils under cyclic moisture conditions. In *Proceedings of the 5th European Conference on Unsaturated Soils*. Conference presentation.

Faizi, H. and A. El-Zein (2025b, June). Predicting dry density of subgrades stabilised with agro-industrial byproducts: Comparison of tree-based ensembles and genetic programming. In *Proceedings of the 6th International Conference on Environmental Geotechnology, Recycled Waste Materials, and Sustainable Engineering (EGRWSE-2025)*, Vigo, Spain. Conference presentation.

Faizi, H. and A. El-Zein (2025c). Predicting soil swelling stabilisation using agro-industrial by-products: Comparing deep learning with genetic programming. In *Proceedings of the 3rd International Conference on Geomechanics and Geoenvironmental Engineering (iCGMGE)*. Conference presentation.

Faizi, H. and A. El-Zein (2025d). Tree-based ensembles and genetic programming for predicting strength of aluminosilicate waste-stabilised soils using an expanded dataset. *Neural Computing and Applications*. Submitted, under review.

Gireesh Kumar, P. and S. Harika (2021). Stabilization of expansive subgrade soil by using fly ash. *Materials Today: Proceedings 45*, 6558–6562.

Géron, A. O. (2019, June). *Hands-on machine learning with Scikit-Learn and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (Second ed.), Volume 1. Beijing and Boston and Farnham and Sebastopol and Tokyo: O'Reilly.

Habert, G., J. d'Espinose de Lacaillerie, and N. Roussel (2011, July). An environmental evaluation of geopolymer based concrete production: reviewing current research trends. *Journal of Cleaner Production 19*(11), 1229–1238.

Hanandeh, S., A. Ardah, and M. Abu-Farsakh (2020, September). Using artificial neural network and genetics algorithm to estimate the resilient modulus for stabilized subgrade and propose new empirical formula. *Transportation Geotechnics 24*, 100358.

Hastie, T., R. Tibshirani, and J. Friedman (2009, August). *The Elements of Statistical Learning* (2 ed.), Volume 1 of *2197-568X*. New York, USA: Springer.

Hebib, S. and E. R. Farrell (2003, February). Some experiences on the stabilization of Irish peats. *Canadian Geotechnical Journal 40*(1), 107–120.

Herman, J. and W. Usher (2017, jan). SALib: An open-source python library for sensitivity analysis. *The Journal of Open Source Software 2*(9), 97.

Hoy, M., S. Horpibulsuk, and A. Arulrajah (2016, August). Strength development of Recycled Asphalt Pavement – Fly ash geopolymer as a road construction material. *Construction and Building Materials 117*, 209–219.

Ishibashi, I. and H. Hazarika (2015). *Soil Mechanics Fundamentals and Applications*. CRC Press.

Iwanaga, T., W. Usher, and J. Herman (2022, May). Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environmental Systems Modelling 4*, 18155.

Jiao, W., W. Zhou, Z. Huang, R. Lan, and M. Ma (2024, April). Study on static and dynamic mechanical behavior of expansive soil modified by oyster shell powder. *Heliyon 10*(8), e29699.

Jones, D., A. Rahim, S. Saadeh, and J. Harvey (2010, July). Guidelines for the Stabilization of Subgrade Soils in California. Technical Report UCPRC-GL-2010-01, University of California Pavement Research Center; California Department of Transportation, Davis and Berkeley, CA; Sacramento, CA.

Karami, H., J. Pooni, D. Robert, S. Costa, J. Li, and S. Setunge (2021, July). Use of secondary additives in fly ash based soil stabilization for soft subgrades. *Transportation Geotechnics 29*, 100585.

Khadka, S. D., P. W. Jayawickrama, S. Senadheera, and B. Segvic (2020, June). Stabilization of highly expansive soils containing sulfate using metakaolin and fly ash based geopolymer modified with lime and gypsum. *Transportation Geotechnics 23*, 100327.

Khanday, S. A., M. Hussain, and A. K. Das (2021, December). A Review on Chemical Stabilization of Peat. *Geotechnical and Geological Engineering 39*(8), 5429–5443.

Khazaei, J. and H. Moayedi (2019, October). Soft Expansive Soil Improvement by Eco-Friendly Waste and Quick Lime. *Arabian Journal for Science and Engineering 44*(10), 8337–8346.

Kohavi, R. (1995). Proceedings of the fourteenth international joint conference on artificial intelligence. Morgan Kaufmann. Meeting Name: IJCAI.

Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection*. Complex adaptive systems. Cambridge, Mass: MIT Press.

Lakkimsetti, B. and S. Nayak (2022, February). Experimental Investigation and Performance Evaluation of Lithomargic Clay Stabilized with Granulated Blast Furnace Slag and Calcium Chloride. *International Journal of Geosynthetics and Ground Engineering 8*(1), 12.

Leong, H. Y., D. E. L. Ong, J. G. Sanjayan, and A. Nazari (2016, March). The effect of different Na 2 O and K 2 O ratios of alkali activator on compressive strength of fly ash based-geopolymer. *Construction and Building Materials 106*, 500–511.

Leong, H. Y., D. E. L. Ong, J. G. Sanjayan, and A. Nazari (2018). Strength Development of Soil–Fly Ash Geopolymer: Assessment of Soil, Fly Ash, Alkali Activators, and Water. *Journal of Materials in Civil Engineering 30*(8), 63.

Leong, H. Y., D. E. L. Ong, J. G. Sanjayan, A. Nazari, and S. M. Kueh (2018, July). Effects of Significant Variables on Compressive Strength of Soil-Fly Ash Geopolymer: Variable Analytical Approach Based on Neural Networks and Genetic Programming. *Journal of Materials in Civil Engineering 30*(7), 04018129.

Li, S., J. Yin, and G. Zhang (2020, July). Experimental investigation on optimization of vegetation performance of porous sea sand concrete mixtures by pH adjustment. *Construction and Building Materials 249*, 118775.

Liu, Y., D. E. L. Ong, E. Oh, Z. Liu, and R. Hughes (2022, feb). Sustainable cementitious blends for strength enhancement of dredged mud in Queensland, Australia. *Geotechnical Research 9*(2), 65–82.

Lundberg, S. and S.-I. Lee (2017, November). A unified approach to interpreting model predictions.

Mahedi, M., B. Cetin, and A. Y. Dayioglu (2019, July). Leaching behavior of aluminum, copper, iron and zinc from cement activated fly ash and slag stabilized soils. *Waste Management 95*, 334–355.

Mandal, S. and J. P. Singh (2007). Stabilization of Soil using Ground Granulated Blast Furnace Slag and Fly Ash. *International Journal of Innovative Research in Science, Engineering and Technology 5*(12), 6.

Marvila, M. T., A. R. G. d. Azevedo, and C. M. F. Vieira (2021). Reaction mechanisms of alkali-activated materials. *Revista IBRACON de Estruturas e Materiais 14*(3), e14309.

Mathiluxsan, S., S. Venuja, and M. Nasvi (2016, 12). Compressibility behaviour of peat stabilized with low calcium fly ash: An experimental study.

Millán Ramírez, G. P., H. Byliński, and M. Niedostatkiewicz (2021, June). Deterioration and Protection of Concrete Elements Embedded in Contaminated Soil: A Review. *Materials 14*(12), 3253.

Miranda, T., D. Leitão, J. Oliveira, M. Corrêa-Silva, N. Araújo, J. Coelho, A. Fernández-Jiménez, and N. Cristelo (2019). Application of alkali-activated industrial wastes for the stabilisation of a full-scale (sub)base layer. *Journal of Cleaner Production 242*, 118427.

Mozumder, R. A. and A. I. Laskar (2015, September). Prediction of unconfined compressive strength of geopolymer stabilized clayey soil using Artificial Neural Network. *Computers and Geotechnics 69*, 291–300.

Munawar, M. A., A. H. Khoja, S. R. Naqvi, M. T. Mehran, M. Hassan, R. Liaquat, and U. F. Dawood (2021, October). Challenges and opportunities in biomass ash management and its utilization in novel applications. *Renewable and Sustainable Energy Reviews 150*, 111451.

Murmu, A. L., N. Dhole, and A. Patel (2020, April). Stabilisation of black cotton soil for subgrade application using fly ash geopolymer. *Road Materials and Pavement Design 21*(3), 867–885.

Mypati, V. N. K. and S. Saride (2022, January). Feasibility of Alkali-Activated Low-Calcium Fly Ash as a Binder for Deep Soil Mixing. *Journal of Materials in Civil Engineering 34*(1), 04021410.

Nalbantoğlu, Z. (2004, July). Effectiveness of Class C fly ash as an expansive soil stabilizer. *Construction and Building Materials 18*(6), 377–381.

Navin, P. K., S. Kumar, K. Murari, and M. Mathur (2018). A Review on the Utilization and Kinetics of Fly Ash. *International Journal of Engineering Research 6*(11), 5.

Nelson, J. D., K. C. Chao, D. D. Overton, and E. J. Nelson (2015). *Foundation engineering for expansive soils*. Hoboken, New Jersey: John Wiley and Sons, Inc.

Nissanka, N. A. N. M., K. M. D. Nimesha, and M. C. M. Nasvi (2021). Prediction of Geotechnical Properties of Stabilized Soil Using Fly Ash-Based Stabilizer Systems. In R. Dissanayake, P. Mendis, K. Weerasekera, S. De Silva, S. Fernando, and C. Konthesingha (Eds.), *12th International Conference on Structural Engineering and Construction Management*, Volume 266, pp. 297–314. Singapore: Springer Nature Singapore. Series Title: Lecture Notes in Civil Engineering.

Nordmark, D., J. Vestin, L. Hansson, and J. Kumpiene (2022, September). Long-term evaluation of geotechnical and environmental properties of ash-stabilised road. *Journal of Environmental Management 318*, 115504.

Odeh, N. A. and A. H. Al-Rkaby (2022, June). Strength, Durability, and Microstructures characterization of sustainable geopolymer improved clayey soil. *Case Studies in Construction Materials 16*, e00988.

Oderji, S. Y., B. Chen, C. Shakya, M. R. Ahmad, and S. F. A. Shah (2019, December). Influence of superplasticizers and retarders on the workability and strength of one-part alkali-activated fly ash/slag binders cured at room temperature. *Construction and Building Materials 229*, 116891.

Onyelowe, K. C., F. F. Mojtahedi, A. M. Ebid, A. Rezaei, K. J. Osinubi, A. O. Eberemu, B. Salahudeen, E. W. Gadzama, D. Rezazadeh, H. Jahangir, P. Yohanna, M. E. Onyia, F. E. Jalal, M. Iqbal, C. Ikpa, I. I. Obianyo, and Z. U. Rehman (2023, December). Selected AI optimization techniques and applications in geotechnical engineering. *Cogent Engineering 10*(1), 2153419.

Parthiban, D., D. S. Vijayan, E. Koda, M. D. Vaverkova, K. Piechowicz, P. Osinski, and B. V. Duc (2022, June). Role of industrial based precursors in the stabilization of weak soils with geopolymer – A review. *Case Studies in Construction Materials 16*, e00886.

Patel, A. (2019). Case examples of some geotechnical applications. In *Geotechnical Investigations and Improvement of Ground Conditions*, pp. 167–191. Elsevier.

Pham, V., E. Oh, and D. Ong (2022, 06). Effects of binder types and other significant variables on the unconfined compressive strength of chemical-stabilized clayey soil using gene-expression programming. *Neural Computing and Applications 34*, 9103—-9121.

Pham, V.-N., E. Oh, and D. E. L. Ong (2021, December). Gene-Expression Programming-Based Model for Estimating the Compressive Strength of Cement-Fly Ash Stabilized Soil and Parametric Study. *Infrastructures 6*(12), 181.

Phanikumar, B. R. (2022, February). Behaviour of expansive clay beds with fly-ash-stabilised clay cushion. *Proceedings of the Institution of Civil Engineers - Ground Improvement 175*(1), 73–81.

Phetchuay, C., S. Horpibulsuk, A. Arulrajah, C. Suksiripattanapong, and A. Udomchai (2016, July). Strength development in soft marine clay stabilized by fly ash and calcium carbide residue based geopolymer. *Applied Clay Science 127-128*, 134–142.

Phummiphan, I., S. Horpibulsuk, R. Rachan, A. Arulrajah, S.-L. Shen, and P. Chindaprasirt (2018, January). High calcium fly ash geopolymer stabilized lateritic soil and granulated blast furnace slag blends as a pavement base material. *Journal of Hazardous Materials 341*, 257–267.

Phummiphan, I., S. Horpibulsuk, P. Sukmak, A. Chinkulkijniwat, A. Arulrajah, and S.-L. Shen (2016, October). Stabilisation of marginal lateritic soil using high calcium fly ash-based geopolymer. *Road Materials and Pavement Design 17*(4), 877–891.

Pourakbar, S., A. Asadi, B. B. Huat, and M. H. Fasihnikoutalab (2015, June). Stabilization of clayey soil using ultrafine palm oil fuel ash (POFA) and cement. *Transportation Geotechnics 3*, 24–35.

Pratama, I. T., A. Y. Arif, and B. Widjaja (2021). Linear Shrinkage and its Correlation to the Shrinkage Limit and Index Properties of Kaolinite, Bentonite, and Eight Bandung Fine- Grained Soils. *25nd Annual National Conference on Geotechnical Engineering*, 7.

Provis, J. L. and S. A. Bernal (2014, July). Geopolymers and Related Alkali-Activated Materials. *Annual Review of Materials Research 44*(1), 299–327.

Puppala, A. J., A. Pedarla, and T. Bheemasetti (2015). Soil Modification by Admixtures. In *Ground Improvement Case Histories*, pp. 291–309. Elsevier.

Ram, A. K. and S. Mohanty (2022, December). State of the art review on physiochemical and engineering characteristics of fly ash and its applications. *International Journal of Coal Science & Technology 9*(1), 9.

Renjith, R., D. Robert, S. Setunge, S. Costa, and A. Mohajerani (2021, April). Optimization of fly ash based soil stabilization using secondary admixtures for sustainable road construction. *Journal of Cleaner Production 294*, 126264.

Rios, S., N. Cristelo, A. Viana da Fonseca, and C. Ferreira (2016, February). Structural Performance of Alkali-Activated Soil Ash versus Soil Cement. *Journal of Materials in Civil Engineering 28*(2), 04015125.

Rogers, C. D. F. and S. Glendinning (1996). Modification of Clay Soils Using Lime. In *Lime Stabilisation*, pp. 99–114. Published Online. _eprint: https://www.icevirtuallibrary.com/doi/pdf/10.1680/ls.25639.0010.

Roy, W., R. Thiery, R. Schuller, and J. Suloway (1981, April). Coal fly ash : a review of the literature and proposed classification system with emphasis on environmental impacts. Technical report, Illinois Institute of Natural Resources.

Sargent, P. (2015). The development of alkali-activated mixtures for soil stabilisation. In *Handbook of Alkali-Activated Cements, Mortars and Concretes*, pp. 555–604. Elsevier.

Sari Ahmed, B., H. Gadouri, M. Ghrici, and K. Harichane (2020, June). Best-fit models for predicting the geotechnical properties of FA–stabilised problematic soils used as materials for earth structures. *International Journal of Pavement Engineering 21*(7), 939–953.

Saride, S. and V. N. Mypati (2024, February). Effect of area improvement ratio of geopolymer-based deep mixing columns on swell-shrink behavior of expansive soils. *Construction and Building Materials 417*, 135163.

Sata, V. and P. Chindaprasirt (2020). 19 - Use of construction and demolition waste (CDW) for alkali-activated or geopolymer concrete. In F. Pacheco-Torgal, Y. Ding, F. Colangelo, R. Tuladhar, and A. Koutamanis (Eds.), *Advances in Construction and Demolition Waste Recycling*, Woodhead Publishing Series in Civil and Structural Engineering, pp. 385–403. Woodhead Publishing.

Schaefer, V. R., J. K. Mitchell, R. R. Berg, G. M. Filz, and S. C. Douglas (2012, May). Ground Improvement in the 21st Century: A Comprehensive Web-Based Information System. In *Geotechnical Engineering State of the Art and Practice*, Oakland, California, United States, pp. 272–293. American Society of Civil Engineers.

Seed, H. B. and R. Lundgren (1962). Prediction of Swelling Potential for Compacted Clays. *Journal of the Soil Mechanics and Foundations Division 88*, 53–88.

Shah, S. F. A., B. Chen, S. Y. Oderji, M. A. Haque, and M. R. Ahmad (2020, June). Improvement of early strength of fly ash-slag based one-part alkali activated mortar. *Construction and Building Materials 246*, 118533.

Sharma, A. and R. K. Sharma (2021, March). Sub-grade characteristics of soil stabilized with agricultural waste, constructional waste, and lime. *Bulletin of Engineering Geology and the Environment 80*(3), 2473–2484.

Sharma, K. and A. Kumar (2020, December). Utilization of industrial waste—based geopolymers as a soil stabilizer—a review. *Innovative Infrastructure Solutions 5*(3), 97.

Singhi, B., A. I. Laskar, and M. A. Ahmed (2016, February). Investigation on Soil–Geopolymer with Slag, Fly Ash and Their Blending. *Arabian Journal for Science and Engineering 41*(2), 393–400.

Sithole, N. T. and T. Mashifana (2020, December). Geosynthesis of building and construction materials through alkaline activation of granulated blast furnace slag.

Construction and Building Materials 264, 120712.

Skempton, A. W. (1984, January). The Colloidal "Activity" of Clays. *The Third International Conference of Soil Mechanics and Foundation Engineering 1*, 57–61.

Smith, M. (1993). *Neural networks for statistical modeling*. New York Van Nostrand Reinhold. Includes bibliographical references and index.

Song, M., S. Lin, and F. Takahashi (2020, May). Coal fly ash amendment to mitigate soil water evaporation in arid/semi-arid area: An approach using simple drying focusing on sieve size and temperature. *Resources, Conservation and Recycling 156*, 104726.

Sposito, G. (2016, December). Soil Minerals. In *The Chemistry of Soils*. Oxford University Press.

Standards Australia (2000, January). Australian standard as 1289.0-2000 methods of testing soils for engineering purposes: General requirements and list of methods. Accessed: 2023-05-04. Available at: `https://www.standards.org.au/`.

Standards Australia Limited (2007). *AS 3798 - 2007: Guidelines on earthworks for commercial and residential developments.* Sydney, N.S.W.: Standards Australia. OCLC: 271531101.

Sukprasert, S., M. Hoy, S. Horpibulsuk, A. Arulrajah, A. S. A. Rashid, and R. Nazir (2021, February). Fly ash based geopolymer stabilisation of silty clay/blast furnace slag for subgrade applications. *Road Materials and Pavement Design 22*(2), 357–371.

Suksiripattanapong, C., S. Horpibulsuk, C. Yeanyong, and A. Arulrajah (2021, March). Evaluation of polyvinyl alcohol and high calcium fly ash based geopolymer for the improvement of soft Bangkok clay. *Transportation Geotechnics 27*, 100476.

Taheri, S. (2023). Resistance to desiccation cracking in clay-polymer mixtures.

Tastan, E. O., T. B. Edil, C. H. Benson, and A. H. Aydilek (2011, September). Stabilization of Organic Soils with Fly Ash. *Journal of Geotechnical and Geoenvironmental Engineering 137*(9), 819–833.

Tiwari, N. and N. Satyam (2021, October). Coupling effect of pond ash and polypropylene fiber on strength and durability of expansive soil subgrades: An integrated experimental and machine learning approach. *Journal of Rock Mechanics and Geotechnical Engineering 13*(5), 1101–1112.

Toksöz Hozatlıoğlu, D. and I. Yılmaz (2021, January). Shallow mixing and column performances of lime, fly ash and gypsum on the stabilization of swelling soils. *Engineering Geology 280*, 105931.

Toufigh, V., S. Ghasemalizadeh, and M. Karamian (2022, April). Experimental Investigation of Mixture Design and Durability Performance of Alkali-Activated Rammed Earth. *International Journal of Geomechanics 22*(4), 04022029.

Transport for NSW (2022). QA Specification R44: Earthworks.

Turkane, S. D. and S. K. Chouksey (2022). Design of low volume road pavement of stabilized low plastic soil using fly ash geopolymer. *Materials Today: Proceedings 65*, 1154–1160.

Turton, J. L., G. D. Brinkworth, H. M. Parker, D. Lim, K. Lee, A. Rush, R. Johnson, and K. B. Rooney (2023). Effects of a low-carbohydrate diet in adults with type 1 diabetes management: A single arm non-randomised clinical trial. *PLOS ONE 18*(7), e0288440. Publisher: Public Library of Science (PLoS).

U.S. EPA (2018, March). 2018 Edition of the Drinking Water Standards and Health Advisories Tables. EPA Technical Report EPA 822-F-18-001, U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Varma, N., T. Kumar, and V. Nagaraju (2021). Compressive Strength of High Plastic Clay Stabilized with Fly Ash-Based Geopolymer and Its Synthesis Parameters. In C. N. V. Satyanarayana Reddy, S. Saride, and S. Haldar (Eds.), *Transportation, Water and Environmental Geotechnics*, Volume 159, pp. 25–37. Singapore: Springer Singapore. Series Title: Lecture Notes in Civil Engineering.

Vijayan, D. and D. Parthiban (2020, November). Effect of Solid waste based stabilizing material for strengthening of Expansive soil- A review. *Environmental Technology & Innovation 20*, 101108.

Vorwerk, S., D. Cameron, and G. Keppel (2015). Clay Soil in Suburban Environments. In *Ground Improvement Case Histories*, pp. 655–682. Elsevier.

Vu, M. C., T. Satomi, and H. Takahashi (2020, February). Influence of initial water, moisture, and geopolymer content on geopolymer modified sludge. *Construction and Building Materials 235*, 117420.

Wan, X.-m., F. H. Wittmann, T.-j. Zhao, and H. Fan (2013, January). Chloride content and pH value in the pore solution of concrete under carbonation. *Journal of Zhejiang University SCIENCE A 14*(1), 71–78.

Wardhono, A. (2018, January). Comparison Study of Class F and Class C Fly Ashes as Cement Replacement Material on Strength Development of Non-Cement Mortar. *IOP Conference Series: Materials Science and Engineering 288*, 012019.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin 1*(6), 80–83.

Xu, Y., Y. Zhou, P. Sekula, and L. Ding (2021, May). Machine learning in construction: From shallow to deep learning. *Developments in the Built Environment 6*, 100045.

Yaghoubi, M., A. Arulrajah, M. M. Disfani, S. Horpibulsuk, M. W. Bo, and S. Darmawan (2018, June). Effects of industrial by-product based geopolymers on the strength development of a soft soil. *Soils and Foundations 58*(3), 716–728.

Yang, T., X. Yao, and Z. Zhang (2014, October). Quantification of chloride diffusion in fly ash–slag-based geopolymers by X-ray fluorescence (XRF). *Construction and Building Materials 69*, 109–115.

Yao, Z., X. Ji, P. Sarker, J. Tang, L. Ge, M. Xia, and Y. Xi (2015, February). A comprehensive review on the applications of coal fly ash. *Earth-Science Reviews 141*, 105–121.

Zeini, H. A., D. Al-Jeznawi, H. Imran, L. F. A. Bernardo, Z. Al-Khafaji, and K. A. Ostrowski (2023, January). Random Forest Algorithm for the Strength Prediction of Geopolymer Stabilized Clayey Soil. *Sustainability 15*(2), 1408.

Zha, F., C. Hu, B. Kang, L. Qin, J. Li, and C. Chu (2024, February). Formulation of PG-FA-L composite modifier for repairing expansive soil based on the statistical mixed design method. *Chemosphere 349*, 140974.

Zhang, P., Z.-Y. Yin, Y.-F. Jin, T. H. Chan, and F.-P. Gao (2021, January). Intelligent modelling of clay compressibility using hybrid meta-heuristic and machine learning algorithms. *Geoscience Frontiers 12*(1), 441–452.

Zhang, W., H. Li, Y. Li, H. Liu, Y. Chen, and X. Ding (2021, December). Application of deep learning algorithms in geotechnical engineering: a short critical review. *Artificial Intelligence Review 54*(8), 5633–5673.

Zhang, X., M. Trame, L. Lesko, and S. Schmidt (2015, February). Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. *CPT: Pharmacometrics & Systems Pharmacology 4*(2), 69–79.

Zhu, J., A. El-Zein, D. W. Airey, and G. Miao (2024, January). An experimental study on root-reinforced soil strength via a steel root analogue in unsaturated silty soil. *Acta Geotechnica 19*(1), 255–272.

Zhuang, H., X. Wang, M. Bendersky, and M. Najork (2020, July). Feature Transformation for Neural Ranking Models. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event China, pp. 1649–1652. ACM.

Ünsever, Y. and M. L. Diallo (2019, 07). Stabilization of clay soils using fly ash. *Black Sea Journal of Engineering and Science 2*, 10–11.

APPENDIX A

MATHEMATICAL EQUATIONS OF THE PREVIOUS MODELS

A1 Conventional Regression Models

Model RM-Moz

$$
\begin{aligned}
UCS = & -19.383 + 0.677 \times (LL\_0) - 0.803 \times (PI\_0) + 0.446 \times (slag\%) \\
& - 0.159 \times (ash\%) + 0.010 \times (M) + 1.559 \times (A\_to\_B) + 1.929 \\
& \times (Na\_to\_Al\_mix) + 1.835 \times (Si\_to\_Al\_mix)
\end{aligned}
\tag{A.1}
$$

Model RM-Con

$$
\begin{aligned}
UCS = & \lambda(a + b \times \text{ash\%}) \\
& \times \left[1 - \left(\frac{DUW\_mix}{1 + vol\_lime}\right) \times (c_1 + c_2 \times \text{ash\%} + c_3 \times vol\_lime)\right]^{-\alpha} \\
& \times \left[\frac{DUW\_mix \times vol\_lime}{1 + vol\_lime}\right]^{\alpha\beta}
\end{aligned}
\tag{A.2}
$$

Equation A.2 estimates UCS in kPa under a curing temperature of 23°C. The volumetric lime content, *vol_lime*, used in the equation is calculated using Equation A.3:

$$
vol\_lime = 100 \left[\frac{DUW\_mix}{1 + (\frac{\text{lime\%}}{100})}\right] \left[\frac{\text{lime\%}}{100 \times DUW\_lime}\right]
\tag{A.3}
$$

The constants used in Equation A.2 are as follows:

$$
c_1 = 0.0379 \qquad c_2 = -0.0043 \qquad c_3 = 0.0402
$$

$$
\alpha = 4.6 \qquad \beta = 0.12 \qquad \lambda = \left(\frac{100}{26.4}\right)^{\alpha\beta} \times 100^{-\alpha}
$$

The effect of curing time in the RM-Con model is incorporated through the coefficients a and b, which vary depending on the duration of curing. For a 28-day curing period, the model uses $a = 0$ and $b = 5.89 \times 10^8$. For 60 days, a remains at 0, while b increases to 9.15×10^8. For longer curing periods, both coefficients increase: at 90 days, $a = 7.42 \times 10^8$ and $b = 14.1 \times 10^8$; at 180 days, $a = 12.8 \times 10^8$ and $b = 15.3 \times 10^8$; and for 360-day strength predictions, the values rise to $a = 31.42 \times 10^8$ and $b = 15.4 \times 10^8$.

Models RM-Sar-1 to RM-Sar-6

$$LL = \frac{LL\_0}{(1+\sqrt{0.00133 \times ash\%})} \tag{A.4}$$

$$PI = \frac{PI\_0}{(1+\sqrt{0.00376 \times ash\%})} \tag{A.5}$$

$$UCS = \frac{UCS\_0}{(1+ln(0.114 \times ash\% + 1))^{-a}}, \tag{A.6}$$

where,

$$a = log_{10}(0.278 \times t\_curing + 1) + 1$$

$$FSI = FSI\_0 \times e^{(-0.022 \times \text{ash\%})} \tag{A.7}$$

$$MDD = \frac{1}{(1+0.00263 \times ash\%)} \times MDD\_0 \tag{A.8}$$

$$OMC = (1+0.00386 \times ash\%) \times OMC\_0 \tag{A.9}$$

A2 Evolutionary Methods

Model GP-Leo

$$
\begin{aligned}
UCS = & \frac{ash\% - soil\%}{\left[(0.84 \times soil\% - ash\%) + \frac{\sqrt[3]{water\%}}{SS\_to\_SH}\right]} \\
& \times \left[(exp\sqrt[3]{0.89 \times ash\%}) + (\sqrt[3]{0.89 - SH\%}) \times (A\_to\_ash \times water\% - 0.45)\right] \\
& \times \sqrt[3]{\left[(SH\% \times water\% \times \sqrt[3]{SS\%}) + \left(\left[\frac{ash\%}{-5.48}\right] \times (ash\% + SS\_to\_SH)\right)\right]} \\
& \times \frac{A\_to\_ash}{2 \times SS\% + \left[\frac{1.51 - SS\_to\_SH}{4.51 - water\%}\right]}
\end{aligned}
\tag{A.10}
$$

Model GEP-Han

$$
\begin{aligned}
M_{\mathrm{r}} = & \, log\lfloor\lfloor e^{a}\rfloor\rfloor + \frac{1}{\left[\frac{-7.59 + ash\%}{2}\right]} + lime\% + log\left[\frac{OMC\_0}{clay\%}\right] \\
& + \sqrt[3]{clay\% \times MC\_0 - (clay\%)^2} \times cement \times \left[\frac{-6.388}{MC\_0}\right],
\end{aligned}
\tag{A.11}
$$

where,

$$
a = \left[\frac{\frac{clay\%}{min(MC\_0, PI\_0)}}{6.39 - cement\%}\right]
$$

Model GEP-Pha-1

$$
\begin{aligned}
UCS = & \, (d_0 \times d_4)^{1/2} \times (d_7 - d_8 - c_{18})^{1/3} + d_2 \times d_{11} \\
& \times \left[\frac{d_7 \times ({d_9}^2 - c_{23} - c_{28}}{d_3 \times c_{26}}\right]^{1/2} + d_7 \times d_{12} \times \left[\frac{d_4 - 1}{d_4}\right] \\
& + d_8 \times \left[d_2 + \frac{{c_{56}}^{1/2} \times (d_7 + d_9 - d_8 - d_{11})}{d_1 - d_8}\right] - \frac{d_6}{d_4} \\
& - 2c_{70} + d_0 \times d_8 \times d_5 + 5d_1 + d_3 - d_5 - 6d_6 - 2d_7 \\
& + 3d_8 + 4d_9 + d_{10} + d_{11} - 6d_{12} + c_{74}
\end{aligned}
\tag{A.12}
$$

In Equation A.12, UCS is calculated in kPa. The values for the *'c'* constants include: c_{18} = -17.103; c_{23} = -6.696; c_{28} = -3.451; c_{26} = 6.617; c_{56} = 477.882; c_{74} = 8.646; and c_{70} = -11.715. The predictors used are denoted as *'d'* and they include: d_0: *PL_0*; d_1: *LL_0*; d_2: *clay%*; d_3: *silt%*; d_4: *sand%*; d_5: *organic%*; d_6: *water%*; d_7: *t_curing*; d_8: *cement%*; d_9: *ash%*; d_{10}: *CaO_ash%*; d_{11}: *Ca_to_SiO_ash*; d_{12}: *LOI_ash*.

Model GEP-Pha-2

$$\begin{aligned} UCS = {} & d_0(d_0-2)+d_3+3(d_5-d_4-d_{10}) \\ & -2(d_1+d_2-d_7)+d_6(d_2+d_6+2)+d_5\times d_9 \\ & +2c_{59}-\left[\frac{e^{c_{51}-c_{55}-c_{56}}}{c_{59}}\right]-e^{c_{51}}+d_7\times c_{21}\times ln(d_5) \\ & +d_8(d_4\times d_3)^{1/3}+c_{50}+c_{64} \\ & +c_{40}\times ln\left[\left(\frac{{d_5}^{d_{10}}}{d_4}+{d_4}^2+e^{d_9}\right)^{-1/2}\right] \\ & +d_7\left[\frac{d_3}{c_{72}+d_4+c_{78}}\right]\left[2\times d_2+\frac{d_4}{c_{78}}\right]+c_{10} \end{aligned} \tag{A.13}$$

Equation A.13 estimates *UCS* in kPa. The values for the *'c'* constants include: c_{10} = -9.204; c_{21} = 9.910; c_{40} = -15.170; c_{59} = -9.800; c_{51} = 5.616; c_{50} = -9.6502; c_{55} = 8.759; c_{56} = -5.498; c_{64} = -9.812; c_{72} = -7.635; c_{78} = -3.177. Model predictors are denoted as *'d'* and they include: d_0: *PI_0*; d_1: *clay%*; d_2: *silt%*; d_3: *sand%*; d_4: *water%*; d_5: *t_curing*; d_6: *lime%*; d_7: *cement%*; d_8: *ash_C%*; d_9: *ash_F%*; d_{10}: *slag%*.

APPENDIX B

DETAILS OF THE PROPOSED NEW GENETIC PROGRAMMING MODEL FOR STRENGTH PREDICTION

Expression tree representations of the Genetic Programming models for strength prediction:

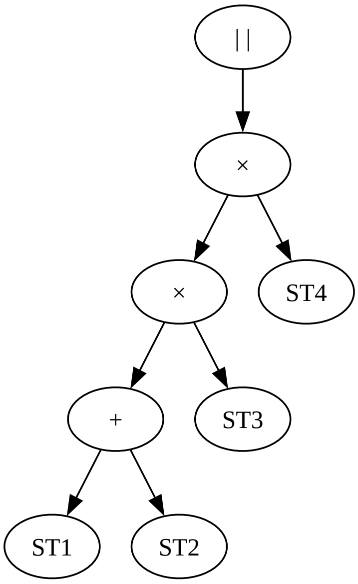
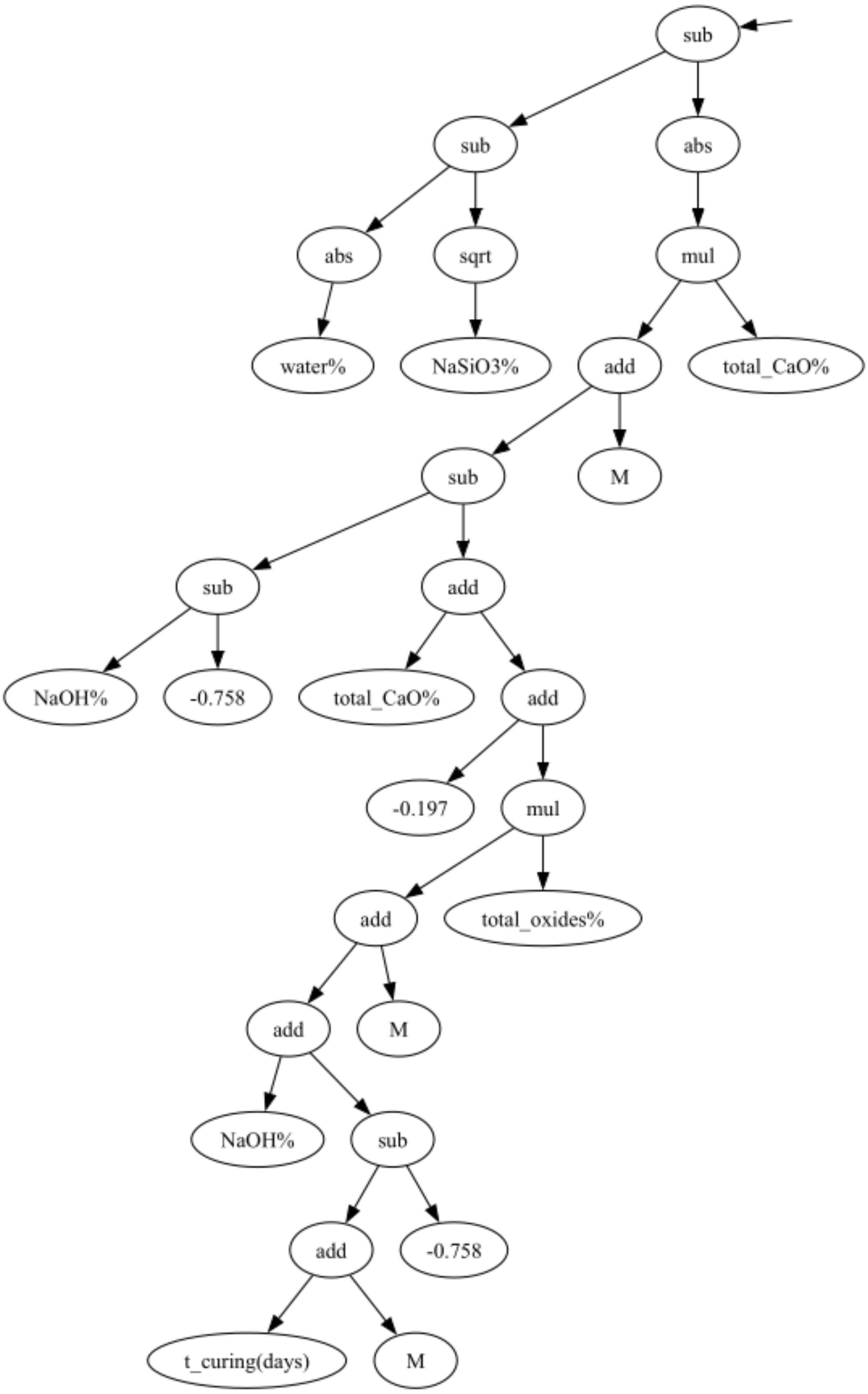
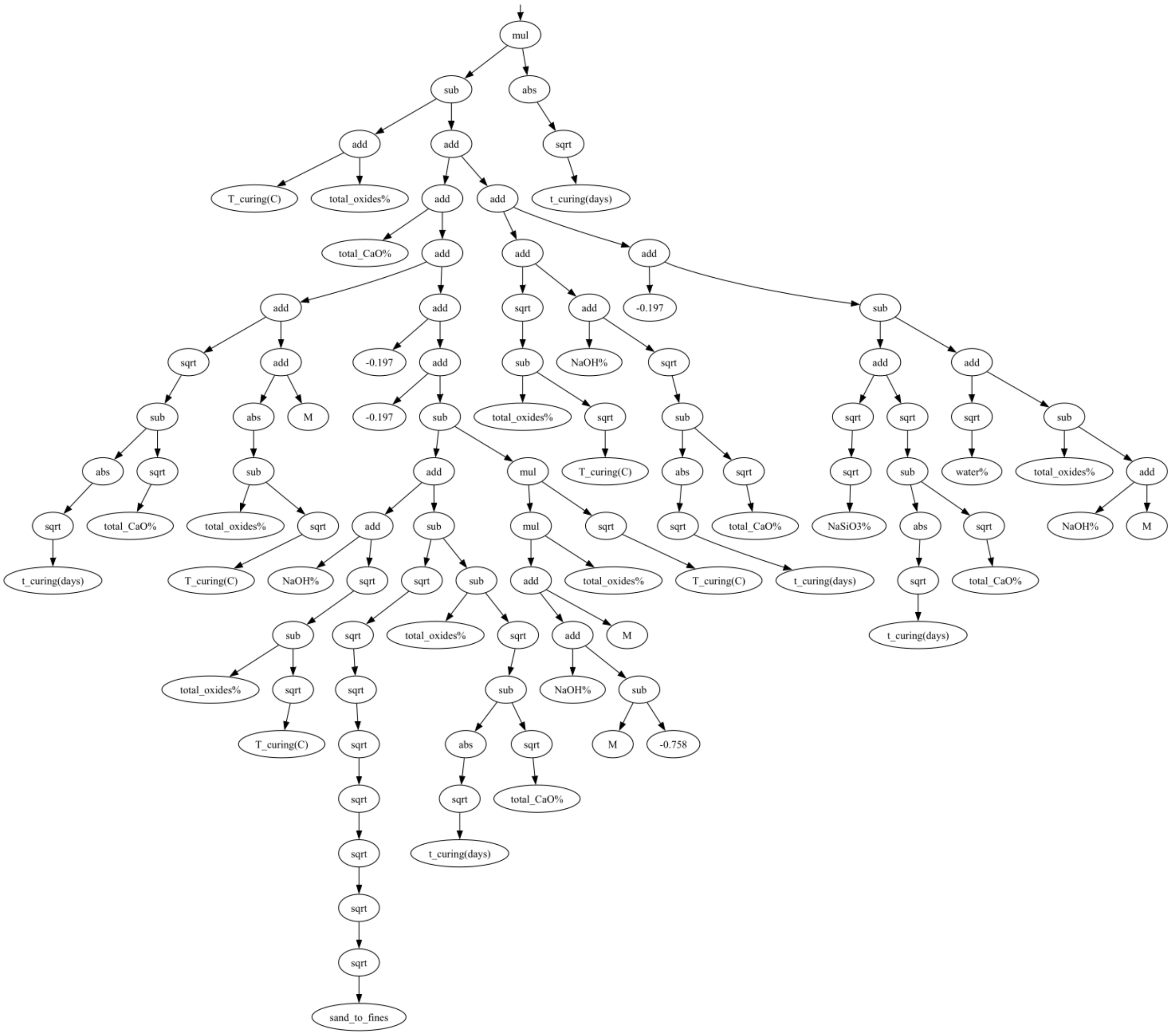


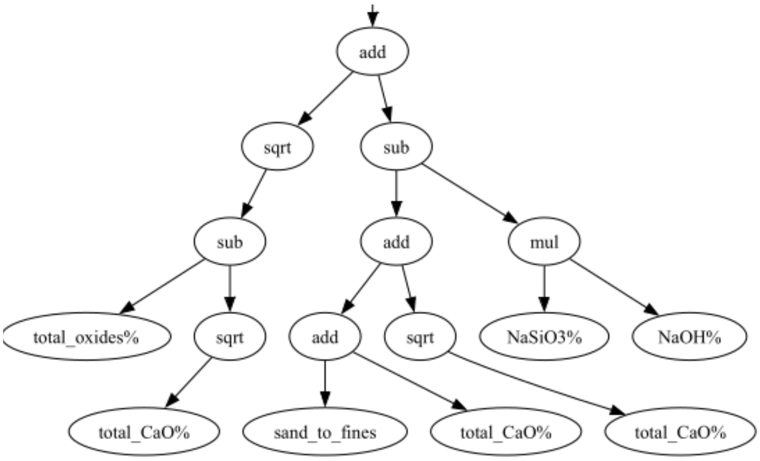
Figure B.1. The optimised expression tree segmented into 4 sub-trees (ST1 to ST4)


sub
sub
abs
abs
sqrt
mul
water%
NaSiO3%
add
total_CaO%
sub
M
sub
add
NaOH%
-0.758
total_CaO%
add
-0.197
mul
add
total_oxides%
add
M
NaOH%
sub
add
-0.758
t_curing(days)
M

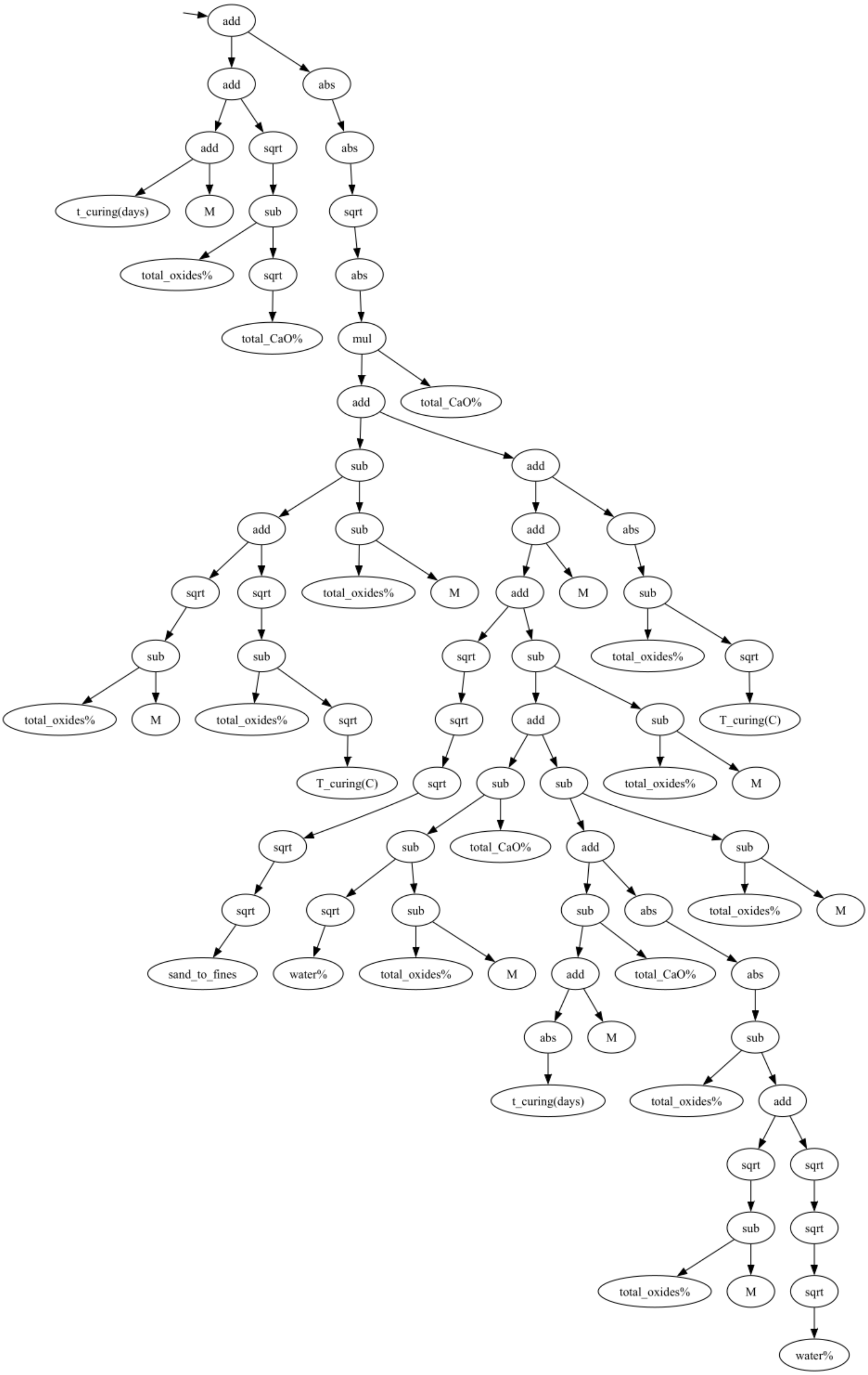
Sub-tree 1



Sub-tree 2



Sub-tree 3


add
add
abs
add
sqrt
abs
t_curing(days)
M
sub
sqrt
total_oxides%
sqrt
abs
total_CaO%
mul
add
total_CaO%
sub
add
add
sub
add
abs
sqrt
sqrt
total_oxides%
M
add
M
sub
sub
sub
sqrt
sub
total_oxides%
sqrt
total_oxides%
M
total_oxides%
sqrt
sqrt
add
sub
T_curing(C)
T_curing(C)
sqrt
sub
sub
total_oxides%
M
sqrt
sub
total_CaO%
add
sub
sqrt
sqrt
sub
sub
abs
total_oxides%
M
sand_to_fines
water%
total_oxides%
M
add
total_CaO%
abs
abs
M
sub
t_curing(days)
total_oxides%
add
sqrt
sqrt
sub
sqrt
total_oxides%
M
sqrt
water%

Sub-tree 4

Mathematical representations of the above expression trees:

$$\text{UCS (MPa)} \;=\; \left|\left(\text{ST}_1 + \text{ST}_2\right) \times \text{ST}_3 \times \text{ST}_4\right| \tag{B.1}$$

$$\text{ST1} = |w| \;-\; \sqrt{N} \;-\; \Big|C \times \left[\, M + \text{Na} + 0.561 - C + O \times (2M + \text{Na} + t + 0.758)\right]\Big| \tag{B.2}$$

$$\begin{aligned}
\text{ST2} = \sqrt{t}\,\Big[&T - C + t^{1/4} - \sqrt{C} + 2M + \left|O - \sqrt{T}\right| - 0.591 + \text{Na} + \sqrt{O + \sqrt{T}} \\
&+ s^{1/256} - \sqrt{\sqrt{t} - \sqrt{C}} - \sqrt{T}\,\left[\, O\,(2M + \text{Na} + 0.758)\right] + \sqrt{O - \sqrt{T}} \\
&+ 2\sqrt{\sqrt{t} + \sqrt{C}} + N^{1/4} - \sqrt{w}\Big]
\end{aligned} \tag{B.3}$$

$$\text{ST3} = \sqrt{O - \sqrt{C}} \;+\; s \;+\; C \;+\; \sqrt{C} \;-\; N \times \text{Na} \tag{B.4}$$

$$\begin{aligned}
\text{ST4} = {}& t + M + \sqrt{O - \sqrt{C}} \\
&+ \Bigg[\bigg|C \times \Big(\sqrt{O - M} + 2\sqrt{O - \sqrt{T}} - 4O - 2M + s^{1/32} + \sqrt{w} - 2C \\
&\qquad + |t| + \left|O - \sqrt{O - M} + w^{1/8}\right|\Big)\bigg|\Bigg]^{1/2}
\end{aligned} \tag{B.5}$$

APPENDIX C

MATHEMATICAL REPRESENTATION OF THE PROPOSED NEW ARTIFICIAL NEURAL NETWORKS MODEL FOR STRENGTH PREDICTION

The new ANN model can replicated by following the formulas and matrices below:
Let:

- $X_1, X_2, \ldots, X_6$: Normalised input values using normalisation Equation 3.7
- $f(\cdot)$: Activation function (ReLU) applied at each node, excluding input nodes
- $W_{jk}^{(1)}$: Weight connecting input node k to node j in the first hidden layer
- $W_{lj}^{(2)}$: Weight connecting node j in the first hidden layer to node l in the second hidden layer
- $W_l^{(3)}$: Weight connecting node l in the second hidden layer to the output node
- $b_j^{(1)}$: Bias term for node j in the first hidden layer
- $b_l^{(2)}$: Bias term for node l in the second hidden layer
- $b^{(3)}$: Bias term for the output node
- $h_j^{(1)}$: Calculated output of node j in the first hidden layer ($j = 1, 2, \ldots, 8$)
- $h_l^{(2)}$: Calculated output of node l in the second hidden layer ($l = 1, 2, \ldots, 6$)
- Y: Calculated model output, i.e., the predicted swelling%

Calculation of $h_j^{(1)}$, $h_l^{(2)}$, and Y combines the aggregation ($\sum$) and activation ($f(\cdot)$) steps at each node as demonstrated by the following equations:

For the first hidden layer nodes:

$$h_j^{(1)} = f\left(\sum_{k=1}^{6} X_k W_{jk}^{(1)} + b_j^{(1)}\right), \quad j = 1, \ldots, 9 \tag{C.1}$$

For the second hidden layer nodes:

$$h_l^{(2)} = f\left(\sum_{j=1}^{8} h_j^{(1)} W_{lj}^{(2)} + b_l^{(2)}\right), \quad l = 1, \ldots, 8 \tag{C.2}$$

For the output node:

$$Y = f\left(\sum_{l=1}^{6} h_l^{(2)} W_l^{(3)} + b^{(3)}\right) \tag{C.3}$$

Weights matrix for connections between the input and the first hidden layer:

$$W^{(1)} \in \mathbb{R}^{9\times 9}$$

$$W^{(1)} = \begin{bmatrix} -0.1471 & 0.8978 & 1.1414 & 0.2872 & 0.8992 & -0.4076 & 0.9836 & 0.3581 & 0.9330 \\ 0.3378 & -3.7306 & 0.2031 & -0.2441 & 0.2722 & 0.0774 & 0.5887 & -0.1786 & 0.9305 \\ -1.4346 & -1.9829 & -0.5792 & -0.2834 & 0.0387 & 2.0355 & -0.1284 & 0.2698 & -0.5044 \\ 0.8018 & 0.4982 & 0.0578 & 1.1418 & 0.1838 & -0.3637 & 0.2271 & -0.7111 & 0.6713 \\ 0.6100 & 0.2529 & 0.6936 & 1.0291 & 0.7557 & -0.2941 & 1.0532 & 0.0727 & 0.9566 \\ -0.3098 & 0.8457 & 0.3723 & -0.4598 & 0.1328 & -0.2578 & 0.2521 & -0.4827 & 0.5169 \\ -0.9456 & -0.8949 & -1.1428 & -1.3950 & -1.0589 & 1.1047 & -0.9038 & -0.0160 & -0.1252 \\ 0.3523 & -0.8480 & -0.2057 & -0.3170 & -0.6628 & 0.5744 & 0.2097 & -0.1456 & 0.2887 \\ 0.7220 & 0.1859 & 1.3253 & 0.9719 & 0.7941 & -1.7198 & 1.7127 & -0.1626 & 1.1884 \end{bmatrix} \tag{C.4}$$

Bias vector for nodes in the first hidden layer:

$$b^{(1)} \in \mathbb{R}^{9} = \begin{bmatrix} -0.0337 \\ 1.1056 \\ -0.3685 \\ -0.2839 \\ -0.1229 \\ 0.7848 \\ 0.0037 \\ -0.3102 \\ -0.0242 \end{bmatrix} \tag{C.5}$$

Weights matrix for connections between the first and the second hidden layers:

$$W^{(2)} \in \mathbb{R}^{9\times 8}$$

$$W^{(2)} = \begin{bmatrix} -0.2106 & -0.3606 & -0.5586 & 0.9109 & 0.1524 & -0.1885 & 0.1353 & -0.0530 \\ -0.3365 & -0.2115 & -2.3491 & -2.5976 & -0.5088 & -0.5562 & -4.0172 & 0.4529 \\ -0.5121 & 0.2693 & 0.3771 & 1.1879 & -0.3719 & 0.2846 & -0.4014 & -0.4378 \\ -0.2298 & 0.0173 & 0.5433 & 1.1706 & -0.4904 & 0.1067 & 0.7207 & 0.3487 \\ 0.2842 & -0.5941 & 0.6450 & 0.5110 & 0.2297 & -0.7282 & -0.0494 & -0.2353 \\ -0.2939 & -0.1874 & -0.9432 & -2.6044 & -0.4781 & -0.3004 & -0.6611 & -0.5805 \\ -0.5438 & -0.4169 & 1.2493 & 0.8780 & 0.1712 & -0.0195 & 0.5518 & -0.4111 \\ 0.1565 & -0.2692 & 0.2749 & 0.0529 & -0.3729 & -0.2818 & -0.2839 & -0.3530 \\ -0.5908 & 0.0261 & 0.8703 & 0.7551 & -0.2258 & 0.2144 & 0.7182 & -0.2943 \end{bmatrix} \tag{C.6}$$

Bias vector for nodes in the second hidden layer:

$$b^{(2)} \in \mathbb{R}^{8} = \begin{bmatrix} 0.0000 \\ 0.0000 \\ 0.1638 \\ 0.0924 \\ -0.0173 \\ -0.1880 \\ 0.7144 \\ 0.0000 \end{bmatrix} \tag{C.7}$$

Weights matrix for the connections between the second hidden layer and the output node:

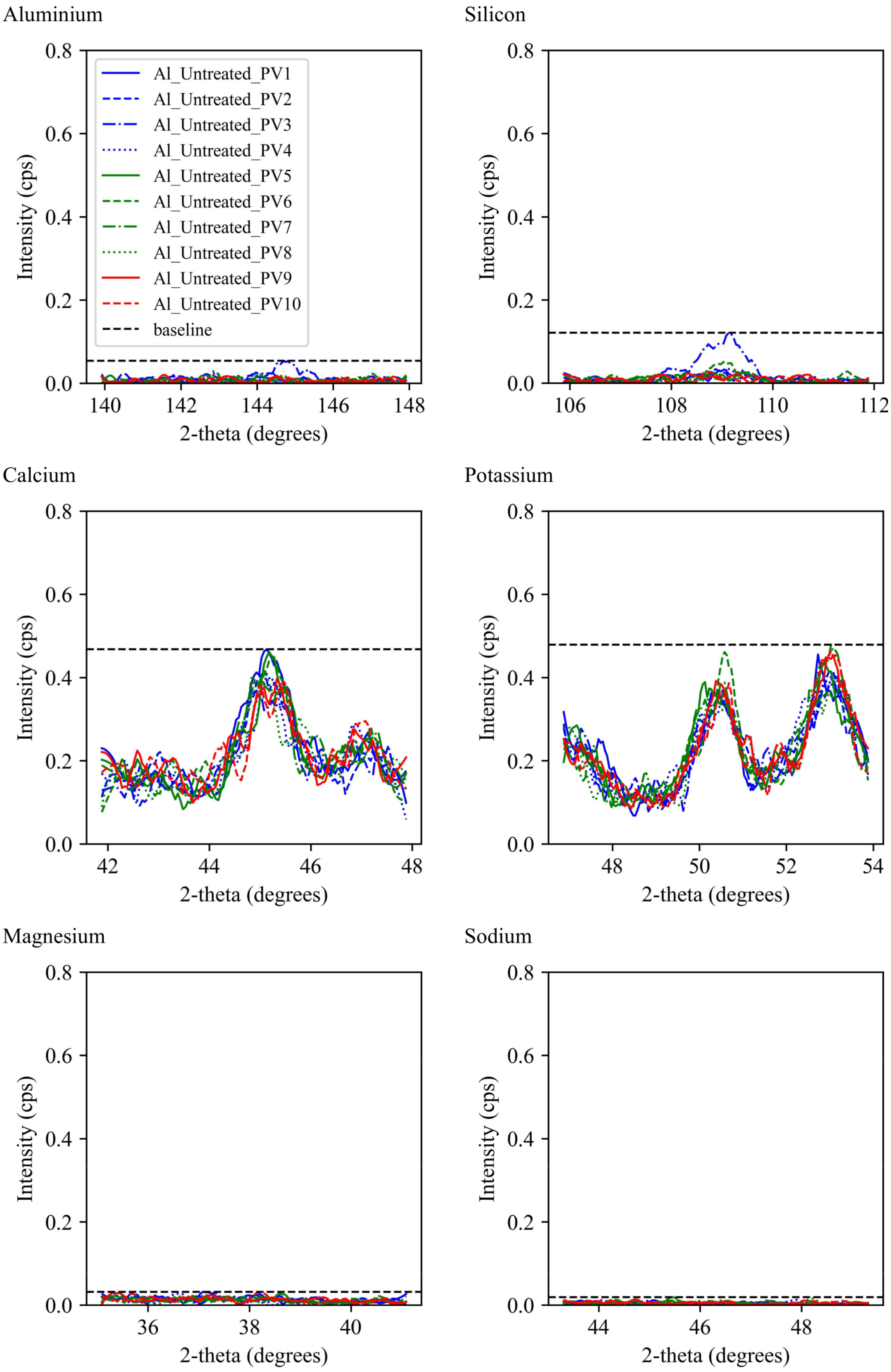
$$W^{(3)} \in \mathbb{R}^{8\times 1} = \begin{bmatrix} 0.2600 \\ -0.6828 \\ 1.6386 \\ 2.0511 \\ 0.2182 \\ -0.1055 \\ 1.6711 \\ -0.7936 \end{bmatrix} \tag{C.8}$$

Bias of the output node:

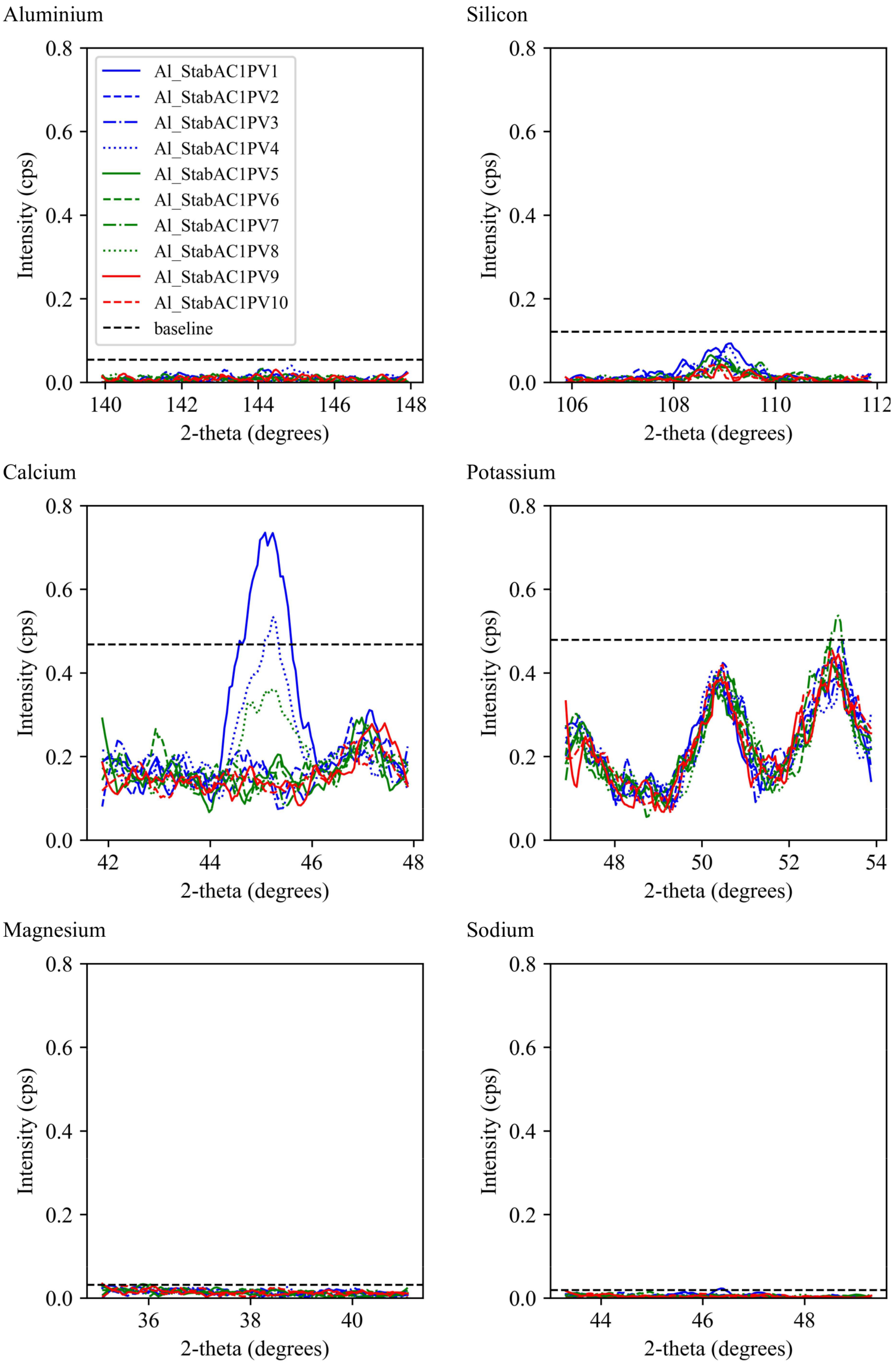
$$b^{(3)} \in \mathbb{R}^{1} = 0.3346 \tag{C.9}$$

APPENDIX D

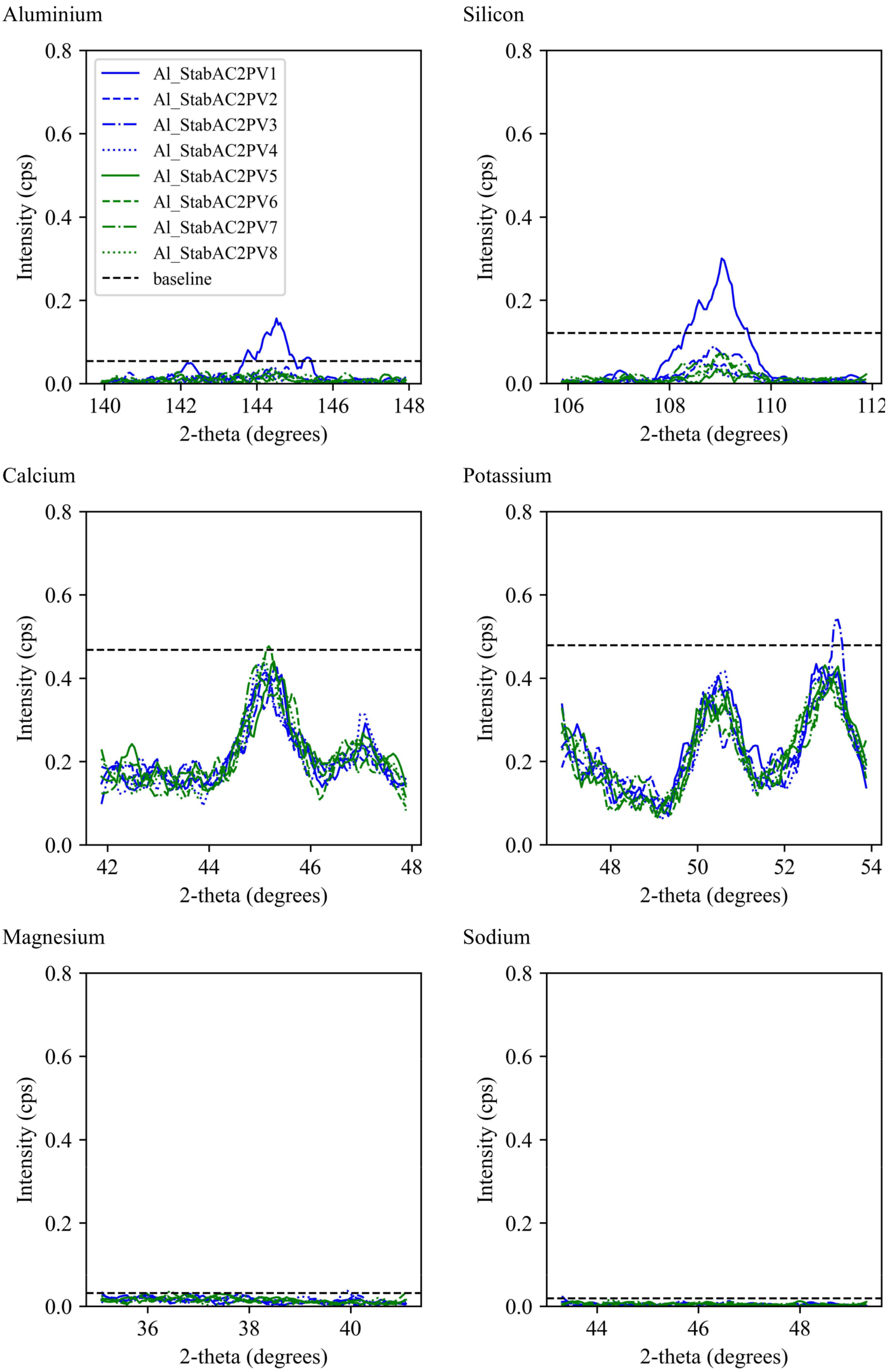
XRF RESULTS OF THE EFFLUENT SAMPLES



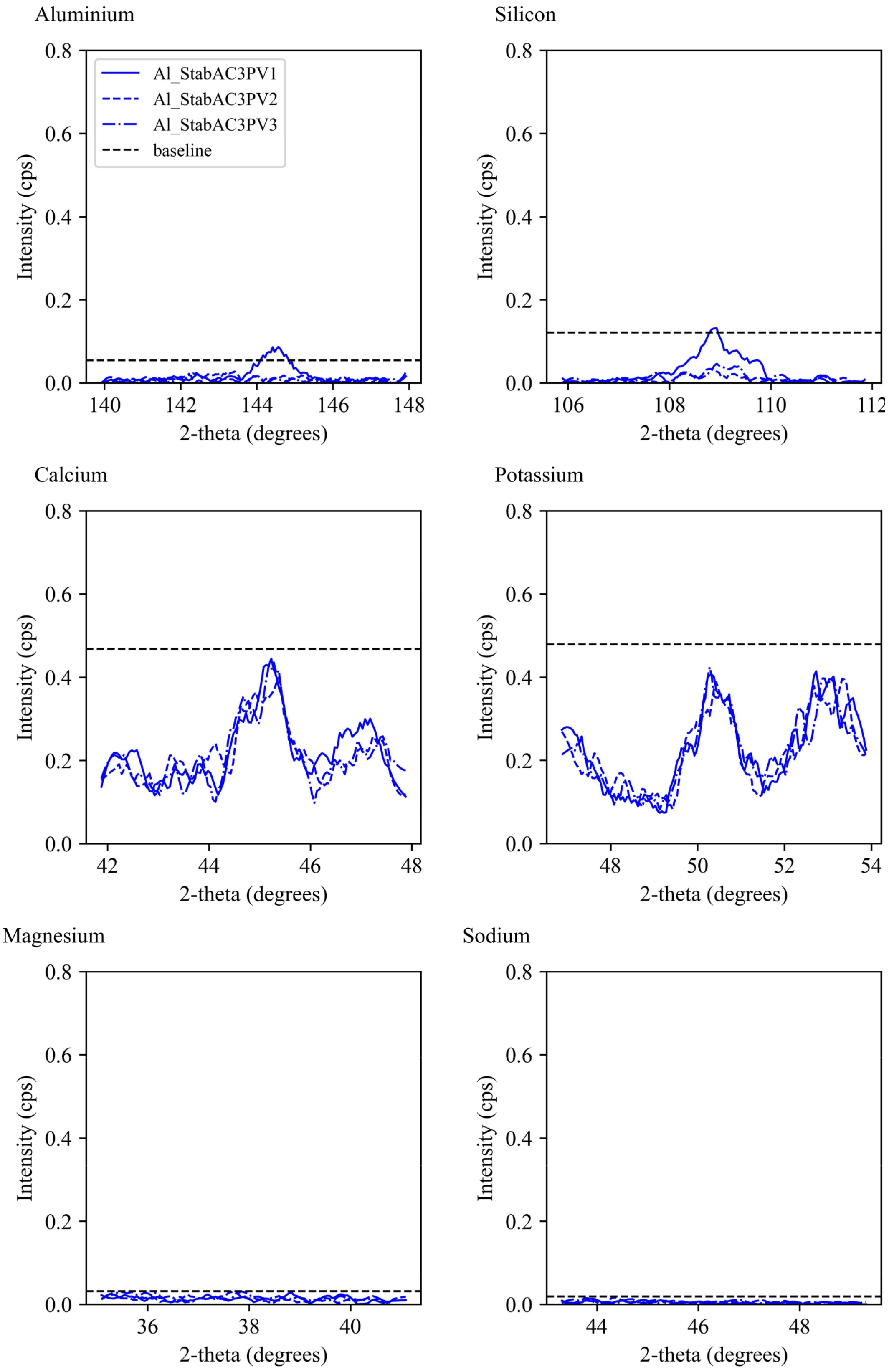
XRF (baseline) results of the effluent samples collected from Untreated_A, which underwent a single wetting cycle



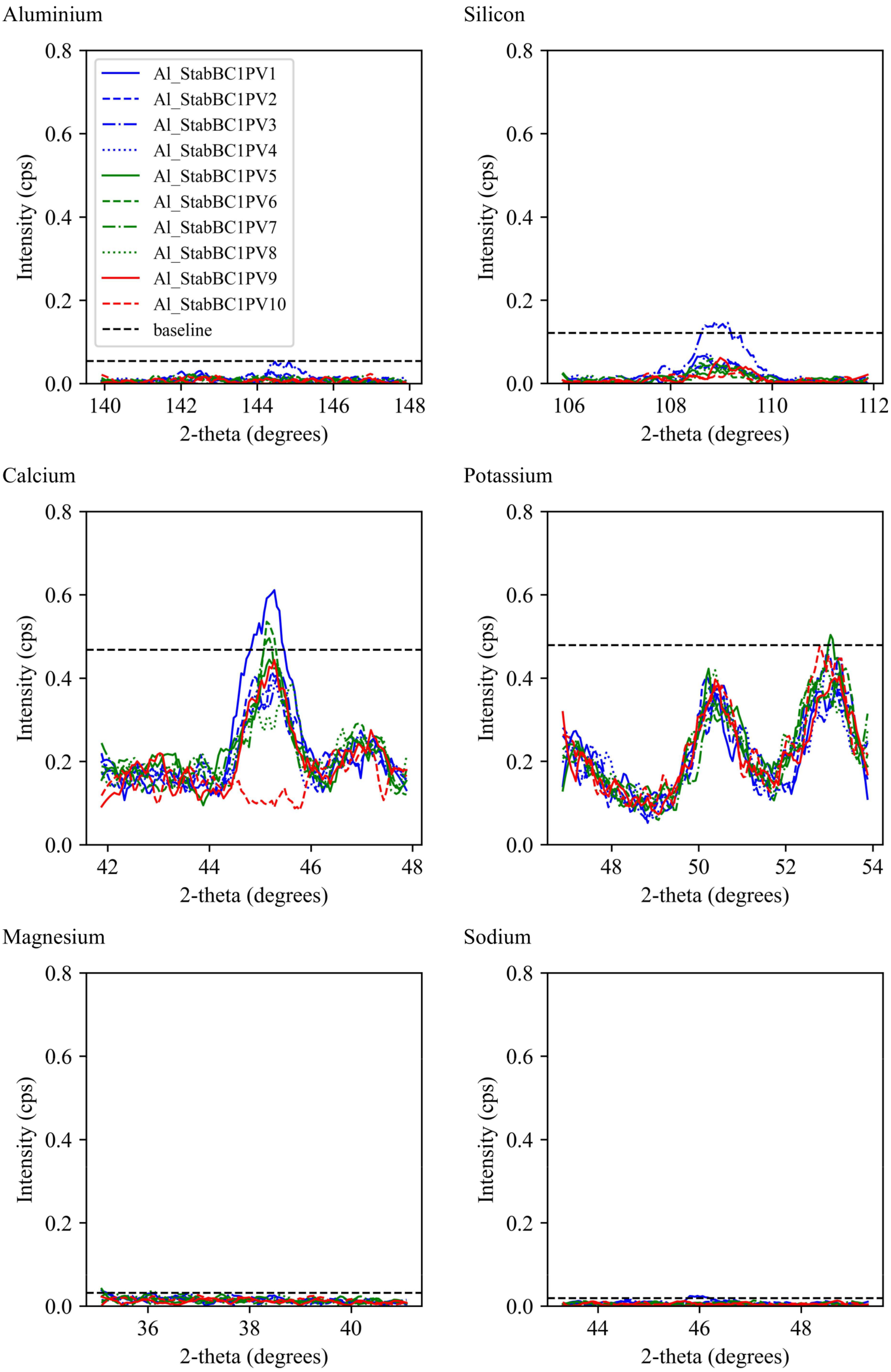
XRF results of the effluent samples collected from the stabilised sample Stabilised_A during the first wetting cycle



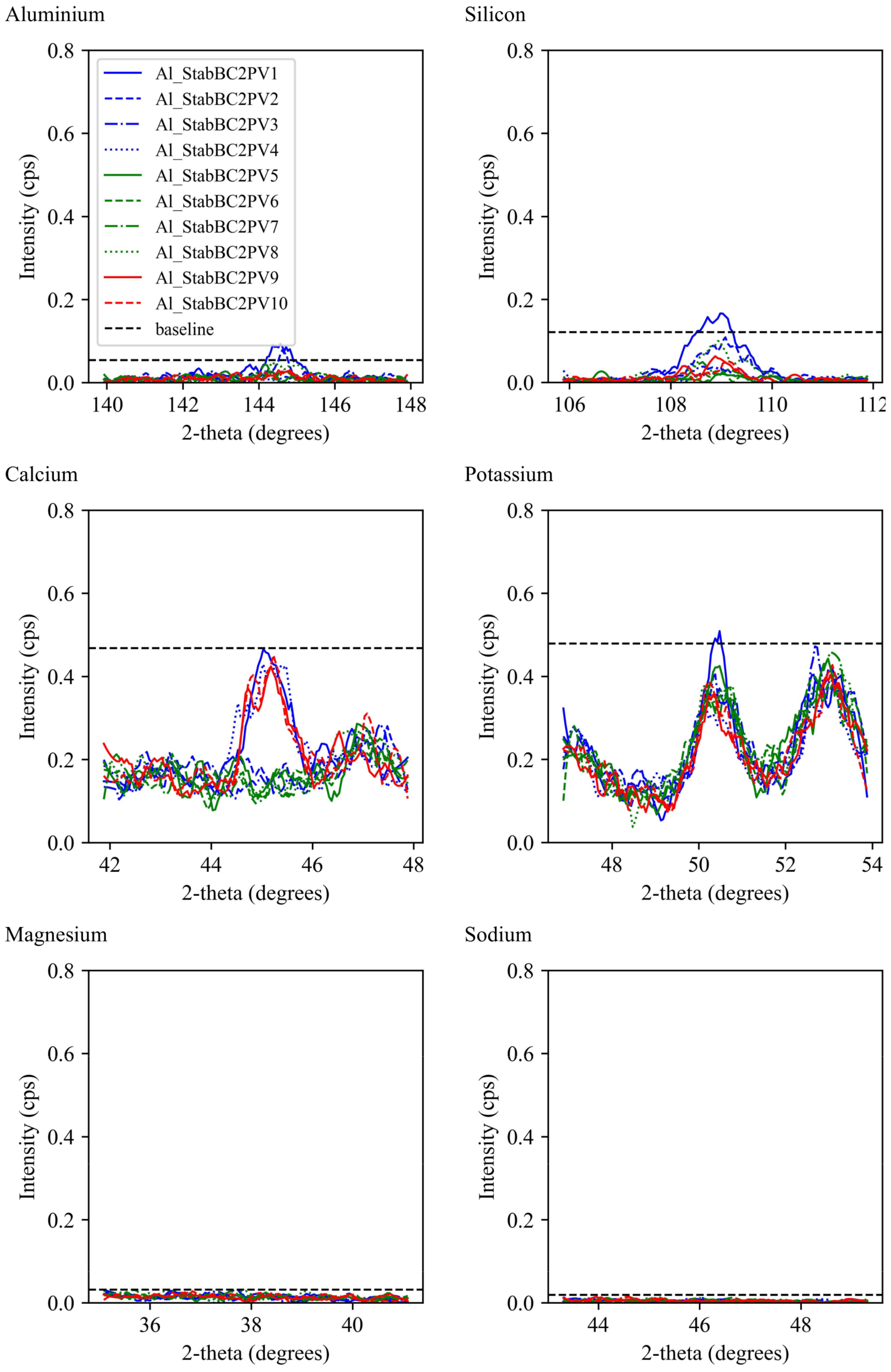
XRF results of the effluent samples collected from the stabilised sample Stabilised_A during the second wetting cycle



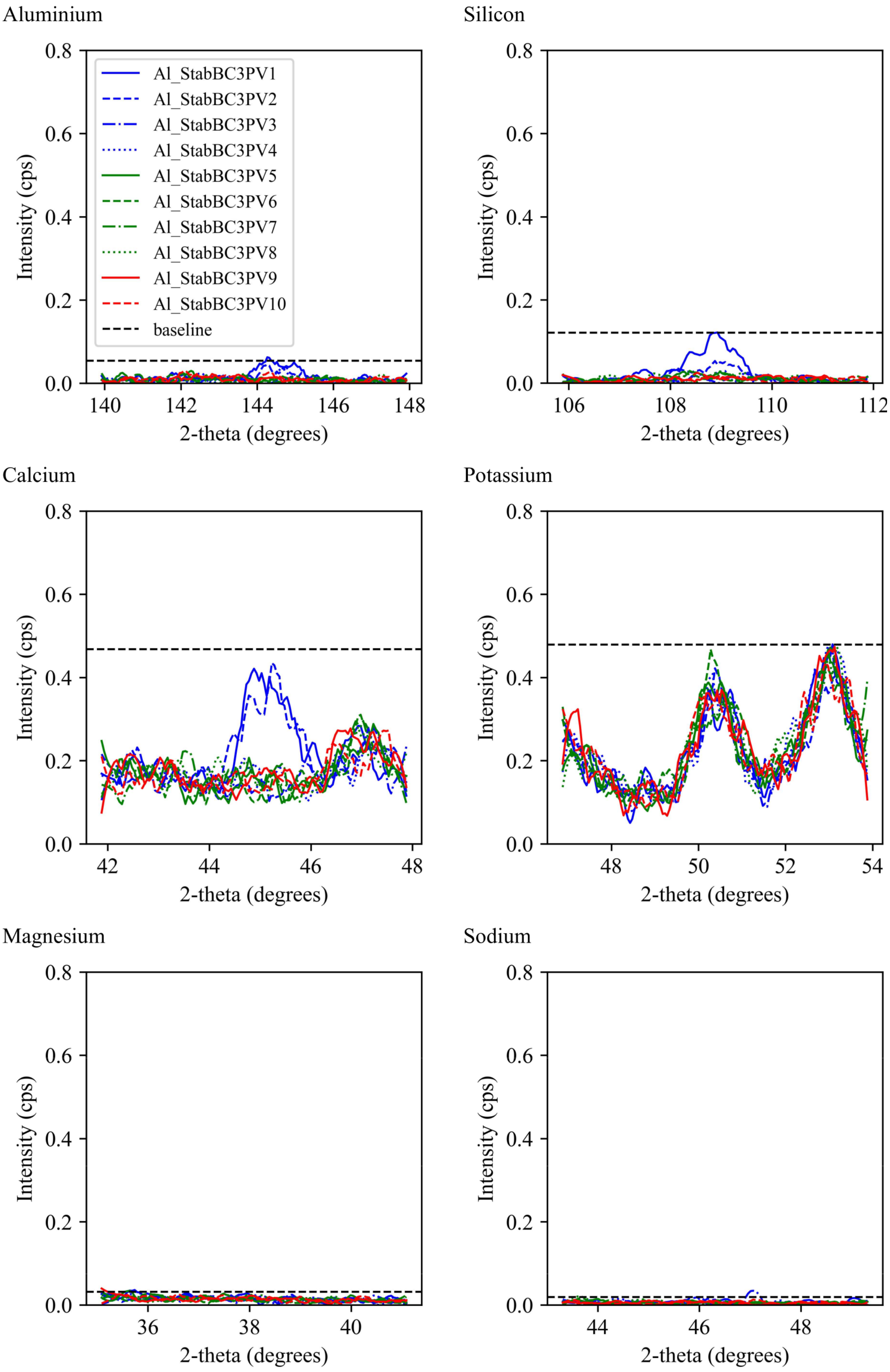
XRF results of the effluent samples collected from the stabilised sample Stabilised_A during the third wetting cycle



XRF results of the effluent samples collected from the stabilised sample Stabilised_B during the first wetting cycle



XRF results of the effluent samples collected from the stabilised sample Stabilised_B during the second wetting cycle


Aluminium

Silicon

Calcium

Potassium

Magnesium

Sodium

XRF results of the effluent samples collected from the stabilised sample Stabilised_B during the third wetting cycle

APPENDIX E

MATHEMATICAL REPRESENTATION OF THE PROPOSED NEW ARTIFICIAL NEURAL NETWORKS MODEL FOR SWELLING PREDICTION

The new ANN model can replicated by following the formulas and matrices below:
Let:

- $X_1, X_2, \ldots, X_6$: Normalised input values using normalisation Equation 3.7
- $f(\cdot)$: Activation function (ReLU) applied at each node, excluding input nodes
- $W_{jk}^{(1)}$: Weight connecting input node k to node j in the first hidden layer
- $W_{lj}^{(2)}$: Weight connecting node j in the first hidden layer to node l in the second hidden layer
- $W_{l}^{(3)}$: Weight connecting node l in the second hidden layer to the output node
- $b_j^{(1)}$: Bias term for node j in the first hidden layer
- $b_l^{(2)}$: Bias term for node l in the second hidden layer
- $b^{(3)}$: Bias term for the output node
- $h_j^{(1)}$: Calculated output of node j in the first hidden layer ($j = 1, 2, \ldots, 8$)
- $h_l^{(2)}$: Calculated output of node l in the second hidden layer ($l = 1, 2, \ldots, 6$)
- Y: Calculated model output, i.e., the predicted swelling%

Calculation of $h_j^{(1)}$, $h_l^{(2)}$, and Y combines the aggregation ($\sum$) and activation ($f(\cdot)$) steps at each node as demonstrated by the following equations:

For the first hidden layer nodes:

$$h_j^{(1)} = f\left(\sum_{k=1}^{6} X_k W_{jk}^{(1)} + b_j^{(1)}\right), \quad j = 1, \ldots, 8 \tag{E.1}$$

For the second hidden layer nodes:

$$h_l^{(2)} = f\left(\sum_{j=1}^{8} h_j^{(1)} W_{lj}^{(2)} + b_l^{(2)}\right), \quad l = 1, \ldots, 6 \tag{E.2}$$

For the output node:

$$Y = f\left(\sum_{l=1}^{6} h_l^{(2)} W_l^{(3)} + b^{(3)}\right) \tag{E.3}$$

Weights matrix for connections between the input and the first hidden layer:

$$W^{(1)} \in \mathbb{R}^{6\times 8}$$

$$W^{(1)} = \begin{bmatrix} -0.1457 & -0.2500 & 0.3231 & -0.2350 & 0.2060 & -0.4162 & -0.4368 & 0.1896 \\ 0.1545 & -0.2897 & 0.4375 & 0.3501 & 0.4897 & -0.2535 & -2.4975 & 0.2209 \\ -0.8693 & 0.5267 & -1.2173 & -1.0302 & 2.3653 & 0.0898 & -0.5554 & 0.4797 \\ -1.4478 & 0.6101 & -2.0317 & -1.8037 & 2.8124 & 0.4191 & 0.8189 & -6.2004 \\ -2.6774 & -0.0811 & -4.4390 & -2.0043 & 3.8169 & 0.5114 & 0.5577 & -0.4656 \\ -0.4087 & -0.5753 & -0.2022 & -0.5883 & 0.3251 & -0.2204 & 0.9612 & 1.1110 \end{bmatrix} \tag{E.4}$$

Bias vector for nodes in the first hidden layer:

$$b^{(1)} \in \mathbb{R}^{8} = \begin{bmatrix} 0.8238 \\ -0.2083 \\ 0.7534 \\ 1.0540 \\ -0.3963 \\ -0.3337 \\ -0.3378 \\ -0.3442 \end{bmatrix} \tag{E.5}$$

Weights matrix for connections between the first and the second hidden layers:

$$W^{(2)} \in \mathbb{R}^{8\times 6}$$

$$W^{(2)} = \begin{bmatrix} 0.5341 & -0.3987 & -0.2880 & 1.0827 & 0.7616 & -0.5880 \\ -0.1643 & -0.0584 & -0.3709 & -0.2019 & 0.5798 & -0.5798 \\ 0.7776 & -0.5095 & -0.4547 & 1.4736 & 0.4146 & 0.2044 \\ 0.4218 & 0.3660 & -0.2804 & 1.5884 & 0.4923 & -0.0103 \\ -1.9700 & -0.1745 & -0.2773 & -0.9845 & -2.1223 & 0.5050 \\ 0.0009 & 0.2626 & 0.3132 & -0.5487 & 0.1154 & 0.3254 \\ 0.8777 & -0.6881 & -0.4180 & -0.7233 & 0.6572 & -1.2671 \\ 2.5432 & -0.5720 & -0.0747 & 1.1008 & 2.4518 & -0.5953 \end{bmatrix} \tag{E.6}$$

Bias vector for nodes in the second hidden layer:

$$b^{(2)} \in \mathbb{R}^{6} = \begin{bmatrix} 0.6281 \\ -0.1959 \\ 0.0000 \\ 0.9269 \\ 0.7793 \\ -0.6906 \end{bmatrix} \tag{E.7}$$

Weights matrix for the connections between the second hidden layer and the output node:

$$W^{(3)} \in \mathbb{R}^{6\times 1} = \begin{bmatrix} 2.2582 \\ -0.7454 \\ 0.7738 \\ 2.0642 \\ 2.1332 \\ -0.7819 \end{bmatrix} \tag{E.8}$$

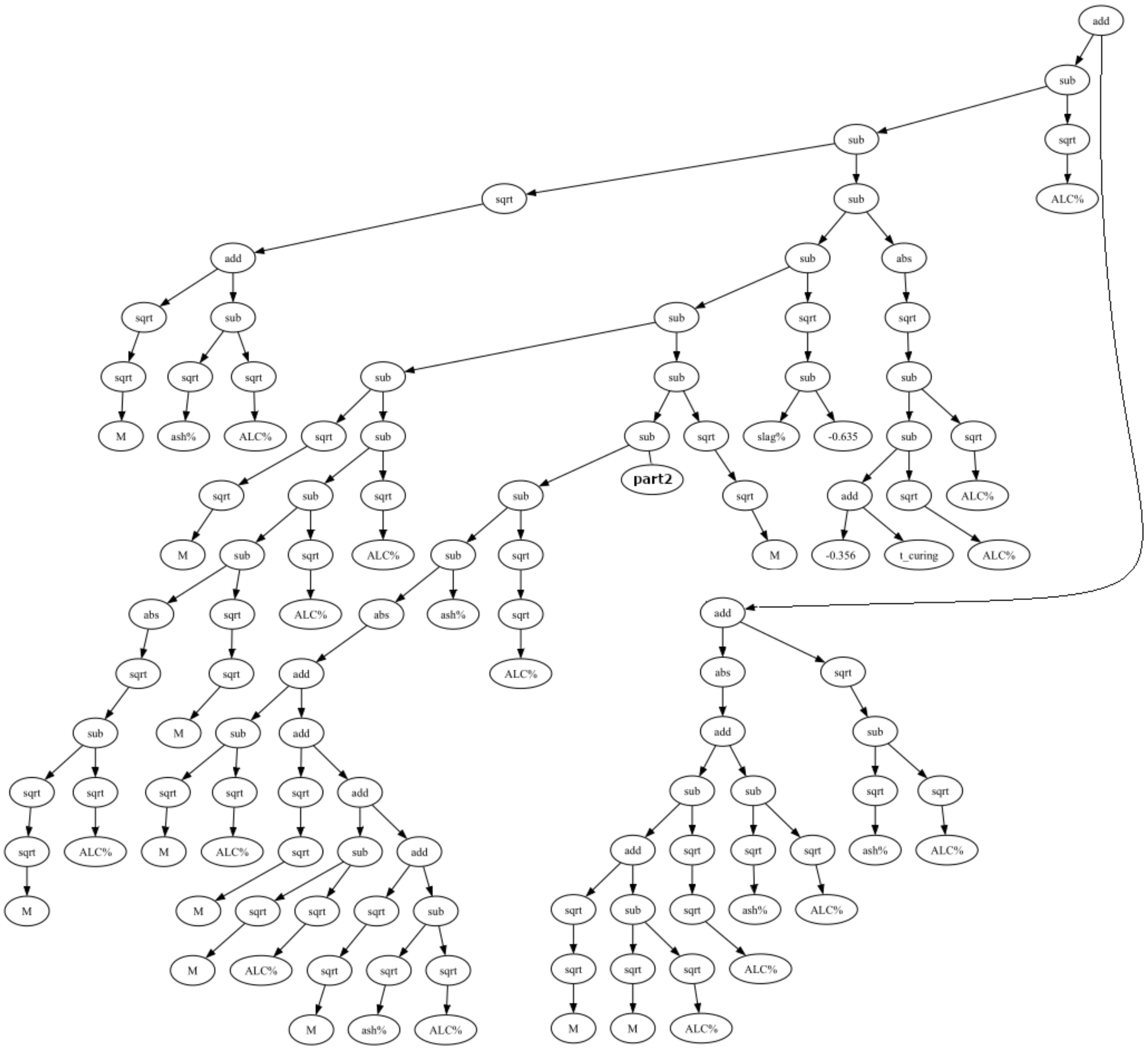
Bias of the output node:

$$b^{(3)} \in \mathbb{R}^1 = 0.8488 \tag{E.9}$$

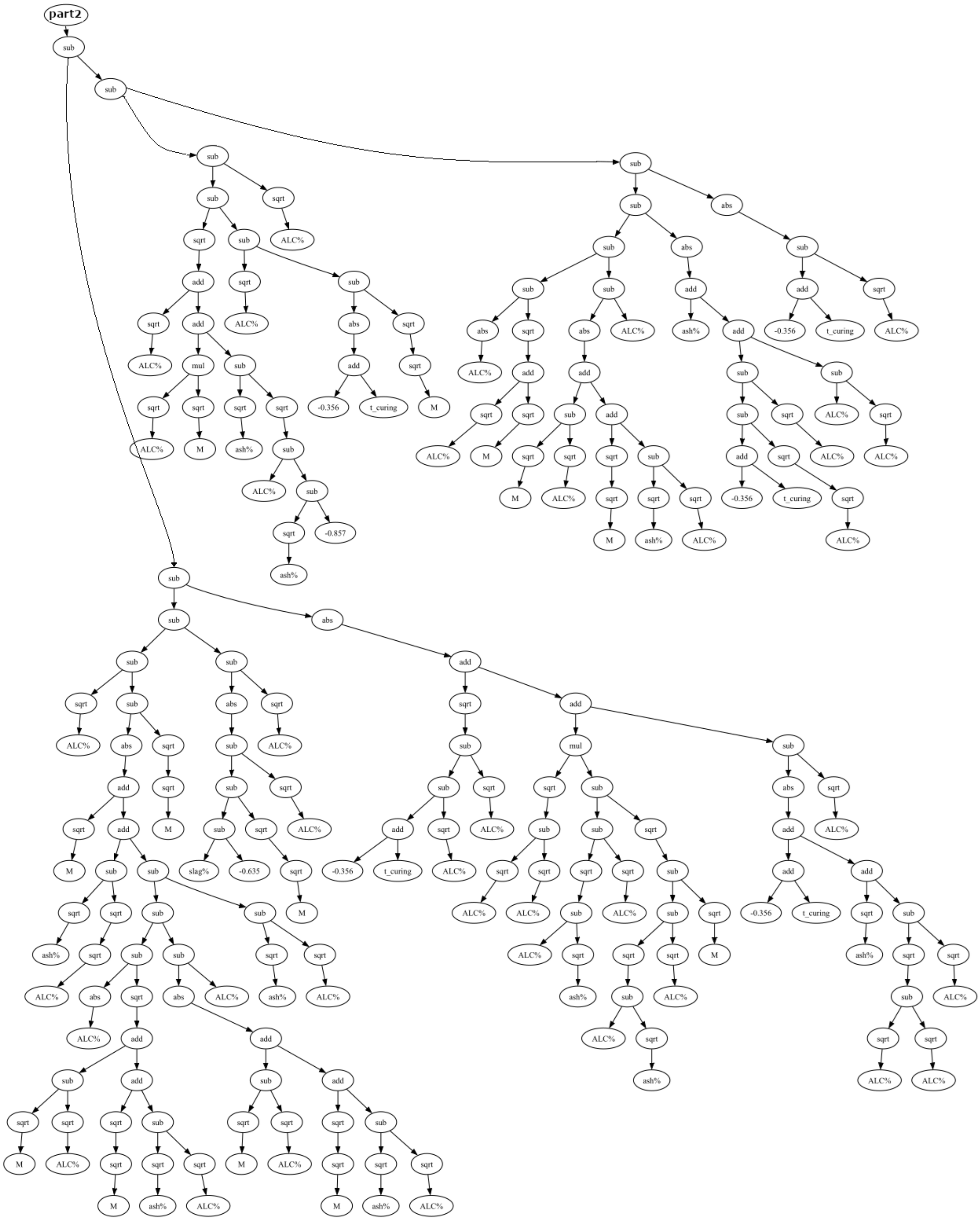
APPENDIX F

EXPRESSION TREE AND MATHEMATICAL REPRESENTATION OF THE PROPOSED NEW GENETIC PROGRAMMING MODEL FOR SWELLING PREDICTION

Expression tree representations of the Genetic Programming models for strength prediction:



Sub-tree 1


part2
sub
sub
sub
sub
sub
sqrt
sub
abs
sqrt
sub
ALC%
sub
abs
sub
add
sqrt
sub
sub
sub
add
add
sqrt
sqrt
add
ALC%
abs
sqrt
abs
sqrt
abs
ALC%
ash%
add
-0.356
t_curing
ALC%
ALC%
mul
sub
add
sqrt
ALC%
add
add
sub
sub
sqrt
sqrt
sqrt
sqrt
-0.356
t_curing
M
sqrt
sqrt
sub
add
sub
sqrt
ALC%
sqrt
ALC%
M
ash%
sub
ALC%
M
sqrt
sqrt
sqrt
sub
add
sqrt
ALC%
ALC%
ALC%
sub
M
ALC%
sqrt
sqrt
sqrt
-0.356
t_curing
sqrt
sqrt
-0.857
M
ash%
ALC%
ALC%
ash%
sub
sub
abs
sub
sub
add
sqrt
sub
abs
sqrt
sqrt
add
ALC%
abs
sqrt
sub
ALC%
sub
mul
sub
add
sqrt
sub
sqrt
sub
sqrt
sqrt
sub
abs
sqrt
sqrt
add
M
sub
sqrt
ALC%
add
sqrt
ALC%
sub
sub
sqrt
add
ALC%
M
sub
sub
slag%
-0.635
sqrt
-0.356
t_curing
ALC%
sqrt
sqrt
sqrt
sqrt
sub
add
add
sqrt
sqrt
sub
sub
M
ALC%
ALC%
sub
ALC%
sub
sqrt
-0.356
t_curing
sqrt
sub
ash%
sqrt
sub
sub
sqrt
sqrt
ALC%
sqrt
sqrt
sqrt
M
ash%
sqrt
sqrt
ALC%
abs
sqrt
abs
ALC%
ash%
ALC%
ash%
sub
ALC%
sub
ALC%
ALC%
add
add
ALC%
sqrt
sqrt
sqrt
sub
add
sub
add
ash%
ALC%
ALC%
sqrt
sqrt
sqrt
sub
sqrt
sqrt
sqrt
sub
M
ALC%
sqrt
sqrt
sqrt
M
ALC%
sqrt
sqrt
sqrt
M
ash%
ALC%
M
ash%
ALC%

Sub-tree 2

Mathematical representation of the above expression trees:

$$
\begin{aligned}
\text{swelling\%} = {} & \left| M^{\frac{1}{4}} + \sqrt{M} - 2\sqrt{A} - A^{\frac{1}{4}} + \sqrt{a} \right| + \sqrt{\sqrt{a} - \sqrt{A}} + \sqrt{A} - \sqrt{M^{\frac{1}{4}} + \sqrt{a} - \sqrt{A}} \\
& - \left(-\left|\sqrt{M^{\frac{1}{4}} - \sqrt{A}}\right| - 2\sqrt{A} \right) - \left| 2\sqrt{M} - 3\sqrt{A} + 2M^{\frac{1}{4}} + \sqrt{a} \right| - a - A^{\frac{1}{4}} + \sqrt{A} \\
& - \left| 2\sqrt{a} - A^{\frac{1}{4}} + |A| - 2\sqrt{A} + \sqrt{M^{\frac{1}{4}} - \sqrt{a} - \sqrt{A}} - \left|\sqrt{M} + M^{\frac{1}{4}} + \sqrt{a}\right| - A \right| \\
& - M^{\frac{1}{4}} - \left| s - 0.635 - M^{\frac{1}{4}} - \sqrt{A} \right| - \sqrt{A} \\
& - \left| \sqrt{-0.356 + t - 2\sqrt{A}} + \left|-0.356 + t + \sqrt{a} - \sqrt{A}\right| - \sqrt{A} \right| \qquad \text{(F.1)} \\
& - \sqrt{\sqrt{A}\,\sqrt{M} + \sqrt{a} - \sqrt{A - \sqrt{a} + 0.857} + \sqrt{A}} - \sqrt{A} - \left|-0.356 + t\right| - M^{\frac{1}{4}} - \sqrt{A} - |A| \\
& - \sqrt{\sqrt{A} + \sqrt{M}} - \left| \sqrt{M} - 2\sqrt{A} + M^{\frac{1}{4}} + \sqrt{a} \right| - A \\
& - \left|a - 0.356 + t - A^{\frac{1}{4}} - 2\sqrt{A} + A\right| - \left| -0.356 + t - \sqrt{A} \right| \\
& - \left| \sqrt{-0.356 + t - 2\sqrt{A}} \right| - \sqrt{s + 0.635} - M^{\frac{1}{4}}
\end{aligned}
$$