The design optimisation of cementitious mixtures using artificial intelligence algorithms for sustainable mining and construction projects

Ehsan Sadrossadat

BSc (Civil Eng.) MSc (Geotech. Eng.)

This thesis is presented for the degree of

Doctor of Philosophy

The University of Western Australia

Department of Civil, Environmental and Mining Engineering

2022

THESIS DECLARATION

I, Ehsan Sadrossadat, certify that:

This thesis has been substantially accomplished during enrolment in this degree.

This thesis does not contain material which has been submitted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

In the future, no part of this thesis will be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text and, where relevant, in the Authorship Declaration that follows.

This thesis does not violate or infringe any copyright, trademark, patent, or other rights whatsoever of any person.

This thesis contains published work and/or work prepared for publication, some of which has been co-authored.

Signature: 

Date: 01/06/2022

ABSTRACT

Cementitious mixtures (CMs) are widely-used materials for different applications in mining and construction projects. Common CMs are normal concrete (NC), high-performance concrete (HPC), self-compacting concrete (SCC), cemented paste backfill (CPB) and hydraulic backfill (CHB) and so forth. CMs consist of precise proportions of materials (i.e. cement, water, sand and/or aggregates) with specific properties which are mixed up based on well-defined methods and conditions. Such composite materials are exceptional due to their valuable and designable properties in fresh or hardened states such as strength, workability and durability.

Emphasis on sustainability issues with respect to excessive use of natural resources and environmental concerns has risen to importance in recent projects. In addition to functional properties, CMs must also satisfy many other objectives and constraints in current projects such as production cost and carbon footprint to strive for a higher level of sustainability. Cement is the main component of CMs and CO_2 emitted from its production only accounts for at least 5-7% of the total greenhouse gas emissions leading to global warming and climate change. Furthermore, constituents of CMs are basically acquired from natural resources and prepared for use after costly and energy-consuming processes, but they might not be recyclable after use. These issues are not economically and environmentally lucrative and lack sustainable development goals.

Recent research recommends that some waste materials and by-products can be used as the main material or partial replacement of the costly elements such as cement in order to produce CMs. Tailings are the major mining waste materials left over after extracting valuable minerals from the mined ore. Cemented paste backfill (CPB) is a CM designed to send tailings back into the mine in order to fill the mined-out stopes and provide local support and stability allowing for a safer and more efficient mining operation. Some waste materials and by-products such as fly ash (FA) and ground-granulated blast-furnace slag (GGBS or GGBFS) and silica fume (SF) are proposed to use in CMs such as self-compacting concrete (SCC) or ultra-high-performance concrete (UHPC) due to their self-cementing or hydraulic, and pozzolanic properties. Admixtures such as plasticisers or fibres also support the sustainability of CMs by improving their properties and extending their applications. These strategies may lead to improvement of the properties of CMs, reduction of the consumption of natural resources and energy, a cleaner production, a profitable waste management and accordingly a more sustainable industry.

The design of CMs is complex and challenging due to the large number of influential factors, objectives and constraints, particularly when it comes to sustainability. Traditionally, CMs are designed based on trial-and-error experiments with the help of experience and recommendations of standards and specifications. However, experimental mixture design is often time-consuming, rigorous and requires expensive equipment and expertise. Yet, experimental methods may lead to suboptimal solutions that are satisfactory only within the intervals of testing.

Many experiments have already been performed to investigate the performance and properties of CMs allowing for collection of experimental datasets and observations. Such datasets are valuable to analyse and make profitable data-driven decisions. It is noteworthy that specifications and standards are also established based on experimental studies and recorded data. Another smart strategy is to use such data to create mathematical methods for predictive modelling and optimisation aims. In this regard, researchers have been attempting to develop and propose appropriate mathematical methods to augment the process and reduce the costs and efforts of mixture design of CMs. Various exciting approaches proposed include statistical mixture design (SMD), design of experiment (DOE), response surface methodology (RSM), artificial intelligence (AI) algorithms, machine learning (ML), and optimisation algorithms. However, reviewing the literature reveals that such proposals have several gaps and even fundamental issues that need to be addressed or improved.

There is an increasing demand for designers to consider multiple requirements in design process and make the best possible decisions at a faster pace while decreasing costs. In fact, the mixture design of CMs can be converted to a mathematical design optimisation problem. For this aim, it is crucial to precisely describe the project, determine influential factors, objectives and constraints and apply robust mathematical methods. This research aims to engineer the concept and propose improved frameworks for multi-objective mixture design optimisation (MOMDO) of CMs for sustainable mining and construction projects using AI techniques.

An appropriate project statement and description is the most important step of the MOMDO procedure. Design variables can be any combinations of influential factors. Objectives are converted to models to be maximised or minimised, and constraints can be any relationship of design variables, models, or inequalities or equalities that must be met. Given this context, properties of CMs such as strength and workability, and production cost and CO_2 emission or any other criteria can be considered as objectives and or constraints. However, proposed methods commonly consider the proportions of constituents as variables only and ignore

unique features of materials such as physical and chemical properties, or other influential factors such as experimental conditions and programs.

In the present thesis, comprehensive and reliable datasets have been collected which contain international experimental results conducted on different CMs such as CPB, SCC and UHPC. The performance of various ML algorithms such as Genetic Programming (GP), Support Vector Regression (SVR), Artificial Neural Network (ANN) and Gaussian Process Regression (GPR) has been investigated and compared for development of high-accuracy models. The accuracy of models used as objectives or constraints is of paramount significance. Detailed investigations are presented for data pre-processing, dimensionality reduction, feature transformation, variable selection, development and accuracy assessment of models. After converting the descriptive statement of pre-defined projects to mathematical optimisation problems, robust optimisation and metaheuristic algorithms such as particle swarm optimisation (PSO) and non-dominated sorting genetic algorithm called NSGA-II are used to find the best mixture designs meeting the considered requirements. In this thesis, various single or multiple objective design scenarios subject to different constraints are considered to represent the feasibility and applicability of a proper MOMDO framework using AI algorithms. The optimisation results are also validated which confirms the robustness of proposed approach. More importantly, GPR was found to be robust for predictive modeling of the complex properties of CMs, and NSGA-II as a robust optimisation algorithm to convert design scenarios to MOMDO of CMs.

This thesis presents a number of advances and novel ideas for MOMDO of different CMs with further discussions, validation and verification and improvement. The underpinning workflow introduced in this thesis can be extended for process design, multi-objective design and advanced manufacturing, and discovery of new CMs or any composite material such as engineered cementitious composites (ECC) and 3D printed concrete. The proposed MOMDO approach considers multiple objectives in design to identify optimum amounts of material or influential factors without the need to overtest the used materials. Finally, this approach certainly reduces the cost and effort of experimental designs, enables optimum consumption of resources, which leads to a cleaner production and more sustainable mining and construction industries. Future studies are recommended to improve the performance of models through more comprehensive international datasets and the potential use of the proposed framework for MOMDO of materials.

ACKNOWLEDGEMENTS

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

I also acknowledge the facilities, and the scientific and technical assistance of Microscopy Australia at the Centre for Microscopy, Characterisation & Analysis, The University of Western Australia, a facility funded by the University, State and Commonwealth Governments.

I cannot be more grateful to have Hakan Basarir, Ali Karrech, and Mohamed Elchalakani as my supervisors. My sincerest gratitude to them for all their kindness, valuable support, technical assistance, best advice and inspirations to this research and more importantly shaping my way in my life which I will cherish forever. I feel extremely privileged to have had the opportunity of working with and learning from them. Thank you very much for everything!

I would like to appreciate and acknowledge The University of Western Australia and the professional and supportive Professors and staff at Department of Civil, Environmental and Mining Engineering and Graduate Research School.

To all my friends who have helped me and made great memories during my research.

My deepest gratitude goes to my father, mother, my lovely wife, Saeedeh Osouli, and my brother for all their always unconditional love, motivations, protection, patience and support. All my accomplishments even little have not been achieved without your love and support.

Ehsan Sadrossadat

AUTHORSHIP DECLARATION: CO-AUTHORED PUBLICATIONS

This thesis contains work that has been published and/or prepared for publication.

Details of the work: "An evolutionary-based prediction model of the 28-day compressive strength of high-performance concrete containing cementitious materials" Published in Advances in Civil Engineering Materials https://doi.org/10.1520/ACEM20190016
Student contribution to work: 80%
Location in thesis: [Chapter 2]
Co-author signatures and dates: Hakan Basarir [redacted] 17/04/2022

Details of the work: "The optimisation of cemented hydraulic backfill mixture design parameters for different strength conditions using artificial intelligence algorithms." Published in International Symposium on Mine Planning & Equipment Selection https://doi.org/10.1007/978-3-030-33954-8_28
Location in thesis: [Chapter 3]
Student contribution to work: [80%]
Co-author signatures and dates:

Hakan Basarir	Ali Karrech	Richard Durham	Andy Fourie	Han Bin
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
17/4/2022	16/2/22	26/04/22	18/4/22	18/4/20022

Details of the work:

"Multi-objective mixture design of cemented paste backfill using particle swarm optimisation algorithm"

Published in Minerals Engineering

https://doi.org/10.1016/j.mineng.2020.106385

Location in thesis:

[Chapter 4]

Student contribution to work:

[80%]

Co-author signatures and dates:

Hakan Basarir	Ganhua Luo	Ali Karrech	Richard Durham	Andy Fourie	Mohamed Elchalakani
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
17/04/2022	26/04/22	16/04/22	26/04/22	18/4/22	16/04/22

Details of the work:

"Multi-objective mixture design and optimisation of steel fiber reinforced UHPC using machine learning algorithms and metaheuristics"

Published in Engineering with Computers

https://doi.org/10.1007/s00366-021-01403-w

Location in thesis:

[Chapter 5]

Student contribution to work:

85%

Co-author signatures and dates:

Hakan Basarir	Ali Karrech	Mohamed Elchalakani
[redacted]	[redacted]	[redacted]
17/04/2022	16/04/22	16/04/22

Details of the work:

"Multi-objective design of composite materials through an engineered AI framework: Sustainable SCC"

Prepared for submission.

Location in thesis:

[Chapter 6]

Student contribution to work:

[80%]

Co-author signatures and dates:

Hakan Basarir	Ali Karrech	Mohamed Elchalakani
[redacted]	[redacted]	[redacted]
17/04/2022	16/04/22	16/04/22

Details of the work:

"An engineered ML model for prediction of the compressive strength of Eco-SCC based on type and proportions of materials"

Published in Cleaner Materials

https://doi.org/10.1016/j.clema.2022.100072

Location in thesis:

[Chapter 7]

Student contribution to work:

70%

Co-author signatures and dates:

Hakan Basarir	Ali Karrech	Mohamed Elchalakani
[redacted]	[redacted]	[redacted]
17/04/2022	16/04/22	16/04/22

Student signature: [redacted] Date: 01/06/2022
I, Ali Karrech certify that the student's statements regarding their contribution to each of the works listed above are correct. Coordinating supervisor signature: [redacted] Date: 01/06/2022

NOTES ON VARIATIONS TO PUBLICATION TEXTS

- Format and style of texts have been updated according to the thesis.
- Number of tables, figures and equations have been updated and reformatted.
- References in all papers have been compiled and moved to the end of thesis under heading "References".
- Minor corrections have been done to improve the coherence of thesis.
- Minor modifications may exist between chapters and their published form as papers due to the comments of peer reviewers at journals.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1-1 Introductory background

Cementitious mixtures (CMs) are the most widely-used composite materials in civil, mining and construction projects due to their exceptional rheological, mechanical and durability properties in fresh or hardened states. Common CMs are concrete, cement paste, high-performance concrete (HPC), self-compacting concrete (SCC), cemented paste backfill (CPB), cemented hydraulic backfill (CHB) and so forth. These mixtures are mainly composed of aggregates bound by a cementitious paste consisting of water and cement. Other materials and admixtures such as superplasticizers or fibres may also be used to improve specific properties. CMs are commonly produced by adjusting the proportion and type of constituents based on requirements, objectives and constraints which may change from one project to another. Fig. 1.1 illustrates some typical factors in the design of CMs design of CMs.

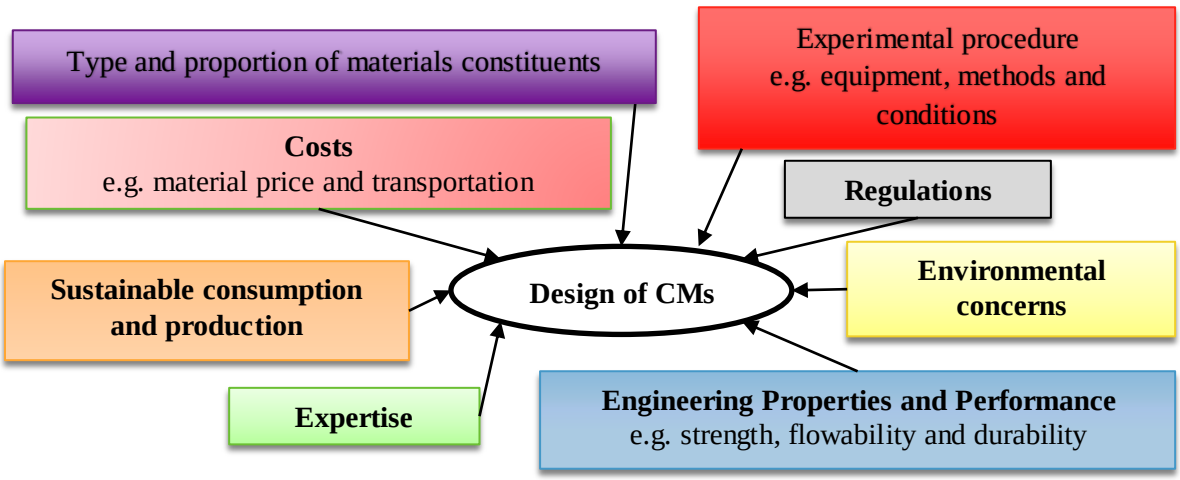


Fig. 1.1 Typical factors in the design of CMs in construction and mining projects.

Emphasis on sustainability issues with respect to inefficient use of natural resources and environmental concerns has risen to importance in recent projects. Constituents of CMs are basically obtained from natural resources and cannot be recycled after use. The production of cement as the main element of CMs requires a rigorous process resulting in considerable carbon dioxide (CO_2) emission, which is not environmentally-friendly. CO_2 emission coming from cement production only accounts for at least 5-7% of the total greenhouse gas emissions, approximately1.5 billion tons each year resulting in global warming and climate change [2, 3]. Each year, large amounts of CMs are produced and used for different applications which require improvements in order to achieve sustainable development goals [4, 5].

Recent research recommends that some waste materials and by-products can be used as the main material to produce CMs, or they can partially substitute for the main elements of CMs for various reasons. The use of such materials may not only improve the properties of the CMs but it may also

reduce the costs and environmental concerns. These strategies give rise to a more sustainable production and consumption, profitable waste management and reduction of environmental footprints [5-7].

Mining is the main generator of solid waste in Australia and tailings are the main wastes of mining operations left over after the extraction of ores and mineral processing, being about 60% of the mined ores [8-10]. The most important sustainability challenge and environmental concern in the management of tailings and waste rock is their long-term storage, due to their environmental effects caused by their chemical and physical properties. It has been demonstrated that tailings and waste rocks can be used as aggregates to produce CMs [7-9]. Particularly in underground mining projects, cemented paste backfill (CPB) is a CM designed to place tailings back into the mine and to reduce their surface storage [7, 8, 10]. CPB is a mixture of fine tailings, about 75 to 85% solids by total weight, cement as a binder, typically ranging from 3 to 9% by total solid weight and water for an approximate solid density of 70 to 80% by weight [11, 12]. The underground voids or mined out stopes can be filled by fresh CPB that is delivered using pipelines, and the strength of hardened CPB can be considered to improve the stability and safety of the mine for a specific period. Bearing these issues in mind, CPB production cost typically accounts for almost 20% of all mining costs, with binder cost reaching up to 75 % of that amount [13]. 50-75 % of total CPB operation cost is the cost of CPB mix with 5-9 % binder [14]. It can be stated that only 1% reduction of cement may reduce the overall mining cost by 1.6-2% [13, 14]. In addition to economic benefits, any reduction of cement substantially improves the sustainability of mining projects with respect to costs, resources and Carbon footprint.

Moreover, much research has been undertaken to investigate the applicability and influence of several by-products, such as fly ash (FA), ground granulated blast furnace slag (GGBS or GGBFS), and silica fume (SF) to reduce the amount of cement in CMs [6, 15]. The ground granulated blast furnace slag (GGBFS or GGBS) is a by-product of steel manufacturing which can be used as a cement replacement in concrete so as to enhance its durability, strength-gain and workability. The CO_2 emissions due to the production of concrete decreases by about 45% if 60% of the ordinary Portland cement (OPC) is replaced by GGBFS and recycled water is used, but the cost increases by around 1% per cubic meter [6]. The increase of the cost can be compensated by improving the productivity or sustainability of the considered CM. FA is a coal combustion's waste product which has also been proposed for use in different CMs. The appropriate amount of FA in the mix can improve the strength, workability and durability of concretes such as self-compacting concrete (SCC) [16]. As a partial replacement of OPC, FA contributes to the reduction of OPC in CMs and this accordingly leads to reducing the CO_2 emission and environmental footprints of OPC production [15]. However, if more FA is used than the optimum amount, this may reduce the compressive strength [6, 16]. It is stated that cement and fly ash contents can be adjusted in a concrete mix by considering the hardened concrete properties such as its 28-day compressive strength and durability [4]. The replacement of cement content by fly ash leads to reducing the energy consumption and environmental effects such as CO_2 emission during the production of cement [4].

Other materials and minerals such as meta-kaolin (MK) and lime powder (LP) can also be partially substitute for cement to improve the properties of the target CM or reduce the need for cement and accordingly reduce the relevant carbon footprint. Such materials are also known as supplementary cementitious materials (SCMs) or partial cement replacement (PCR) materials. Overall, they can contribute to the properties of a cementitious mixture like SCC through their properties such as pozzolanic and or hydraulic features, fine and powdery size of particles. Admixtures such as plasticisers or fibres also support the sustainability of CMs by improving their properties and extending their useful applications.

As already noted, several materials have been proposed to be used in CMs to strive for a more sustainable production and consumption. Previous research demonstrates that merely a proper design of such CMs leads to reducing the costs of projects and use of natural resources and reducing the pertinent environmental footprints of projects, thereby improving the sustainability of civil, mining and construction projects [17, 18]. Traditionally, CMs are designed through a trial-and-error process of experiments based on specific experimental methods and conditions. In each trial test, one or some parameters of the system such as proportion and or type of the constituents may change to compare and see how the target objective such as strength or cost of the final mixture changes. This procedure continues until the behaviour of the mixture is well understood or the expected performance and properties are achieved. Experimental mixture design of CMs is often time-consuming, and costly due to the expensive equipment and expertise required. It is also challenging due to the large number of variables and influential factors, and particularly when it comes to multiple objectives and constraints which may contradict. Even after a large number of trials, this way might not lead to an optimum product which can satisfy a myriad of requirements.

In order to reduce the costs and efforts, researchers have been attempting to propose efficient mathematical approaches for indirect estimations [19]. Furthermore, many experiments have already been performed to investigate the performance and properties of CMs, allowing for building experimental datasets. Such datasets are impressively valuable to analyse and make profitable data-driven decisions. It is noteworthy that specifications and standards are also developed based on experimental studies and their recorded data. A smart strategy is to use such data to create mathematical methods for predictive modelling and optimisation aims. Various exciting approaches have been proposed for mathematical mixture design of CMs and composite materials such as statistical mixture design (SMD), design of experiment (DOE), response surface methodology (RSM), and artificial intelligence (AI) algorithms including various machine learning (ML) and optimisation models. These methods may cumulatively provide designers with better ideas about the mixture design leading to reducing the number of trial tests which is considerably cost-effective in projects [19].

Researchers have demonstrated the capability of some artificial intelligence (AI) algorithms and machine learning (ML) methods, for data analysis and modeling of various properties of different composite materials and cementitious mixtures such as concrete [30-34]. These algorithms are inspired

from human intelligence in learning, natural systems, and phenomena such as the behaviour of animals or their neural system and so forth. They enable computers to solve different classification, regression, and optimisation problems in supervised or unsupervised forms. In terms of regression and predictive modeling, several ML algorithms have been proposed such as Support Vector Regression (SVR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR), Genetic Programming (GP), and Ensemble Trees (ET) [32, 35-37]. The concept of developing a model by each algorithm is different. For example, ANN is motivated from the neural system of the body of animals while GP is inspired and computerised based on the theory of evolution. ML algorithms are data-driven and self-adaptive which means that their parameters are iteratively tuned during a training process on a given dataset until the required fitness is obtained. ML algorithms can learn from data to generate high accuracy models particularly when a large number of variables, irregular data and nonlinear relationships exist, each of which makes the problem sophisticated to analyse using other approaches.

In fact, mixture design of CMs can be converted to mathematical design optimisation problems. More importantly, mathematical optimisation frameworks enable the designer to reflect multiple objectives and constraints where experimental design approaches are costly or impossible. In a design optimisation procedure, the requirements of the project such as production cost, strength and slump, or environmental concerns can be considered as objectives or constraints. Such requirements can be modeled using regression and ML methods based on the considered variables, influential factors and experimental data and observations. Finally, an appropriate optimisation algorithm can be used to discover the optimum values of design variables from a pool of solutions, while one or more objectives are maximised or minimised and some constraints are met. In terms of mathematical design optimisation, finding optimal solutions, even when dealing with a single objective, would be a difficult task due to the high variability, stochastic, randomness, and high nonlinearity of relationships and parameters of problems. Importantly, objectives and constraints are commonly contradicting when it comes to multi-objective multi-constraint optimisation. In this regard, metaheuristic algorithms such as particle swarm optimisation (PSO) and genetic algorithm (GA) have become valuable choices for solving complex optimisation problems. Such nature-inspired algorithms are designed to generate and search a space of solutions to find the optimal one iteratively rather than using a deterministic approach. After converting the descriptive statement of each project to a mathematical optimisation problem, such robust optimisation and metaheuristic algorithms can be used to find the best mixture designs meeting the considered objectives and constraints. It is noteworthy that any relationship or issue can be considered in this procedure that make it flexible and more efficient.

Integrating ANNs and optimisation methods was proposed as a good method to find the optimum concrete mixture designs with the lowest production cost and required properties, such as strength and workability [20]. AI-based algorithm such as ANNs and SMD have been applied to estimate the compressive strength and slump flow of ultra-high-performance concrete (UHPC). An optimisation process was conducted to find the batch from within the experimental results where the design variables

were, cement, silica fume, and quartz flour [21]. It was also stated that ANNs can be used as promising prediction tools as they are able to find nonlinear relationships between objectives and design variables for the design of CMs [21]. A nice review of applications of different mixture design methods, both experimental and mathematical, states that AI methods are good choices to find optimum solutions consistent to the needs of projects [19]. According to another research, a framework based on using ML methods can be established to improve the design optimisation of composites [7]. However, reviewing the literature implies that such proposals have not been well conceptualised or applied, and several open questions need to be addressed with respect to multi-objective mixture design optimisation (MOMDO) of CMs.

Overall, there is an increasing demand for engineers and designers to consider multiple objectives and make the best possible decisions at a faster pace while decreasing costs and efforts. Mathematical design optimisation techniques aims to augment the design process and allow for thousands of designs to be analysed and experimented on a computer to find the optimal solution. Traditional mathematical design optimisation procedures such as DOE and SMD are dependent on identifying models and combination of design variables that is extremely difficult due to the large number of influential factors and unknown or nonlinear relationships. CMs consist of precise proportions of materials with specific physical and chemical properties and features that are mixed based on determined methods and conditions. Proposed mathematical optimisation methods commonly consider the proportions of constituents only and ignore their unique features such as chemical and physical properties, or other influential factors such as experimental conditions and programs. Indeed, this concept is correct until the only influential factors as design variables are the proportion of constituents of the mixture.

The predictability of models used as objectives and or constraints in a mathematical optimisation framework is of paramount significance where ML algorithms can be useful. Exciting applications of multiple optimisation algorithms, metaheuristics and AI algorithms have also been represented for different applications and benchmarks. Reflecting constraints properly plays an important role in optimisation algorithms as they narrow down the space of solutions to more optimal ones and enable find the ideal solutions. However, the feasibility and applicability of recent AI and ML algorithms for data-driven predictive modeling and optimisation purposes require more investigations and comparisons in this regard. Although AI methods have been considered as robust tools to find high-accuracy solutions, they are merely methods that must be used to augment the design process. Although several methodologies have been proposed so far, the feasibility of optimisation algorithms and their performance need to be more investigated and validated in finding proper solutions. Validation of the results and solutions, on the other hand, is an important step which is commonly ignored due to its difficulty and complexity. These issues confirm that the MOMDO of CMs still requires to address many issues and open questions. This research aims to engineer the concept of MOMDO of CMs and address gaps and deficiencies of current approaches and come up with new ideas to improve the MOMDO of CMS augmented with AI algorithms for sustainable mining and construction industries.

1-2 **Research objectives and scope**

Given this context, this thesis aims to address the main objectives in terms of MOMDO of CMs for sustainable mining and construction industries as follows:

1. To investigate and compare the applicability of ML algorithms for predictive modelling of objectives for MOMDO of CMs using comprehensive collections of experimental data
2. To investigate the applicability of metaheuristics for solving multi-objective multi-constraint design optimisation problems considered for MOMDO of CMs
3. To properly reflect influential factors such as chemical and physical properties of materials into predictive modeling of properties and MOMDO of CMs
4. Multi-objective design based on multiple economic, environmental and functional objectives and constraints such as workability and compressive strength, production costs and CO_2 emission for mixture design and production of sustainable CMs
5. To recommend engineered MODMO frameworks using integrated ML models and optimisation algorithms
6. To minimise the costs and efforts of mixture design of CMs
7. To optimise the consumption of natural resources, minimise the use of costly materials such as cement for sustainable production of CMs in mining and construction projects

1-3 **The original contribution of the thesis**

In this research, the mixture design of CMs was considered as a mathematical optimisation problem. This research aimed to engineer the concept and propose improved approaches using AI techniques. The robustness of this approach was investigated and validated for the design of various CMs such as cemented paste and hydraulic backfill, HPC, SCC and UHPC. In all cases, different design scenarios and examples were considered based on different functional, economic, environmental factors such as slump and compressive strength, production costs and CO2 emission as objectives and or constraints and the results were analysed and discussed.

Various experimental datasets were collected, and the applicability of different ML algorithms such as Genetic Programming (GP), Support Vector Regression (SVR), Artificial Neural Network (ANN) and Gaussian Process Regression (GPR) was investigated and compared for predictive modelling of complicated objectives such as compressive strength. The accuracy of models used as objectives or constraints is very important and this is why ML algorithms are useful as they are data-driven techniques that can produce high accuracy models.

The mixture design problem is commonly a multi-objective multi-constraint optimisation problem with a set of linear and/or nonlinear objectives and constraints which may conflict. The applicability of metaheuristics such as PSO and GA was verified for mixture design of sustainable CMs in different scenarios. Importantly, the proposed approach enables the designer to pick the best mixture design

amongst Pareto-optimal solutions based on priorities and engineering judgement and expertise rather than one solution which may not be the optimal.

Reviewing the literature revealed that proposed methods commonly consider the proportions of constituents as variables only and ignore other influential factors such as physical and chemical properties, or other influential factors such as experimental conditions and programs. Another contribution of this research was to investigate how an appropriate universal approach using international experimental datasets can be developed in terms of both predictive modeling and multi objective mixture design using AI techniques.

This research and the proposed approach certainly contributes to reduce the cost and effort of experimental designs, particularly when it comes to multiple requirements in terms of sustainability. It also contributes to the optimisation of the consumption of natural resources, a cleaner production with less costs and accordingly a more sustainable project and industry.

1-4 **Thesis structure**

This thesis is presented as a series of papers that has already been published or submitted for publication in order to achieve the research objectives given and discussed previously. Each following chapter reviews the literature, introduces the applied methodology to address the considered problem and discusses the results and improvements with respect to the topic.

Chapter 2 discusses and presents the applicability of linear genetic programming (LGP) for the prediction of the uniaxial compressive strength (UCS) of high-performance concrete (HPC) after 28 days of curing. A comprehensive database is used to develop models. The proportions of constituents of HPC mixture are assumed and considered as the only influential factors. The superiority of LGP over other AI techniques is that LGP can create explicit models that can be converted to mathematical formulae for manual calculations. A sensitivity analysis is done to find the importance of material proportions to prepare high strength HPC mixture.

Chapter 3 presents the mixture design and optimisation of cemented hydraulic backfill (CHB) for different strength conditions in underground mine backfilling as a constrained single-objective optimisation problem. CHB is expected to have a certain UCS at a specific time so that the adjacent stope can be mined out according to the production plan. Genetic programming (GP) is used to develop UCS models that are then used as the objective to maximise or constraints to meet for CHB mixture design. The proportions of constituents of CHB mixture are considered as the only influential factors. A single-objective particle swarm optimisation (PSO) that is programmed to consider any linear or nonlinear constraints is used to find the best mixture designs of CHB that can satisfy the specified strength conditions in a planned time. Some design examples are presented to emphasise the benefits of the optimisation of CHB mixture design using the proposed methodology in terms of production costs and mine planning.

Chapter 4 presents multi-objective mixture design of cemented paste backfill (CPB) using multi-objective particle swarm optimisation (MOPSO) algorithm. With increasing the amount of cement in the mixture, CPB strength and production cost increase together, whereas the workability decreases. This contradict makes the experimental mixture design too costly or impossible to find the best batch. An experimental program is designed and completed to measure UCS, slump and cost of different CPB mixtures. The experimental results are then used to develop models. Objectives are modeled using multiple linear regression approach. The proposed framework here enables the designer to pick the best mixture design amongst Pareto-optimal mixture designs satisfying strength-slump-cost paradox. For this practical application, two hypothetical design examples are considered and the results are discussed. The proposed approach identifies the optimum amount of tailings and cement to prepare a mixture yielding in desired strength and workability with a minimum cost.

Chapter 5 demonstrates MOMDO of steel fiber reinforced ultra-high-performance concrete (UHPC). The production cost of UHPC is considerably high due to the large amount of cement used, and also the high price of other required constituents such as quartz powder, silica fume, fibres and superplasticisers. In order to achieve specific requirements such as desired production cost, strength and flowability, the proportions of UHPC's constituents must be well adjusted. Traditional DOE and SMD methods are discussed and evaluated. This chapter aims to propose a framework by integrating ML algorithms such as Multi-layer Perceptron Neural Networks (MLPNNs) and Gaussian Process Regression (GPR) to develop high-accuracy objective models, and a metaheuristic optimisation algorithm called Particle Swarm Optimisation (PSO) algorithm for MOMDO of UHPC reinforced with steel fibers. The performance of GPR and MLPNN models are compared in creating objective models. GPR is found to be more accurate in prediction. The mixture design of UHPC is converted to a multi-objective mathematical optimisation problem. The mixture designs found by the proposed method for different design scenarios are compared with experimental results and validated to ensure their reliability. The proposed approach not only reduces the efforts in experimental design of UHPC but also leads to the optimal mixtures when the designer faces strength-flowability-cost paradoxes.

Chapter 6 aims to reflect all possible influential factors such as chemical and physical properties of materials and experimental variabilities specimens' size and shape, experimental conditions and methods to develop comprehensive predictive models and MOMDO of sustainable SCC. Proposed ML models and optimisation methods commonly consider the proportions of constituents as variables only and ignore unique features of materials such as physical and chemical properties, or other influential factors such as experimental conditions and programs. A comprehensive database is collected. An optimisation framework is developed by coupling AI algorithms. After comparison of some robust ML algorithms, GPR with exponential kernel function is found to be robust for predictive modeling of complex objectives such as properties of SCC. A non-dominated sorting genetic algorithm called NSGA-II is proposed for MOMDO aims, particularly when it comes to considering multiple objectives and multiple constraints in design scenarios. The proposed method is designed to get accessible

information such as chemical and physical properties of materials, and output the optimal mixture proportions. Various assessments and design examples with different objectives and constraint are considered to confirm the applicability of the proposed framework. It is confirmed that comprehensive models can be developed based on the novel ideas. The underpinning concept of this chapter enables building a universal tool for MOMDO of CMs using international datasets such as sustainable SCC. Using this approach certainly reduces the cost and effort of experimental designs, enables optimum consumption of resources, which leads to a cleaner and more sustainable production.

Chapter 7 aims to provide solution for dimensionality reduction and variable selection as well as the feasibility of developing engineered ML models to use as universal tool for prediction of the properties of CMs. It is strived to compile an experimental dataset with the least uncertainties and the largest number of experimental data. The proportions and type of constituents are reflected properly to develop ML models for prediction of the UCS of an Eco-friendly SCC (Eco-SCC) containing some waste materials and by-products. In terms dimensionality reduction and variable selection, some indices such as hydraulic and pozzlanic reactivity indices are proposed to reasonably to reflect the effects of considered factors in different ML models. The performance of GPR and SVR algorithms are also examined and compared. The GPR model with rational quadratic kernel function outperformed other models. The importance of considering type and properties of materials as well as other influential factors to build comprehensive ML models is precisely demonstrated and the results are discussed. The results confirm that engineered ML models can be developed using international experimental datasets to use as universal tools for indirect estimation of properties of CMs such as Eco-SCC.

Chapter 8 discusses the main findings and key conclusions of this thesis and provides some recommendations for future research.

Chapter 2 An Evolutionary based Prediction Model of the 28-Day Compressive Strength of High Performance Concrete containing Cementitious Materials

2-1 Abstract

High performance concrete (HPC) is a class of concretes which may contain further cementitious materials other than Portland cement such as fly ash and blast furnace slag in addition to chemical admixtures e.g. plasticizers. Strength, durability and rheological properties of the normal concrete are enhanced in HPC. The compressive strength of HPC can be considered as a key factor to identify its quality in concrete technology and construction industry. This parameter can be directly acquired by experimental observations. However, testing methods are often time-consuming, expensive or inefficient. This paper aims to develop and propose a new mathematical equation formulating the compressive strength of HPC samples of 28-day age through a robust artificial intelligence (AI) algorithm known as linear genetic programming (LGP) using a valuable experimental database. The LGP-based model proposed here can be used for manual calculations and is able to estimate the compressive strength of HPC samples with a good degree of accuracy. The performance of the LGP model is confirmed through comparing the results to those provided by other models. The sensitivity analysis is also conducted and it is concluded that the amount of cementitious materials such as cement, furnace slag have more influence than other variables.

2-2 Introduction

Concrete is an extensively-used material in the construction of buildings and civil structures due to its unique properties such as strength, durability, thermal mass and economy. Normal concrete is basically composed of a binder, typically Portland cement, and water which is used to paste fine and coarse aggregates. This mixture hardens after placing and curing and can be utilized for different aims. In order to overcome the deficiencies in the existing normal concrete, various types such as self-compacting concrete (SCC), high strength concrete (HSC) and high performance concrete (HPC) and so forth have been produced through engineering the ingredients or adding new materials. Due to such superior properties HPC is more durable with higher strength and performance concrete which is commonly used in structures subjected to severe exposure conditions and where long-term service-life properties of concrete are expected such as marine structures and highway bridges [22-24]. As is stated by the American Concrete Institute (ACI), HPC is a concrete that satisfies distinct requirements in concrete performance and homogeneity which cannot be typically achieved by traditional materials and mixing, casting and curing procedures [25]. HPC contains some additional cementitious materials such as pozzolanic industrial by-products e.g. fly ash and blast furnace slag in addition to chemical admixtures e.g. plasticizers [23, 26]. Such materials are usually added to the mixture of HPC in an effort to enhance

the engineering properties of normal concrete in both plastic and hardened states such as its flowability, durability and strength. From another perspective, these materials have also been used in HPC to reduce the pertinent production costs and environmental problems. Fly ash might be used in HPC as a replacement of Portland cement to decrease the amount of cement, thereby reducing the production cost and the release of hazardous emissions [23]. Adding such reactive binders reduces the porosity or the permeability of concrete surface. Accordingly, the effects of corrosive and aggressive agents such as chlorides decrease. Moreover, fly ash improves the rheological properties of concrete and moderates its cracking which is due to its less emerged heat of hydration, while the impact of slag is less [24]. A proper use of such cementitious substitutes and fillers also enhances the strength and durability factors through improving the density of concrete. The quality of outcome essentially depend on the interaction of cement and admixture type. Superplasticizers enhance the workability and flowability of concrete with lower water content. It is noteworthy that the compressive strength of concrete increases well, due to the reduction of w/c or w/b ratio [24]. Therefore, these admixtures cause an increase in the compressive strength of HPC to some extent through reducing the water required.

Amongst the properties of concrete, the compressive strength (f'_c) factor is of great significance in structural engineering computations and concrete technology and it usually indicates the quality of concrete. This parameter can be experimentally obtained through lab tests and it is commonly achieved after 28 days from the placing and hardening period. The 28-day compressive strength ($f'_{c\text{-}28}$) of normal concrete is more important than those of other ages since the standard concrete's strength may reach to its 99% in 28 days and the rate and amount of this increase decrease considerably.

Although testing methods provide accurate results and information to make inferences, they are costly and sometimes inapplicable or uneconomical. Besides, there exist several variables and uncertainties which in turn impact the problem and cannot be measured or observed. Hence describing a system or problem by experimental studies would be very difficult and expensive. One of the best alternatives to deal with such difficulties is mathematical modelling. Mathematical models can be generated using various approaches and theoretical concepts. Such models can be used for indirect estimation and approximation aims, e.g. the prediction of f'_c of HPC, with a degree of accuracy and reliability. From another perspective, the established models can be used to evaluate the effect of parameters and to obtain an optimal HPC mixture design such as those provided by [27, 28]. The 28-day compressive strength ($f'_{c\text{-}28}$) of HPC can be formulated as a function of HPC′s ingredients such as the weights of cement (C), fly ash (FA), blast furnace slag (BFS), water (W), fine aggregates (FAg), coarse aggregates (CAg) and superplactisizer (SP). Concrete's age (A) can also be taken as an input variable. Considering these, the f'_c model can be given as follows:

$$f'_{c\text{-}28} = f(C, FA, BFS, W, FAg, CAg, SP, A) \quad (2\text{-}1)$$

In case the mathematical model acquired can be demonstrated as an equation or a formula, its use will not require significant knowledge or expertise or expensive procedure and this is definitely cost-

effective and efficient compared to experimental tests. However, experimental test results and outcomes can be used to compile a database which can be utilized for modeling of a problem using data mining (DM) techniques. DM is the process of determining patterns and models existing between input and output variables in large data sets. For the prediction of $f'_{c\text{-}28}$ of HPC, regression-based analyses have been previously employed to develop predictive models e.g. [26, 29]. It should be noted that in both linear and nonlinear multiple regression analyses (LMR and NLMR, respectively), few models are fitted to the experimental data and this is the major weakness of such methods. The models obtained by regression analyses are not validated for unseen data and the predictability of the model cannot be assessed. Nevertheless, mathematical modeling of the f'_c of HPC as a highly complex composite material would be difficult and challenging due to the existence of many influential parameters and nonlinearity of the problem [23].

Artificial intelligence (AI) algorithms, as a robust subgroup of DM methods, can be selected as useful alternative approaches in terms of estimating and predictive modeling of complicated and nonlinear real-world problems [30-32]. Artificial neural networks (ANNs), Adaptive neuro-fuzzy inference systems (ANFIS), support vector machines (SVMs), genetic programming (GP), gene expression programming (GEP) and the other subgroups are common AI-based prediction tools. These techniques are inspired from the biological activities, analogy of biological systems, and the behavior of natural systems. These data-driven and self-adaptive algorithms are programmed to mathematically simulate the problem and find the structure of the predictive models relating independent input to corresponding output variables.

Researchers have demonstrated the successful applicability of such methods for finding solutions for civil engineering problems [33-40]. Machine learning and artificial intelligence approaches have recently been utilized for predicting the f'_c of concrete using proportions of ingredients considering their significance in industrial projects18. These methods have been utilized for indirect estimation of the f'_c of HPC. Some of the relevant researches such as [23, 26, 39, 41-44] carried out for predicting the f'_c of HPC samples are summarized in Table 2.1. Besides, the methods and the variables incorporated in the modeling process and the HPC sample size existing in their databases are also demonstrated.

Table 2.1 Some AI methods used to model the f'_c of HPC samples of different ages.

Reference	Method	Sample size	Model type
[26]	ANN	727	Black-box
[23]	ANN and SVM	1030	Black-box
[41]	GEP	1133	White-box (Equation)
[42]	Evolutionary fuzzy SVM	1030	Black-box
[43]	ANN	864	Black-box
[44]	M5P model tree algorithm	1912	White-box

It is noteworthy that the AI-based models developed by researchers in Table 2.1 typically account for C, FA, BFS, W, FAg, CAg and A as independent input variables for the prediction of f'_c of HPC samples of different ages. The measurement unit of each variable, i.e. C, FA, BFS, W, FAg, CAg, and SP in SI system, is kg/m^3 and that of A which is the age of HPC sample is in day when the f'_c factor was tested. Although ANNs and SVMs have gained very much attention for solving engineering problems, these approaches have often been criticized due to the fact that they produce black-box models. In other words, a transparent structure of the predictive model generated by such algorithms may not be presented as an explicit formula or an equation so as to be used by others for further calculations. The model constructed by [44] based on M5P model tree algorithm was very complex. In other words, it consists of several equations which should be calculated to acquire the final result for the prediction of f'c of HPC. In Table 2.1, amongst AI methods utilized for the prediction of f'_c of HPC, a formula has been proposed by [41] using GEP algorithm. GEP is also a subgroup of EAs and is used for prediction, pattern recognition and classification purposes. The obtained model by GEP for the approximation of f'_c of HPC samples in different ages is as follows:

$$f'_C = \frac{-5.69(2.89+(2.64BFS-3.19A+FA)/CAg)(8.72SP+1.93A-33.60+C)(16.08SP+A-BFS-196.26-FAg-F)}{FAg(W-(C-CAg+A)/A)} \quad (2\text{-}2)$$

This paper explores the applicability of linear genetic programming (LGP), a DM technique and a subgroup of EAs, as an alternative technique for predictive modeling of the 28-day compressive strength ($f'_{c\text{-}28}$) of HPC samples containing cementitious and chemical admixtures, i.e. BFS, FA and SP. The superiority of LGP approach over other AI techniques is the fact that it is able to create models. The models generated can be converted to mathematical equations or explicit formulae which can be utilized for further design and manual calculation purposes. In addition to the accuracy of LGP models, this is a great superiority of this approach compared to other DM or AI based prediction algorithm. Here, a new equation will be proposed for the prediction of $f'_{c\text{-}28}$ of HPC samples using a reliable database consisting of 425 test results performed on standard samples of HPC of 28 days age. The LGP model's accuracy will be investigated through various validation and verification analyses and the results will be compared to those existing models provided by other researchers.

2-3 Linear genetic programming

Genetic programming (GP) is a subgroup of biologically inspired AI algorithms which is based on the combination of genetic inheritance notions expressed in Darwin's theory of evolution for mathematical modeling of real-world problems. GP and its subclasses aim to recognize models linking their input variables to corresponding output(s) using experimental and observed data. The process of finding the optimal model is dependent on the concept of survival of the fittest model or program. GP outputs are computer programs or mathematical models which are generated in tree-shaped models [45, 46]. Traditional GP or tree GP (TGP) constructs tree-shaped models and these can be read through a rarely used programming language i.e. LISP [47-50]. LGP algorithm can be considered as a branch of GP which develops variable-length sequences of an imperative programming language, C or C++ [48, 51]. The functions commonly used in GP and LGP include arithmetic operations, e.g. +,-,/ and ×, conditionals such as if-then statements, and some of Trigonometric functions, Sin and Cosine. The terminals are constants and variables. A TGP and an LGP model are given in Fig. 2.1.

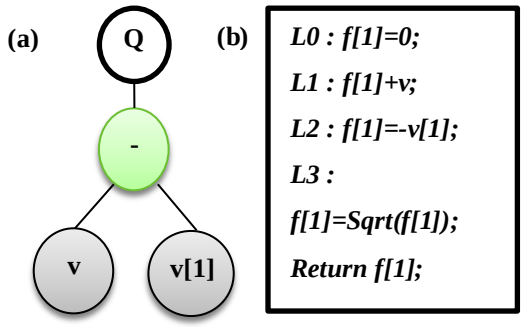


Fig. 2.1 (a) A TGP and (b) an LGP model indicating the equation of $f(v[i]) = \sqrt{v[1] - v[2]}$ where i is the number of input variables.

2-4 Input-processing-output system in LGP

Generally, LGP begins with generating an initial population of programs or models. The accuracy of models is assessed by a fitness function which can be defined as the summation of the residual error (RE) values existing between LGP predicted and experimental outputs. After running a tournament, programs are randomly chosen from the initial individuals. Those members having better fitness acquire a greater opportunity in order to come to be a parent for the next generation. Then, those members selected are reproduced or transformed randomly to new programs by means of genetic operators. Generally, there are three kinds of genetic operators, namely crossover or recombination, mutation, and reproduction, each of which has a modification impact based on preset probability values. Reproduction is basically carried out by duplicating a selected program within the existing population to the next generation without modifications where there is a desired fitness. Mutation operation is applied for one program randomly selected from the pool of members in a generation. Thereafter, a node is exchanged

and a new spring/child is generated which is a new program. Mutation may modify operators or a part of a model, e.g. changing v[1]/v[1] into v[1]+v[1] or v[1]-v[3] to v[3]×v[1] where v[i] is the i^{th} input variable . Hence, new models can be generated using genetic operators and their fitness will be assessed to be used for the next generation. Crossover or recombination performs by replacing one or more parts of programs between two parents to generate new offsprings or models. Crossover takes place between two models, whereas mutation operates on a single randomly chosen model. After the evaluation of the programs in terms of their fitness, the procedure repeats till the stop conditions are reached and the best program is obtained [32, 48, 50].

2-5 **Dataset and model variables**

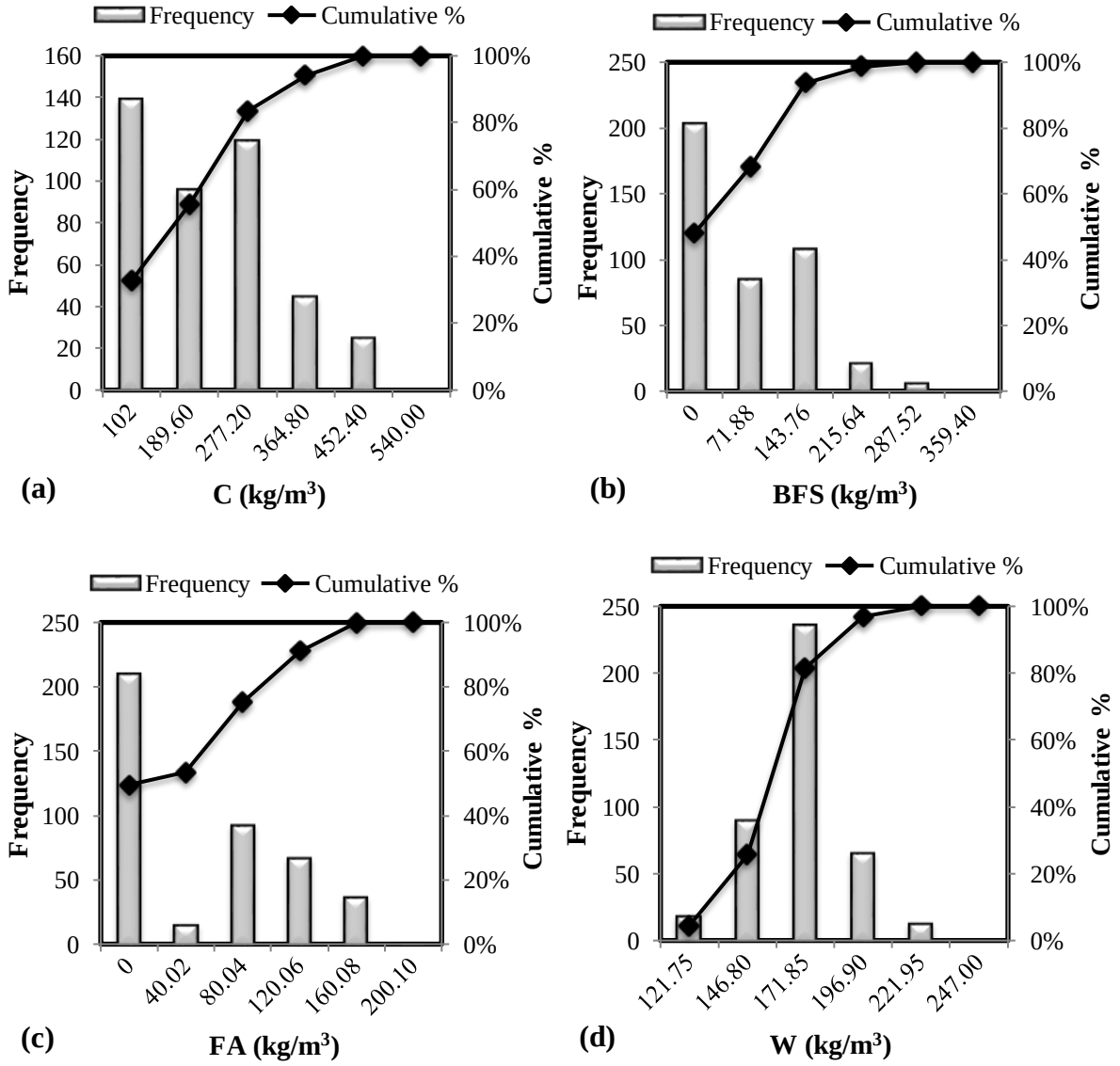
LGP and other prediction AI methods use data to extract the models existing between input and dependent or output variables. A comprehensive database consisting of 1030 experimental tests conducted on HPC samples of various ages between 1 to 365 days, in order to obtain their f'c values was published by [26]. The database contains 9 variables of C, FA, BFS, W, FAg, CAg, SP, A and f'_c. Although the age of HPC sample was given in the database and was considered by some researchers as an input variable, as was demonstrated in Table 2.1, in this study the 28-day f'_c of HPC samples which is illustrated by $f'_{c\text{-}28}$ is considered as an input variable. Therefore, the age parameter (A) is considered to be 28 and the data set was narrowed down into 425 data. It is noteworthy that $f'_{c\text{-}28}$ is a more standard value of the compressive strength of HPC in structural computations and concrete technology. In order to have a general statistical perspective on data set, a descriptive statistical analysis was done and the relevant indicators are summarized in Table 2.2.

Table 2.2 Statistical indices of variables utilized for predictive modeling using LGP.

Parameter	**Input Variable (Predictor)**							**Output (Target)**
	C (kg/m³)	***BFS (kg/m³)***	***FA (kg/m³)***	***W (kg/m³)***	***SP (kg/m³)***	***CAg (kg/m³)***	***FAg (kg/m³)***	***$f'_{c\text{-}28}$ (MPa)***
Mean	265.44	86.29	62.79	183.06	6.99	956.06	764.38	36.75
Standard Deviation	104.67	87.83	66.23	19.33	5.39	83.80	73.12	14.71
Range	438.00	359.40	200.10	125.25	32.20	344.00	398.60	73.22
Minimum	102.00	0.00	0.00	121.75	0.00	801.00	594.00	8.54
Maximum	540.00	359.40	200.10	247.00	32.20	1145.00	992.60	81.75
Sum	112813.44	36671.13	26687.75	77800.11	2972.71	406325.13	324860.07	15618.10
Count	425	425	425	425	425	425	425	425

LGP and other AI-based prediction algorithms are data-driven methods and this means that they calibrate models using data. In addition, the frequency of a particular observation value is the number

of times the data value occurs in tests results. For more details regarding the range of different variables existing in the database, the frequencies of each variables is calculated and are illustrated in Fig. 2.2.



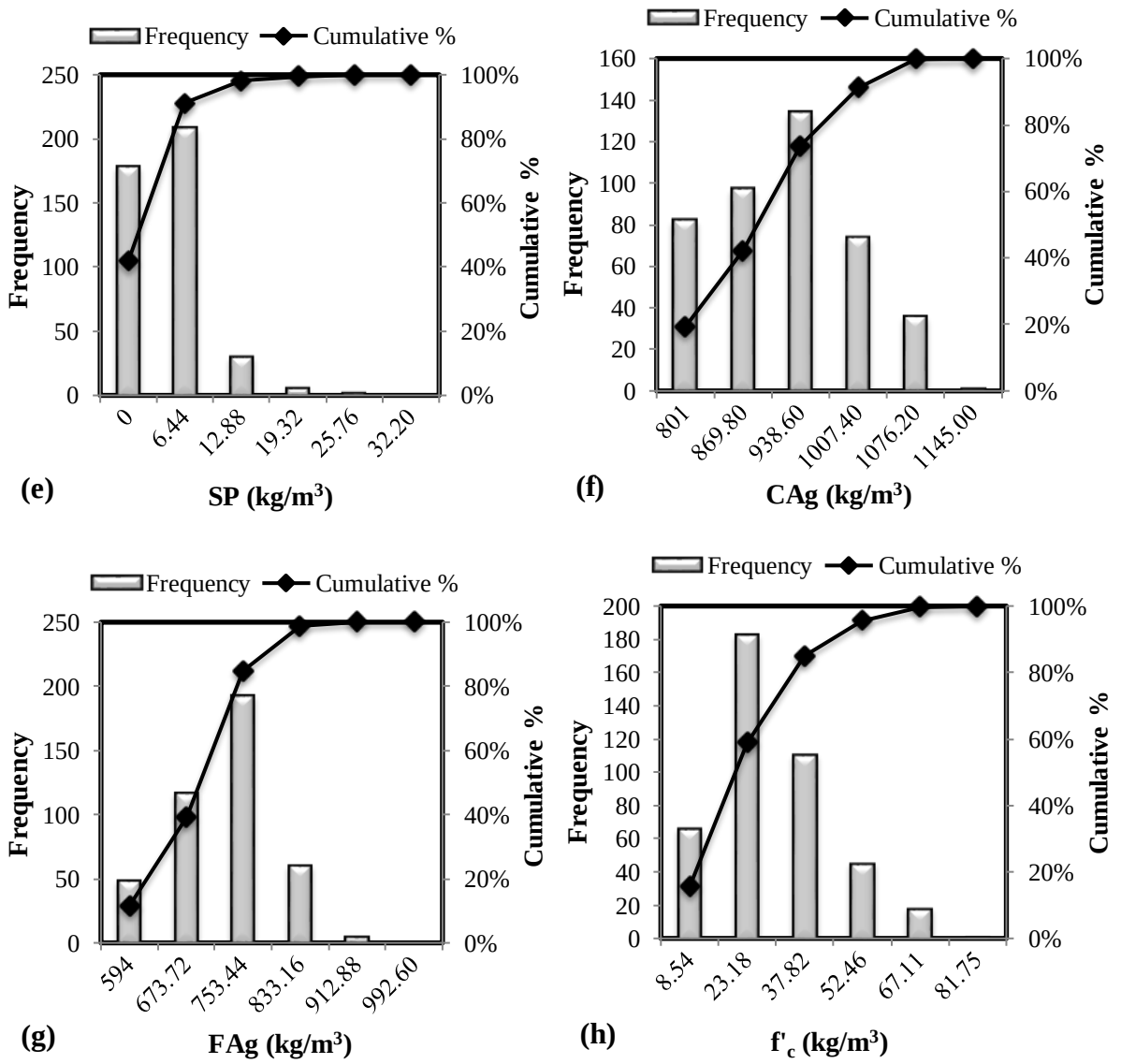


Fig. 2.2 Frequencies of data for each variable existing in the employed database.

The predictability of AI-based models is extremely dependent on the number and the range of datasets used in model development. It is suggested that in case the ratio of the number of objects over the number of selected variables has been more than five, the obtained model can be regarded as a safe model [32-34, 52]. Here, this ratio is obtained considerably higher and equals to 425/5=85 which shows the fact that the number of data can be accepted to be sufficient for the development of LGP model. However, the more number of data, the more accurate model will be achieved.

2-6 **Data division for model development**

In terms of the fitness and predictability of LGP models, researchers recommended that the fitness must be obtained for different divisions of data. In this regard, the data set can be initially classified into two groups, training and testing data [25, 34, 50, 53]. The training group is used to evolve and fit the models and the test one is employed to assess the fitness of trained models. Therefore, the test data set is new

to the process of modeling and it reveals the predictability and the generalization of the model. It is noteworthy that the data can be classified to more groups but two groups would be sufficient. In order to classify the data set, researchers suggest that the 80% of the data are taken as the training and the remaining 20% of all data are used for testing the developed model through soft computing techniques [51, 53]. Similarly, these values are considered in this study to explore the accuracy degree of the generated models.

It is obvious that in case a model can estimate the output with a high degree of accuracy, i.e. the model output results fit on the observed data, it means that the model does not underfit. Additionally, if the model fitness on testing data is obtained acceptable and close to that of training data, it can be concluded that overfitting does not occur [32, 50]. Moreover, it is considerable that the models generated by multiple linear or nonlinear regression analyses are not checked on different classes of data. Regression-based models are often achieved after fitting merely few models on data and it would not be generalized for new data which is the model predictability. Hence, it is highly expected that such models overfits and this is their major weakness.

2-7 **Adjustment of algorithm parameters**

A software [54] was employed to evolve LGP-based models for indirect estimation of the $f'_{c\text{-}28}$ of HPC samples. There exist several parameters in AI algorithms which should be fixed or pre-defined to perform a run. These parameters should be initialized and then adjusted to produce a perfect model. In order to adjust these parameters' values, numerous runs might be required to be done through a process of trial and error. Albeit, those proposed values of the initial parameters by other researchers can also be used [30]. Here, this process was well performed through initiating the parameters given in Table 2.3.

Table 2.3. The optimum values of parameters adjusted for the proposed LGP model development.

Training Parameters	**Values**
Population Size	1000
Arithmetic Operators	+, -, ×, /
Mathematical functions	sqrt (√)
Random constants	-10 to 10
maximum program size	128
number of demes	20
Probability of crossover	0.5
Probability of mutation	0.95

As the process of evolving LGP models is completely random, being successful to achieve an acceptable model may depend on the program or model size. Though, the best model would not be merely the

largest one in size. In this way, it can be expected that all variables are available in the model and a better model can be obtained. The significant problem is the size of produced programs which gradually increases without any improvement in their accuracy after a while and this is known as bloating. To rise above this, LGP restricts the size of programs generated to a number of orders [49, 50, 55]. Additionally, demes, i.e. semi-isolated subgroups, classify program populations of equal size and this results in better progress of generation in each run in LGP [46, 47, 50, 52]. These are some remarkable distinctions of LGP in comparison to other variants of GP which result in speed-up the algorithm progress and generating higher precision models.

2-8 **Fitness evaluation of generated models**

Generally, some statistical indicators are commonly employed to evaluate the precision of prediction models. If the correlation coefficient (R) between the predicted values of the output using LGP model and those measured becomes equal or more than 0.8, it can be proposed that the accuracy of the model generated in estimation is acceptable [25, 51, 56, 57]. In addition to this, the mean absolute error (MAE) and the root mean square error (RMSE) can be used to provide an insight to the overall error. These parameters are sensitive to the errors between the measured and predicted values of the target. Obviously, the less the MAE and RMSE values, the better the accuracy of the model will be. Here, these recognized indicators are chosen as tools for the accuracy assessment of the LGP-based models and to select the fittest model.

2-9 **The proposed LGP model**

A comprehensive process of trial and error was made to develop the best LGP model for the prediction of the $f'_{c\text{-}28}$ of HPC samples in this study. Many runs were made and numerous LGP programs were generated through modifying various combinations of the algorithm parameters. Amongst all models, the selection of a good model initially depends on fulfilling two criteria as a multi-objective strategy [30, 32]. Firstly, the simplicity of the model structure and secondly, the optimal accuracy of the model on the training and testing data sets. It is noteworthy that some other criteria should be considered in order to evaluate the models. This procedure (model development and accuracy analysis) was performed several times to achieve the proposed LGP model in the present. The fittest $f'_{c\text{-}28}$ program generated by LGP is in C or C++ code after the training process and is acquired as follows:

```
float DiscipulusCFunction(float v[])
{
  long double f[8];
  long double tmp = 0;
  int cflag = 0;
```

```
f[1]=f[1]=f[1]=f[3]=f[4]=f[1]=f[6]=f[58]=0;


L0:   f[1]-=v[4];
L1:   f[1]*=f[1];
L2:   f[1]-=v[1];
L3:   f[1]*=f[1];
L4:   f[1]/=v[1];
L5:   f[1]+=v[1];
L6:   f[1]*=-1.9f;
L7:   f[1]*=-1.6f;
L8:   f[1]*=f[1];
L9:   f[1]*=v[4];
L10:  f[1]/=v[3];
L11:  f[1]=sqrt(f[1]);
L12:  f[1]+=v[1];
L13:  f[1]-=v[3];
L14:  f[1]-=v[1];
L15:  f[1]-=v[3];
L16:  f[1]+=v[6];
L17:  f[1]*=0.1f;
L18:  f[1]-=v[3];
L19:  f[1]*=1.5f;
L20:  f[1]/=0.6f;
L21:  f[1]+=v[1];
L22:  f[1]+=v[1];
L23:  f[1]*=f[1];
L24:  f[1]=sqrt(f[1]);
L25:  f[1]+=v[1];
L26:  f[1]+=v[1];
L27:  f[1]+=v[1];
L28:  f[1]/=0.7f;
L29:  f[1]+=v[1];
L30:  f[1]*=0.035f;
L31:

if (!_finite(f[1])) f[1]=0;
```

```
  return f[1];
}
```

In this program, v[1], v[1], …, v[6] are, respectively, C (kg/m^3), BFS (kg/m^3), FA (kg/m^3), W (kg/m^3), SP (kg/m^3), CAg (kg/m^3) and FAg (kg/m^3) where the output is $f'_{c\text{-}28}$ (MPa). Herein, the C++ code is simplified and converted into an equation through replacing the variables. The optimal LGP-based formula of $f'_{c\text{-}28}$ (MPa) is given as follows:

$$f'_{c-28} = 0.035\left[4C + 2.8BFS + FA - 4.12W + 0.4FAg + \left(\frac{(SP^2 - CAg)^2}{CAg} + C\right)\sqrt{\frac{SP}{W}}\right] \qquad (2\text{-}3)$$

The proposed LGP model is acquired after conducting a great process of training and validation. The $f'_{c\text{-}28}$ model is selected among a total of 10, 587,115 programs produced after several runs and iterations. The accuracy of this model is also validated for a group of unseen data in the modelling process which presents the generalization performance and predictability. It is worth remarking that the advantage of LGP approach over other SC techniques, e.g. ANNs and SVMs, is the fact that it can produce predictive models which can be converted to explicit formulae to be utilized for further analyses [50, 51].
Results and Discussions

2-10 **Initial evaluation of the model performance**

As was mentioned previously, during and after the process of modeling and performing the training and validation procedures, several models are generated by LGP. The fitness of LGP models is evaluated by some statistical indices, e.g. R, RMSE and MAE. Additionally, in order to explore underfitting and overfitting problems, the performance of the optimal model should be examined for different classes of data, i.e. training data and test data. This type of model assessment is also called the generalization performance or the predictability assessment of the model [35]. A traditional way to observe the accuracy and fitness of a model is to indicate the model output results versus the measured or experimental values [32, 33, 35]. Fig. 2.3 illustrates a scatter plot of the obtained values of the f'c-28 of HPC samples estimated by the LGP for training and test data sets. Moreover, in order to have a general insight to the accuracy of models, the pertinent statistical indices are also given.

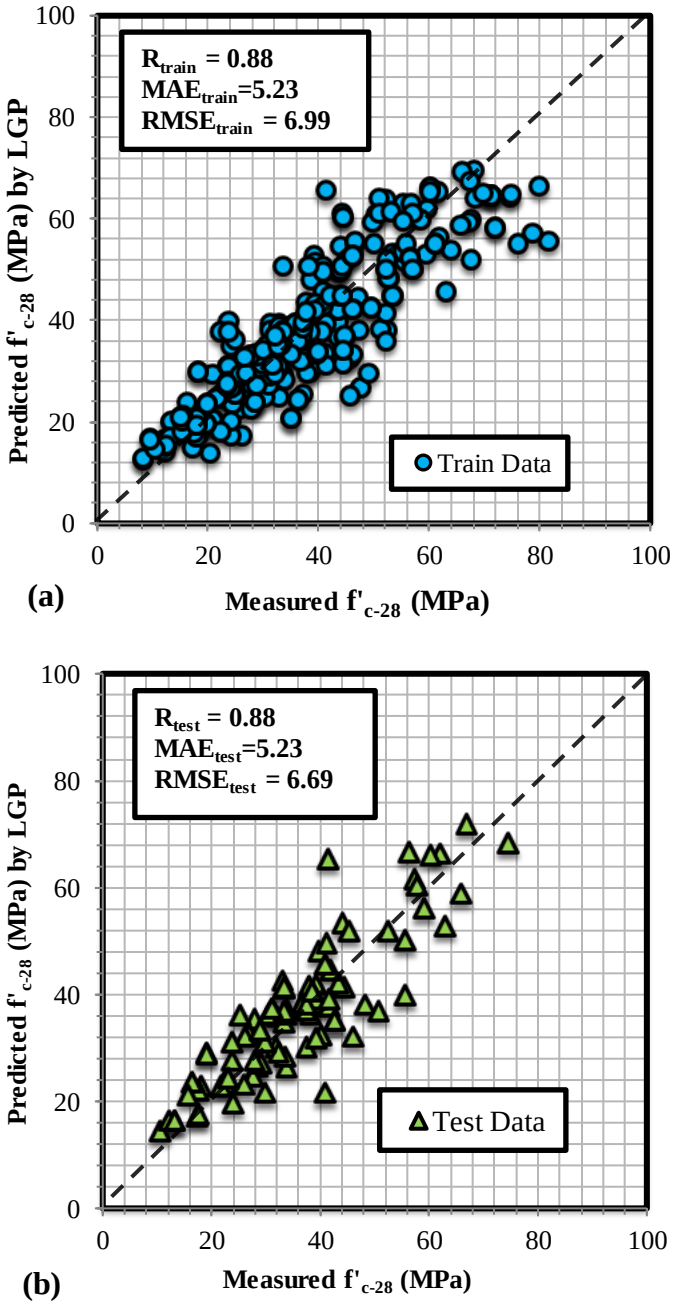


Fig. 2.3 Predicted versus experimental values of f_{c-28} by LGP model (a) training data and (b) test data.

As is shown in Fig. 2.3, the LGP model satisfies the criterion of R>0.8 recommended by researchers which means that the model can be accepted in terms of the accuracy. The obtained values of R, RMSE and MAE for the training and test data given in Fig. 2.3 indicate that the LGP model does not underfit which means that it is capable of estimating f'_{c-28} of HPC samples with an acceptable degree of precision. Additionally, the closeness of R, RMSE and MAE values calculated for training and test data implies the fact that overfitting is avoided. It means that the LGP model is well-trained and also the predictability of the model is acceptable.

2-11 **Additional evaluation of the model**

The model fitness assessment parameters such as R, RMSE and MAE and so forth mainly provide an overall perspective on the fitness and accuracy of the prediction model and these indices are very much sensitive to the scales of error[32, 57]. To tackle this, a proper way of evaluating the mismatches existing between the predictions obtained by models and experimental values of $f'_{c\text{-}28}$ of HPC is observing of the error, defined as the residual error (RE) or fitting deviation. The RE is an obvious estimate of the unclear statistical error generated by the model or the values represented by R, RMSE and MAE. In this regard, Fig. 2.4 shows a line graph of the RE values produced by the $f'_{c\text{-}28}$ of HPC LGP model and those obtained by the GEP mode for all data. Values of R, RMSE and MAE are also calculated for both LGP and GEP models to have a comparative perspective.

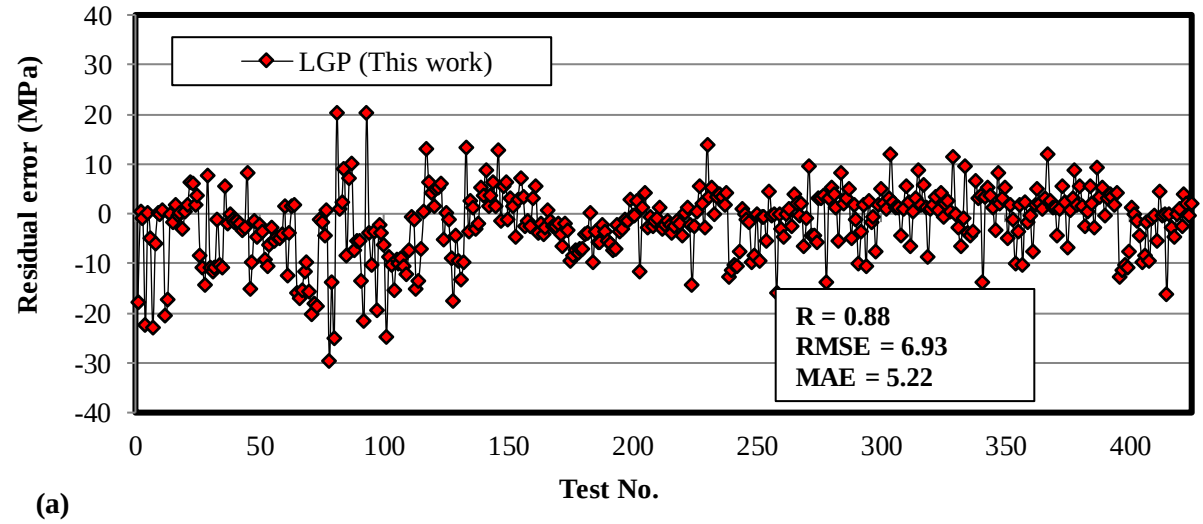


Fig. 2.4 Line graphs expressing the RE values acquired by various $f'_{c\text{-}28}$ estimation models.

Considering the RE values calculated for each model in Fig. 2.4, it can be observed that the LGP model with (R=0.88, RMSE=6.93 and MAE=5.22) estimates the $f'_{c\text{-}28}$ factor better than the GEP model with (R=0.85, RMSE=7.71 and MAE=6.02) which is also and EA prediction algorithm. Obviously, these models are of a good predictability and this confirms the capability of EA and the theory for solving engineering problems.

2-12 **Sensitivity analysis**

A significant question which should be answered prior to the use of a model for estimation aims is the fact that how the predicted values vary as the values of each input variable changes. This also represents the importance rank of variables in a formula and which should be considered by the users. In order to investigate this, a sensitivity analysis (SA) may be of use. This kind of analysis of the model indicate how predictors impacts the changes of the output variable [32, 33, 35, 50]. The sensitivity index (SI) is the output of an SA which can be acquired as follows:

$$N_i = f_{max}(x_i) - f_{min}(x_i) \tag{2-4}$$

$$SI_i(\%) = \frac{N_i}{\sum_{j=1}^{n} N_i} \times 100 \tag{2-5}$$

In these equations, f_{max} (x_i) is maximum value of the LGP model prediction for the i^{th} input variable x, while f_{min} (x_i) is calculated minimum value. These values are calculated by the model where other inputs incorporated, except x_i, are considered constant at their mean values and n is the number of independent input variables in the model.

It is worth noting that those variables with higher SI value are merely more influential than other variables and their changes further change the value of the output. Herein, the SA process is done for the LGP and GEP model in order to have a comparison and the obtained SI values are illustrated in Fig. 2.5.

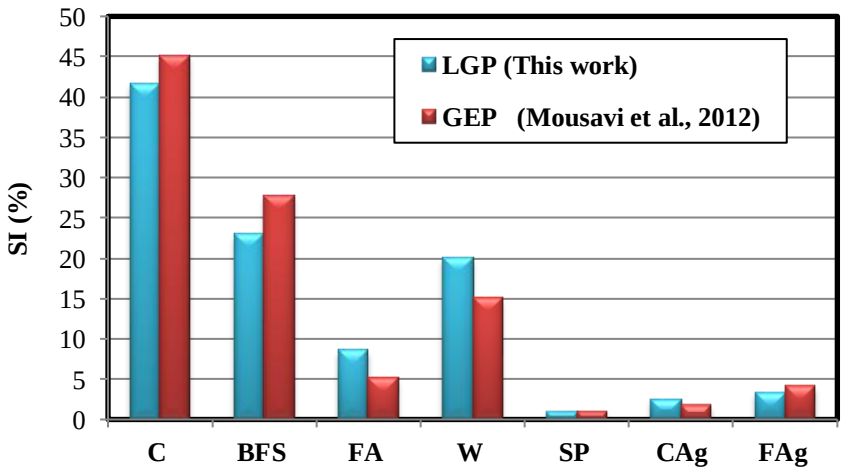


Fig. 2.5 SI (%) values of incorporated variables in LGP and GEP model of $f'_{c\text{-}28}$.

As can be seen in Fig. 2.5, the SI (%) values obtained by LGP and GEP models present that C, BFS, W and FA are respectively the most influential input variables in the models proposed. It is also true theoretically since these materials acts as the binder in concrete. SP, CAg and FAg slightly impact the changes of the LGP and GEP models for the prediction of f'c-28 of HPC samples. Additionally, these results conform to those provided by researchers, e.g. [24, 25, 41]. It should be noted that the SI value for variables incorporated in a model is unique for every single estimation models and this may change for other models. In this regard, the true point is to have a viewpoint before using a model.

2-13 **Conclusion**

This paper aimed to introduce the LGP algorithm as a robust alternative approach for predictive modeling and numerical analysis of $f'_{c\text{-}28}$ of HPC samples having different combinations of cementitious materials and chemical admixtures. The number of variables existing in HPC and the amount of its ingredients impact the quality of HPC and in turn influence on its compressive strength. The LGP model was acquired as a C++ code and was converted to a mathematical equation so as to be explicitly used for hand-calculation and further analyses. The predictability and performance of the proposed LGP model was indicated through performing some analyses and the results were confirmed via comparing to those obtained by another model obtained by GEP approach in the literature. It is notable that the sensitivity of each model to the changes of its inputs is a unique factor which should be considered prior to its use. It was represented that C, BFS, W and FA, respectively, impacts the f'_c of HPC obtained by LGP model in this paper. The proposed LGP model is useful in achieving an optimal HPC mixture design, estimating the f'_c of HPC, particularly for pre-design aims. It is noteworthy that models with better accuracy can be generated by LGP if more data and more number of descriptive variables are available.

Chapter 3 The Optimisation of Cemented Hydraulic Backfill Mixture Design Parameters for Different Strength Conditions Using Artificial Intelligence Algorithms

3-1 Abstract

Cemented hydraulic backfill (CHB) is widely used in underground mine backfilling, especially when regional stability is required. One of the critical backfill properties is uniaxial compressive strength (UCS), the maximum axial compressive stress that the sample can withstand before failing. CHB is expected to have a certain UCS in a specific time so that the adjacent stope can be mined out according to the production plan. To reach the desired strength in a specific time, there exit parameters which must be well adjusted. To increase the strength of CHB, either the cement dosage could be increased or a longer curing time could be allowed using less cement and more tailings. However, increasing cement content significantly increases the operational mining cost. If there is enough curing time for the planned production then less cement and/or more tailings can be added to get the desired strength at a reduced cost. This paper investigates the applicability of artificial intelligence (AI) algorithms to optimise key parameters of CHB design so that the desired strength would be reached in a specific time. Genetic programming (GP) is used to generate models relating the UCS factor to CHB's key parameters using an experimental database. The generated GP models are then used by a particle swarm optimisation (PSO) algorithm in order to determine the amounts of CHB's parameters which can satisfy the specified UCS conditions in a planned time. Some examples are presented to emphasize the benefits of the optimisation of CHB mixture design. Using the presented approach, it is possible to optimise CHB design parameters by considering mine production plan, requiring certain UCS at specific ages, and to reduce the cost.

3-2 Introduction

Tailings and waste rock are the main byproducts of mining operations generated during the extraction of ores and mineral processing. These wastes must be treated according to environmental and safety concerns. A cleaner waste management solution is to send these materials back into the mine in order to fill the mined out stopes [59, 60]. These wastes are mixed with a binder, usually cement, in order to provide local support and stability allowing for a safer and more efficient mining operation. In this regard, cemented hydraulic backfill (CHB) is widely used, which is composed of tailings, crushed waste rock, water and a small dosage of cement to satisfy the strength conditions required in the project [60-62].

In underground mining operations CHB must remain stable during the production of any adjacent stope. If it does not have enough strength, adjacent faces may relax and fall into the open stope. To measure the strength, the uniaxial compressive strength (UCS) test is commonly used. In designing CHB, both

short and long term strengths are important from the mine planning perspective. If CHB does not gain enough strength when the adjacent stope/s are mined out, the possible consequences are mine dilution, collapse of a stope, and even large scale hanging wall or footwall failure threatening the global stability of the mine. The solid content (SC), cement dosage (CD), aggregate to tailing ratio (ATR) and time are the key parameters controlling the UCS of CHB. Cement is the most expensive component of CHB and it can be considered as a notable percentage of the total cost of mining operation [59-61]. To decrease cement cost, it is desirable to use more tailings as replacement material.

From the mine production perspective, to mine out the adjacent stope without jeopardising the overall stability and production plan, two or more requirements may exist in terms of UCS, at different timeframes. For example, a UCS of over Y_1 (kPa) might be required in short term (T_1 days) and the long term (T_2-day) UCS should be more than Y_2 (kPa). In order to have a specific UCS, an appropriate mixture design is required. Conventionally, the amounts of CHB's ingredients for such problems are obtained by doing tests based on a process of trial and error. Given that mine planning is a dynamic process depending on the mine production plan, a large number of trial tests should be carried out until the best batch is formulated for each case or stope. Thus, the recipe of backfill mixture design may be changed or optimised frequently during a mining project depending on the production plan. This is a time consuming and costly operation. However, without such optimisation more cement than necessary can be consumed leading to serious economic loss. Theoretical and mathematical methods can be utilized to find near-optimal solutions, to reduce the number of trial mixtures and to have a more economical and engineered mixture design during a project.

The indicated problem is an optimisation problem which is the process of finding the best solutions when one or more objectives are achieved. There exist several steps for formulating and solving such an optimisation problem. Firstly, a proper problem statement must be developed. Then, the problem should be formulated in which the verbal or qualitative statement of the problem is converted into a mathematical or quantitative form. The next step is to collect data and information relevant to the problem. Then, the designer should utilize appropriate methodologies to solve the problem which may involve different numerical modelling and optimisation methods. This paper aims to demonstrate the applicability of artificial intelligence (AI) algorithms to find feasible solutions to such problems. For this aim, an example problem is formulated to be resolved by AI algorithms. The problem assumed here is to determine possible CHB batches yielding a desired UCS at a specific time with reduced cost for different conditions. Genetic programming (GP) is used to find the models existing between different variables in the database. The generated GP models are then used as objective functions in an optimisation process using a particle swarm optimisation (PSO) algorithm, in order to make a back analysis and find near-optimal quantities of CHB's ingredients.

3-3 Basic concepts of constrained optimisation

Mathematically, a single-objective optimisation is the selection of the best-input values from within a range when an objective is being maximised or minimised. In the process of mathematical optimisation, objectives and constraints are the mathematical models which might be linear or nonlinear in engineering problems. By means of the constraints, the space of solutions is narrowed down, thereby the algorithm is directed to the fittest solution. Constraints can be functions or relationships between two or more design variables considered in a problem. A typical engineering constrained optimisation problem can be defined as follows [63]:

Design variables: $X == [x_1, x_2, x_3, \ldots, x_i, \ldots, x_n]^T$

Objective: minimise or maximise f(X)

Subject to: $a_i \leq x_i \leq b_i$

$g_k(X) = 0$ (k =1,2, …, l)

$h_j(X) \leq 0$ (i =1,2, ..., p)

where a_i and b_i are lower and upper bound of x_i, n is the number of design variables, l is the number of equality constraints which can be functions of X represented here by g_k and p is the number of inequality constraints which can be functions of X represented here by h_j.

3-4 Genetic programming

Here, genetic programming (GP) is utilized to develop models which are then used as objectives in the optimisation process. GP uses the concepts of the theory of evolution and genetic inheritance to recognize and produce the fittest linear or nonlinear models existing between input and output variables in a database [33, 61, 64]. Technically, these models which are computer programs (so-called tree-expressions (ETs)), have tree-shaped structures and can be developed in computer using programming languages e.g. LISP [33, 64]. GP models can be simply converted to mathematical expressions, which can be used for further calculation aims. This is a useful superiority of GP over other AI algorithms such as support vector-machines (SVMs) or artificial neural networks (ANN). Fig. 3.1 illustrates an ET and the converted mathematical equation.

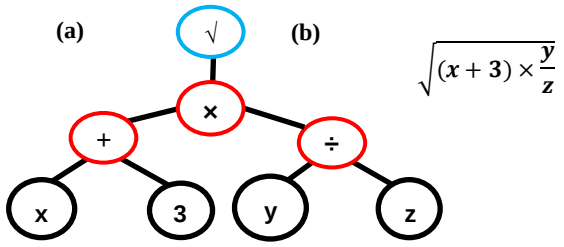


Fig. 3.1 A GP program in the form of (a) an ET (b) a mathematical equation.

As is shown in Fig. 3.1, an ET is composed of different nodes. Operation nodes may do unary operations such as abs, sin, cos, tan, sinh, cosh, exp and log, or binary operations such as add (+), multiply (×), and divide (/). Nodes without children are called leaf nodes or terminals that represent input values or evolved constant values within the system [33, 61, 64].

Generally, GP starts with the initialization of a random number of models. After evaluating the fitness of programs and ranking them, those with better fitness are reproduced or selected as a parent for the next generation. The programs are randomly changed into new program mutations, recombination or crossovers. The crossover operation changes random parts of selected pairs (parents) to generate new and different programs (offspring). Mutation involves changing some random parts in only one program. Finally, the fitness and the accuracy of programs are assessed and the process stops once the desired fitness is reached or repeats until finding the fittest program or model [33, 61, 64].

3-5 **Single-objective particle swarm optimisation algorithm**

In the present study, a single-objective particle swarm optimisation (PSO) is used to find optimal solutions to the problem. PSO is a population-based search algorithm, which is inspired from the social behaviour of birds or fish seeking food [63]. Each particle (member) of a swarm (population size) updates its search pattern in line with its own experience and the success of the others in its group or in the neighbourhood. A particle, p_i, in PSO has a position (x_i) and velocity (v_i). Assume that $\mathbf{x}_{i(t)}$ represents the position of particle p_i, at iteration t, with a velocity of v_i. x_i represents a solution obtained by the particle and v_i shows the rate of changes of the current particle [63]. x_i and v_i are randomly initialised. Then, the position of x_i is updated as follows:

$$x_{i(t)} = x_{i(t-1)} + v_i(t) \tag{3-1}$$

The velocity vector reflects the exchanged information and is defined as the following equation:

$$v_i(t) = Wv_i(t-1) + C_1 r_1\left(x_{pbest} - x_{i(t)}\right) + C_2 r_2\left(x_{gbest} - x_{i(t)}\right) \tag{3-2}$$

where $x_{pbest\ is}$ the best position found so far by the particle and x_{gbest} is the best position found by the neighbouring particles. C_1 and C_2 are learning factors defined as constants, r_1 and $r_2 \in [0, 1]$ are randomly generated values, and W is the inertial weight defined within the algorithm. These definitions and indicators have also been extended for the vectors of X_i and V_i in a N-dimensional space as is represented by [59, 63].

A single objective PSO is basically programmed to find solutions for merely one function. The accuracy of each solution is examined by a fitness function. In the process of searching the solution space, if the fitness of a solution is better than x_{pbes} or x_{gbest} or even both, their value will be replaced. The process of updating is done after each iteration until the termination or if the stopping criterion is reached which

might be the predefined value for the maximum iteration is achieved or target solution is attained. Eventually, the optimal feasible solution is found by PSO [59, 63].

3-6 **Experimental database and parameters**

In this paper a database presented in recently published paper by the authors [60] was used. The database consists of UCS test results conducted on CHB cylindrical specimens after 7, 14 and 28 days curing times, here presented by UCS_7, UCS_{14} and UCS_{28}, respectively. The percentage of solid content in the mixture (SC), aggregate to tailing ratio (ATR) and cement dosage (CD) are the dominant parameters impacting the UCS of CHB at different curing times and they can be used to have an optimal mixture design. For each curing time 547 tests have been conducted where the amount of three parameters were changed, i.e. SC, ATR and CD. The waste rock and tailings used to prepare the specimens were collected from Jinfeng underground gold mine, which is located in Southwest Guizhou province of China. The details of the properties of materials and testing procedure are presented in detail by [60]. A summary of descriptive statistics of the database is given in Table 3.1 for further consideration.

Table 3.1 Statistical description of variables in the database.

Parameter		Mean	St. Dev.	Range	Min.	Max.
Input Variables	SC (%)	75.42	1.25	9	67	76
	ATR	4.88	1.09	9.99	1.46	11.45
	CD (%)	11.66	3.99	28.06	1.92	29.98
Output Variables	UCS_7 (kPa)	551.2	309.19	2438.3	26.67	2465
	UCS_{14} (kPa)	850.5	508.65	3969.7	49.67	4019
	UCS_{28} (kPa)	1260	724.44	6268.3	67.67	6336

3-7 **Problem formulation, analyses and results**

3-7-1 **Problem formulation**

The scenario of the problem considered here is to find feasible and near-optimum CHB mixture batches, i.e. amounts of design variables which are SC, ATR and CD, for a UCS_7 of about Y_1 (kPa), when the UCS_{14} should not be less than Y_2 (kPa) and/or the UCS_{28} should be more than Y_3 (kPa). Considering the main objective of this paper, each UCS is a model, which may be used either as an objective function or as a constraint. Actually, Y_j is considered as a limit of UCS here. The lower and upper bound of design variables are their minimum and maximum values in the database, which is given in Table 3.1. Note that several different other scenarios may be constructed. An example of the considered problem can be presented as follows:

Objective: To acquire ($UCS_7 \approx Y_1$)

Subject to: $67 \leq SC \leq 76$, $1.46 \leq ATR \leq 11.45$ and $1.92 \leq CD \leq 29.98$

$UCS_{14} \geq Y_2$ and $UCS_{14} \geq Y_3$

3-7-2 **Determination of UCS models by GP**

Three different GP models are developed using the key design parameters for predicting the UCS of CHB specimens with 7, 14 and 28 day curing time, respectively, UCS_7, UCS_{14} and UCS_{28}. These models will be used as objective functions or constraints in the optimisation process. The tree-shaped GP models are converted to mathematical equations to be used for hand-calculation and further aims. GP models presented here have been selected after conducting several runs. In each run, the model was developed using a set of data and then it was tested on some unseen data to examine the accuracy of the model. Finally, the optimal models are presented here. The GP-based models are as follows:

$$UCS_7 = CD\sqrt{1.15SC \times CD\sqrt{ATR}} \quad (3\text{-}3)$$

$$UCS_{14} = CD\sqrt{ATR + 5.43(SC + 2ATR)CD} \quad (3\text{-}4)$$

$$UCS_{28} = CD(0.45ATR \times CD + CD + SC) \quad (3\text{-}5)$$

In order to assess the estimation performance of the models, the error existing between the predicted and experimental output values are examined. To do so, some well-known performance indicators, i.e. the correlation coefficient (R), root mean square error (RMSE) and mean absolute error (MAE), are calculated to evaluate the model [59, 60, 64]. To examine the capability of GP, the performance of GP and linear multiple regression (LMR) models are compared. The LMR models are as follows:

$$UCS_7 = -3.74SC + 13.64ATR + 65.84CD \quad (3\text{-}6)$$

$$UCS_{14} = -7.24SC + 27.33ATR + 107.77CD \quad (3\text{-}7)$$

$$UCS_{28} = -10.92SC + 59.18ATR + 153.44CD \quad (3\text{-}8)$$

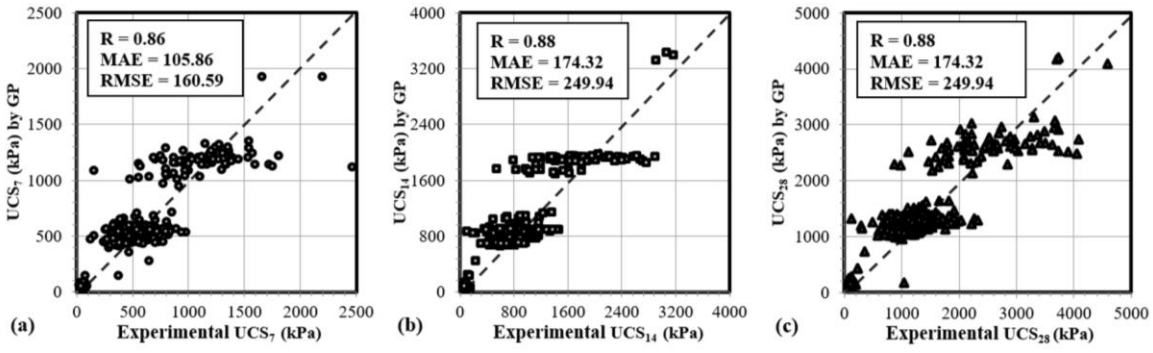


Fig. 3.2 GP versus experimental values of UCS for (a) 7, (b) 14 and (c) 28 day curing time

According to Fig. 3.2 and 3, it can be said that GP models are able to estimate UCS factor better than LMR models. It should also be mentioned that the LMR models give some negative results. GP models are also able to consider the nonlinear relationship between variables. Furthermore, GP algorithm can generate models which may be converted to equations to be used for further aims whereas those models

obtained by other AI algorithms such as ANFIS, SVMs and ANNs are usually black-box or too complex to be converted to mathematical equations [33, 64].

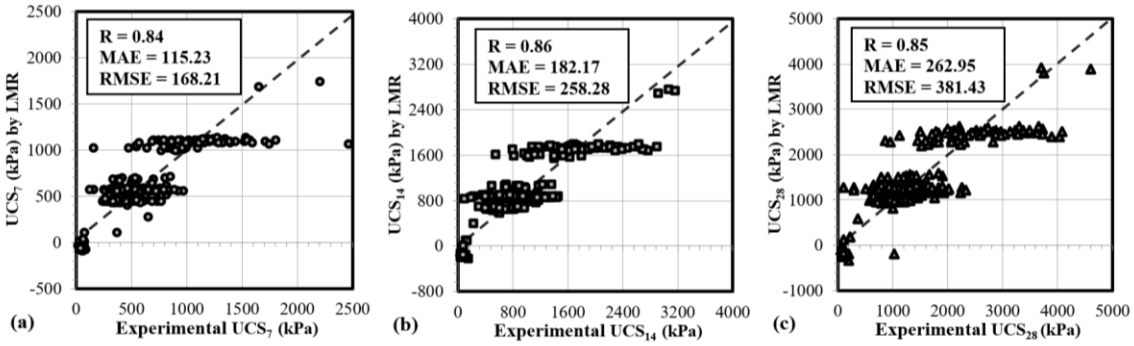


Fig. 3.3 LMR versus experimental values of UCS for (a) 7, (b) 14 and (c) 28 day curing time

3-7-3 **Finding CHB mixture design parameters for a specific UCS by PSO**

After formulating the optimisation problem and finding the objectives, the next step is to perform the process of optimisation with a PSO algorithm. In each run, the PSO algorithm provides one feasible solution, i.e. the amounts of the design parameters, which satisfy the objective and constraints. Since different solutions might be produced in each run, one way is to consider the average of their values as the final solution. Here, a number of runs were performed to obtain feasible CHB mixture design parameters for different problem formulations subject to different UCS limits Y_j. For example, in the third case it is assumed that a UCS_7 around 300 kPa is required subject to that UCS_{14} should not be less than 300 kPa and the UCS_{28} should be greater than 1000 kPa. The average value of design variables after 10 runs is considered as the final PSO solution and some results are represented in Table 3.2.

Table 3.2 Obtained CHB mixture parameters for the considered UCS conditions by PSO.

Item	UCS Limit			Proposed batch by PSO			Predicted UCS_i by GP		
	UCS_7, Y_1	UCS_{14}, Y_2	UCS_{28}, Y_3	SC (%)	ATR	CD (%)	UCS_7	UCS_{14}	UCS_{28}
1	200	300	-	71.78	3.62	6.34	200.06	330.89	560.76
2	300	-	700	73.69	3.02	8.49	300	490.1	776.95
3	500	700	1000	74.58	4.23	11.23	500	798.2	1204.1
4	0	0	2000	74.82	2.52	17.75	873	1555	2000
5	2000	0	0	70.16	8.73	25.54	2000	2824	5020

As can be seen from the presented cases #1-3 the amount of cement and tailings change significantly depending on the required strength and time. As shown by the last two extreme cases, when very high strength is needed in a week (#4) quite high cement (26%) and less tailings (ATR=8.7) should be used, whereas to get the same strength in 28 days (#5) significantly less cement and more tailings can be used. When the proposed batches are compared with the experimental data they are quite close. It should be noted that in optimisation problems, the accuracy of solutions is related to several factors such as the

accuracy of the objective functions and the type of constraints and these can be finally solved by considering a good scenario for solving the problem. Of course, it must always be remembered that the engineer and decision-makers are ultimately responsible for finding and selecting the proper plan and strategy.

3-8 **Conclusion**

This paper aimed to investigate the capability of AI-based prediction and optimisation to find CHB mixture batches yielding a certain UCS at a specified time as are required by a mine planning engineer. It is demonstrated that the optimisation problem should be properly formulated so as to find appropriate solutions. GP is used to produce nonlinear models using an extensive database containing SC, ATR, CD, time and corresponding UCS. The developed GP models are used as objectives and/or constraints to find optimum CHB mixture variables for considered UCS conditions by a PSO algorithm. Several examples, showing the practical application of the presented approach, are provided. The presented procedure can provide the engineer with a better idea of an appropriate mixture design for a desired UCS in order to avoid several trial tests. The optimisation procedure presented will result in reducing the costs of CHB, by using optimum quantities of costly materials during the project.

Chapter 4 Multi-objective mixture design of cemented paste backfill using particle swarm optimisation algorithm

4-1 Abstract

In order to achieve a successful cemented paste backfill (CPB) mixture design, multiple project requirements such as strength, flowability and cost should be met. For this achievement, the key design parameters, solid content (SD) and cement percentage (C), should be well adjusted. With increasing the amount of cement in the mixture, CPB strength and production cost increase together, whereas the workability decreases. In order to reduce the cost, more tailings can be added while keeping the cement amount the same but this will reduce both strength and workability. Therefore, CPB design is in fact a multi-objective optimisation problem. In this study, the particle swarm optimisation (PSO) algorithm is used to design CPB mixture meeting multiple objectives. PSO identifies the optimum set of SD and C yielding in desired strength and workability with a minimum cost. The proposed workflow can be a useful and practical for multiple decision making where CPB designers face strength-workability-cost paradox. In addition to reducing the number of trial experiments, the multi objective mixture design of CPB also provides the optimum use of materials to reduce the incurred costs and ensure cleaner and more sustainable production.

Keywords: Mining Engineering; Tailings management; Cemented paste backfill; Multi-objective design; Particle swarm optimisation.

4-2 Introduction

Mining is one of the main solid waste generating industries. Tailings are fine-grained solid waste materials left over after extracting valuable minerals from the mined ore. The chemical and physical properties of tailings change with the nature of the ore and the processing method. These waste materials must be managed considering their environmental impacts and to have a more sustainable mining industry which is a challenge in present projects. Generally, surface tailing dams are used to store mine tailings. Recent tailing dam failures around the world lead to the loss of human lives, devastation of agricultural and forestry lands, and financial losses for the mining companies. These events triggered serious concerns about tailings management. To handle these concerns, various management approaches have been suggested, including the use of tailing as backfill materials (e.g. in Falconbridge Nickel mines Ltd. at the Hardy mine in Sudbury mine [65]. In the late 1970s, cemented paste backfill (CPB) started to be used in underground mines in Germany for the reduction of solid waste on the surface by returning up to 60% of the extracted material underground to be used as a main support element [66]. Since then, CPB has been considered as a cleaner and more sustainable tailings management strategy in which tailings are placed back into underground mines in many leading countries in mining such as Canada and Australia [62].

The main factors controlling the design of CPB are strength, flowability and cost. The key controlling parameters, strength and flowability, are affected by intrinsic and extrinsic factors [67]. The intrinsic factors are mainly related with the physical, chemical and mineralogical properties of the main components of CPB such as tailings, cement and mixing water. Whereas extrinsic factors are related with the conditions, occurring around an underground stope filled with CPB. Most of these factors are mine specific and uncontrollable that is why the design of CPB is unique for each mine. In practice, the controllable key design parameters such as cement percentage (C), solid content (SD) are adjusted to achieve the desired timely strength, flowability and the cost. In the following sections the dominant intrinsic factors effecting the strength, cost, flowability of CPB and the key design parameters are explained in detail.

4-2-1 **Cemented paste backfill (CPB)**

CPB is a mixture of fine tailings (about 75 to 85% solids by total weight), cement as a binder (typically between 3 and 9% by total solid weight) and water for an approximate solid content of 70 to 80% by weight [67, 68]. The underground voids or mined out stopes can be filled with CPB transferred using pipelines.

The design of cemented paste backfill (CPB) is important in underground mines that use backfill-based production methods such as cut and fill, sublevel stoping. CPB design is specific to each mine due to varying tailings characteristics [67]. Therefore, there is a significant variation in CPB design from mine to mine; the proposed solutions should take into account three main factors, namely the uniaxial compressive strength, workability and cost.

4-2-1-1 Strength

In order to preserve the stope stability in an underground mine, CPB must remain stable during the extraction of ore from adjacent stopes. When CPB is not strong enough, the adjacent faces may relax and collapse into the open stope. The strength of CPB is mostly measured by its uniaxial compressive strength (UCS). In designing CPB, both short term and long term strength are important. Short term strength is important to prevent liquefaction and reduce delays in production. Long term strength reduces the occurrence of ground subsidence in mining operations, safeguards the operation of secondary stopes in the mine and ensures the long term stability of the mine. If the CPB could not gain enough strength within the planned time then it would not be possible to mine out the nearby stopes. The possible consequences of weak CPB can be mine dilution, the collapse of stopes, and large scale hangingwall or footwall failure, threatening the overall mine stability. The minimum strength required at the early stage is 0.15 MPa to prevent liquefaction [69, 70]. For the mines where the mining method is cut and fill or sublevel stoping the suggested strength is 0.7 MPa [71, 72]. For roof support CPB should exceed 4 MPa [73]. Strength gaining time is very important from a mine planning perspective to reduce waiting time for the extraction of the adjacent stope, thereby reducing delays in production.

4-2-1-2 Flowability

Generally, CPB is produced at surface batching plants and transported to underground stopes through pipelines. Therefore, freshly mixed CPB should be “flowable” enough for efficient pumping or delivery. The flowability or workability of CPB is an important parameter as strength in paste design. In order to facilitate the transportation of CPB and prevent pipeline clogging, the workability of CPB is mainly controlled by the rheology of the mixture. Moreover, CPB with low flowability requires high pumping pressure to transfer it to the underground stopes, which increases the costs of pipe/pump maintenance and energy. The flowability or workability of CPB is mostly characterised by a slump test or rheometer. The slump is defined as the difference between the heights of the slump cone and the collapsed material after removal from the cone. When traditional slump cone with 12” height with 8” base and 4” top radius [74] is used, the necessary slump value for CPB ranges from 6” to 10” [67]. The corresponding yield stress can be estimated using the equation suggested by Murata [75] and modified by Christensen [76]. As an alternative to the traditional slump cone, a hollow cylindrical can be used for the same purpose [77]. With the cylinder, it is possible to estimate yield stress, the pressure required to overcome the static friction of the fluid materials, based on the measured slump value. The yield stress should range from 250 to 800 Pa for a workable CPB [78]. In case of centrifugal pumps for pipeline transport it has been reported that yield stress should be less than 200 Pa [79].

4-2-1-3 Cost

The cost of CPB typically represents almost 20% of all mining costs, with binder cost reaching up to 75 % of that amount [78]. According to Naylor et al. [80] as a general rule, the cost of CPB mix with 1% binder is 1$/ton. Other researchers indicated that 42% of the total CPB operational cost is the cost of CPB with 3% cement (in weight) [81]. 50-75 % of the total CPB operational cost is the cost of CPB mix with 5-9 % binder [82]. So far, many attempts have been made to mitigate the cement cost by using chemical agents, alternative binders, cement replacement materials and additives. Such high cement cost puts greater pressure on mine economy. Hence, it can be stated that a 1% reduction of cement content may reduce the overall mining cost by 1.6 to 2%. Therefore, any reduction in the use of excess cement will result in obvious economic benefits.

4-2-2 **CPB design parameters**

All of the listed CPB properties are mainly affected by the same key design parameters such as cement percentage, solid content, type of tailing, curing time [67, 83-87]. Therefore, these key parameters should be included in any model constructed for the prediction of strength, yield stress and cost.

Cement percentage (C): It is the key parameter controlling strength, yield stress and most importantly the cost of CPB. More cement helps improve the strength but also increases the cost. Yield stress is also affected by cement percentage to some extent. The cement content used generally ranges from 4 to 9 % (by weight) and is mainly influenced by the magnitude of the stress expected around the exploited stope, scheduled production time and the cost of paste production i.e. higher binder content yields higher strength but also increases the cost.

Solid content (SD): Yield stress, the key parameter controlling the transportation of CPB, is strongly related to the solid content. To enhance the flowability, fluidity and workability of CPB the water content is increased, thereby decreasing the solid content. For the same amount of cement increasing SD improves both short and long term strength. Moreover, it is well known that the mechanical strength of cementitious materials such as CPB can be improved by reducing the total amount of water. In addition to mechanical improvement high SD is also desirable as it uses more tailings, thus reducing the amount of tailings which needs to be deposited in surface tailing dams [88].

Type of tailings: Density, particle size and chemical composition of the tailings are all important parameters controlling the mechanical properties of CPB such as strength and yield stress. As CPB is designed for a specific type of tailings it is not possible to use the same design parameters for other mines or tailings. Therefore, each type of tailings should be considered separately. As only one type of tailings is used in this study, the tailings related parameters are considered as constant rather than separate parameters.

Time (T): Like other cementitious materials, the strength of CPB is a function of time. From the production planning perspective, time is the most important parameter as it has a direct effect on the proposed mining sequence. For example, sufficient strength in a shorter time will allow the adjacent stopes to be scheduled earlier.

As outlined above, successful design of CPB is a complex optimisation problem. Different optimisation techniques are used in mining and geotechnical engineering applications. As most of the optimisation problems do not perfectly fit exact algorithm, heuristic algorithms such as the genetic algorithm (GA), particle swarm optimisation (PSO), ant colony optimisation (ACO), and simulated annealing (SA) have been used in mining engineering applications such as mine planning, underground stope optimisation, and open pit blast design [89-91]. Among others PSO algorithm has been used recently to solve complex and multidimensional global optimisation problems in mining and geotechnical engineering applications [92-94]. In this paper, a hybrid algorithm combining multiple regression modelling and particle swarm optimisation is presented. A comprehensive experimental program covering slump and

UCS tests was conducted on CPB samples prepared considering C and SD suggested by the mine management. The constructed database is used to generate strength, yield stress, cost predictive models and implement the optimisation algorithm. Proper objective functions were constructed using conventional multiple regression modelling. The constructed models were optimised considering the desired constraints and conditions. Here, a popular global optimiser algorithm, namely multi objective particle swarm Optimisation (MOPSO), is used to find the optimal solutions. MOPSO has also been utilized for solving many geotechnical engineering optimisation problems [93]. The results prove that the approach can be used as a tool for the determination of best key design parameters leading to the aimed strength in specified time and desired yield stress with the possible minimum cost.

4-3 **Experimental studies**

4-3-1 **Material**

For the sample preparation, mine tailings were collected from a hard rock underground gold mine in Western Australia. Chemical composition analysis was conducted on tailings samples (Table 4.1). Particle size distribution of tailing sample was determined using a model 2000 Mastersizer by Malvern Instruments. The tailings' specific gravity and specific surface area are 2.8 and 0.598 m^2/gr, respectively. As can be seen in Fig. 4.1, the tailings contain more than 15% Wt.% of fine particles (<20 μm), so the tailings can be used for CPB production as it will retain sufficient water to form paste.

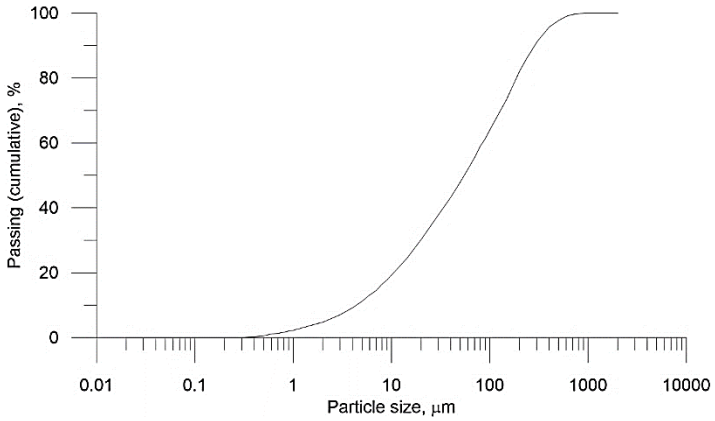


Fig. 4.1 Particle size distribution of the tailing.

Table 4.1 Chemical composition of the tailings used

Element	**%**
Al_2O_3	9.55
CaO	3.16
Fe_2O_3	3.18
K_2O	1.34
MgO	1.56
MnO	0.08
Na_2O	3.31
P_2O_5	0.05
SiO_2	68.49
SO_3	1.13
TiO_2	0.54
Hydration modulus $(Al_2O_3+MgO+CaO)/SiO_2$	0.21
Basicity coefficient $(MgO+CaO)/(Al_2O_3+SiO_2)$	0.06

Scanning electron microscope (SEM) analysis was conducted for the identification of the shape of individual tailing particles (Fig. 4.2).

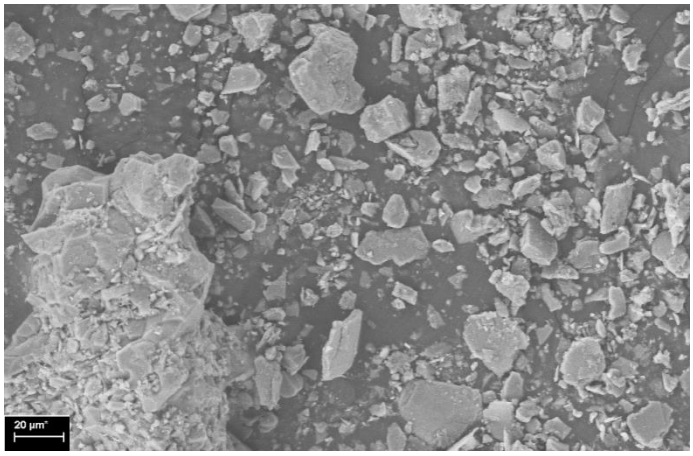


Fig. 4.2 Scanning electron micrograph of the tailing.

The tailings mineralogical composition is determined by X-ray diffraction analysis (XRD), which is a powerful non-destructive technique for characterizing crystalline mineral composition of materials. The diffraction pattern in the 2θ range of 5–90° were recorded using PANalytical Empyrean X-ray diffractometer with C_u K radiation (λ = 1.5404 Å), 40 kV, and 40 mA at a rate of 2°/min. According to the peak intensity of minerals shown in Fig. 4.3, the major mineral identified in tailings is Quartz, which is composed of SiO_2. Other minerals such as Biotite, Gypsum, Albite are identified, but their relative proportion is not significant. These results are consistent with the chemical composition results in Table

4.1 representing that SiO_2 was the major chemical in the tailings. Quartz may increase the early age hydration process and also the compressive strength of CPB and this can be considered by the designer to have a more successful CPB mix design.

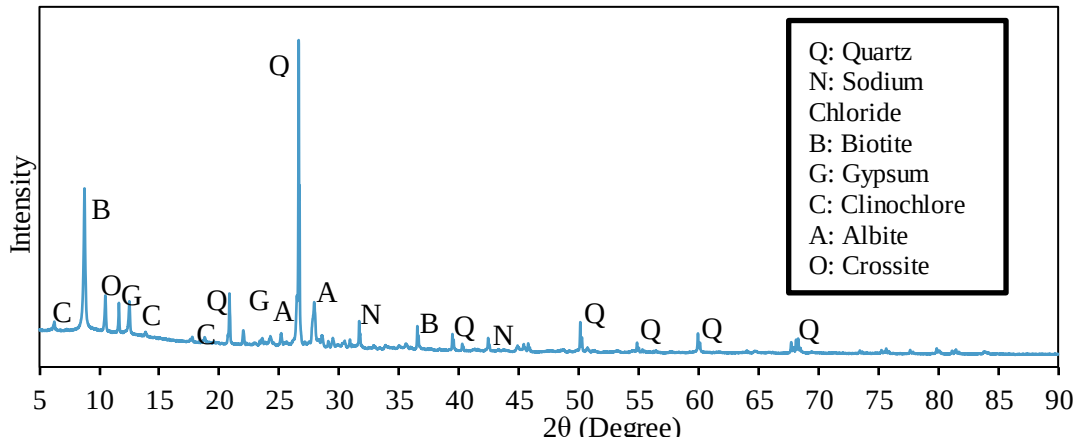


Fig. 4.3 XRD profiles of the tailings used in this study.

In this study, general purpose Ordinary Portland Cement (OPC) was used as the binder material. The mechanical and chemical properties of the used OPC are given in Table 4.2.

Table 4.2 Mechanical and chemical properties of OPC.

Chemical properties		**Physical properties**		**Mechanical properties**	
CaO, %	63.4	Specific gravity, t/m^3	3.0-3.2	3 day UCS, MPa	38.6
SiO_2, %	20.1	Fineness index, m^3/kg	390	7 days UCS, MPa	48.4
Al_2O_3, %	4.6	Normal consistency, %	27	28 days UCS, MPa	58.5
Fe_2O_3, %	2.8	Setting time initial, min	120	Shrinkage 28 days,	640
SO_3, %	2.7	Setting time final, min	210	Compound composition	
MgO, %	1.3	Soundness, mm	2	C_3S	
Na_2O, %	0.6	Loss on ignition, %	3.8	C_2S	
Chloride, %	0.02	Residue 45 mm sieve, %	4.7	C_3A	
				C_4AF	

Samples were prepared by varying the key design parameters, cement content (C) and solid content (SC), to determine yield stress and 7, 28 and 56 days uniaxial compressive strengths. 5-9% OPC was used for sample production. The solid content of CPB samples ranges from 76 % to 79 %. All the laboratory tests were conducted on 16 batches as shown in Table 4.3.

Cement percentage (C), solid content (SD) and water contents (W) are determined using Equations (4-1), (4-2), and (4-3).

$$C = \frac{Mc}{(Mc + Mt)} * 100 \quad (4\text{-}1)$$

$$SC = \frac{(Mt + Mc)}{(Mc + Mt + Mw)} * 100 \quad (4\text{-}2)$$

$$W = \frac{Mw}{(Mc + Mt)} * 100 \quad (4\text{-}3)$$

where, C is the cement percent (%), SC is the solid content (%), W is water content (%), M_t is dry tailing mass (kg), M_c is dry cement mass (kg), M_w is water mass (kg).

Table 4.3 Prepared CPB batches.

Batch No	C, %	SD, %	Batch No	C, %	SD, %
1	5	76.0	9	7	76.0
2		77.0	10		77.0
3		78.4	11		78.4
4		79.0	12		79.0
5	6	76.0	13	9	76.0
6		77.0	14		77.0
7		78.4	15		78.4
8		79.0	16		79.0

The CPB mixtures were thoroughly mixed using a bench top high speed mixer for one minute, then the side of the bowl were cleared and the batch was mixed for two more minutes. Following mixing, slump tests were conducted and samples were moulded in cubic moulds for strength experiments. Pashias [77] suggested the use of a cylinder and established an equation to relate the slump to yield stress. Recent studies demonstrated that the yield stress from the rheometer tests and cylindrical slump test were in agreement for engineering purposes [95].

4-3-2 **Laboratory experiments**

4-3-2-1 Slump test

The Slump test was conducted on CPB samples using a hollow cylinder with 110 mm (diameter) x 110 mm (height). The test method was suggested by [77]; it was used on the 16 batches presented in Table 4.3. Freshly mixed paste was poured into a cylinder to overfill it. Then the top of the cylinder was flattened using a spatula. The cylindrical mould was lifted vertically and slowly at a constant speed. Just like a conventional slump cone, the top of the sample can deviate, in such cases the middle point of the top surface was taken as reference point for slump measurement. Slump value was measured to the nearest mm with a ruler (see Fig. 4.4).

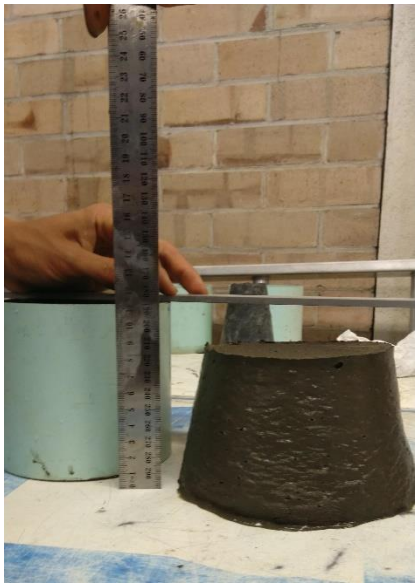


Fig. 4.4 Slump test

Recent studies proved that the alternative method using a cylinder can be successfully applied to CPB samples [95]. The used slump test method requires less material and provides fast results. Moreover, Pashias et al. [77] generated analytical equations for calculating yield stress.

$$\tau_y' = \frac{1}{2} - \frac{1}{2}\sqrt{s'} \tag{4-4}$$

where;

$$s' = \frac{s}{H} \tag{4-5}$$

$$\tau_y' = \frac{\tau}{\rho g} \tag{4-6}$$

where τ_y' is dimensionless yield stress, s' is dimensionless slump height proportion, s is the slump height (mm), H is the initial height (mm) stress, τ is the yield stress (Pa), ρ is paste density (kg/m^3), and g is gravitational constant. The measured slump values and the calculated yield stress values are presented in Table 4.4.

Table 4.4 Measured slump height, density and calculated yield stress values.

Batch No	H, mm	s, mm	Yield stress, Pa
1		84	163
2		64	306
3		28	640
4		28	640
5		76	223
6		69	271
7	110	48	442
8		39	527
9		78	204
10		64	307
11		44	475
12		31	606
13		72	244
14		69	266
15		52	400
16		40	513

4-3-2-2 UCS tests

Over 100 UCS tests were conducted using 5x5x5 cm cubic samples as suggested by ASTM [96] as a function of curing time (7, 28 and 56 days). Each test was repeated at least twice and the average values are presented in Table 4.5. The cubes were allowed to cure in a curing room at a relative humidity of 80%.

At predetermined curing periods of 7, 28 and 56 days, the samples were loaded axially to determine uniaxial compressive strength. Compressive strength tests were conducted using computer controlled mechanical loading press. For each mix, the average compressive strength values are shown in Table 4.5.

4-3-2-3 Cost prediction

As indicated previously, the main cost item for CPB is the cost of cement. By knowing the values of SC and C for each batch, the cement cost per wet cubic metre of CPB was calculated and presented in Table 4.5. The cement cost is assumed as 300 AUD/tonne.

Table 4.5 UCS values and calculated cement cost per wet m^3 of CPB.

Batch No	UCS, kPa			Cement cost,
	7 day	28 day	56 days	AUD /m^3
1	307	496	395	11.40
2	384	501	395	11.55
3	400	347	295	11.76
4	521	610	357	11.85
5	502	646	473	13.68
6	434	565	410	13.86
7	654	834	706	14.11
8	717	928	721	14.22
9	442	590	632	15.96
10	568	767	739	16.18
11	716	912	975	16.47
12	803	1098	995	16.59
13	689	932	979	20.53
14	615	835	850	20.79
15	959	1290	1362	21.18
16	904	1265	1335	21.34

4-4 **Modelling and optimisation**

4-4-1 **Multiple regression modelling**

After collecting data, the next stage of the optimisation procedure is the construction of predictive models for the strength, yield stress and cost. The constructed models will be used as objective functions for the optimisation. In this study, a simple multiple regression modelling technique is used for this purpose. Multiple regression models were developed for the prediction of UCS, yield stress and cost based on key design parameters C and SD. In this study, several linear and nonlinear regression models are developed to describe the objectives mathematically and the best performing models are selected and presented in Table 4.6. The regression and predictive models used in the proposed optimisation process are calibrated and developed based on the results of the laboratory experiments conducted on the samples prepared using the aforementioned materials. It should be noted that, for the cases in which the intrinsic factors affecting the strength, flowability and cost change, the models should be reconstructed and/or updated accordingly.

Table 4.6 Constructed multiple regression models for the prediction of UCS, yield stress and cost.

Predictive model	Adjusted R^2
Cost=-15.74+2.33C+0.203SD	99.97
Yield stress=-9138.48-15.94C+124.17SD	86.81
UCS=EXP(0.185C+0.153SD+0.00388T-6.7415)	79.91

4-4-2 **Multi objective PSO**

Particle swarm optimisation (PSO) is a population-based metaheuristic algorithm proposed by Kennedy an Eberhart [97, 98]. PSO is inspired by the social behaviour of animals e.g. birds or fish seeking food [97]. Each particle (member) of a swarm (population size) updates its search pattern in line with its own experience and the success of the others. Assume that $\mathbf{x}_{i(t)}$ represents the position of particle p_i, at iteration t. The position of p_i is updated by adding the effect of velocity $\mathbf{v}_i(t)$ to the previous position as follows:

$$\boldsymbol{x_{i(t)} = x_{i(t-1)} + v_i(t)} \tag{4-7}$$

The velocity vector reflects the exchanged information and, in general, is defined in the following way:

$$\boldsymbol{v_i(t) = W v_i(t-1) + C_1 r_1 \left(x_{pbesti} - x_{i(t)}\right) + C_2 r_2 \left(x_{leader} - x_{i(t)}\right)} \tag{4-8}$$

where C_1 and C_2 are learning factors an defined as constants, and r_1, $r_2 \in [0, 1]$ are randomly generated values, W is the inertial weight defined within the algorithm. These definitions and indicators have also been extended for the vectors of X_i and V_i in a N-dimensional space as is represented by [63].

PSO is a single function Optimisation algorithm which is basically developed to find optimal solutions for merely one function. A multi-objective PSO (MOPSO), proposed by [99] can be utilized when the problem requires more than one objective function to be optimised simultaneously. Accordingly, it is not possible to find one optimal solution for all the functions and thereby, a set of optimal solutions, named Pareto, can be obtained [63, 100]. It should be considered that there are infinite number of Pareto optimal solutions mathematically and perhaps one is the preferable solution. The best solution can be selected based on the requirements and an expert knowledge about the problem.

As represented in Fig. 4.5, the process of finding solutions in MOPSO starts with the initialization of swarms. Then, a group of swarms called leaders is created with the non-dominated members or particles. Leaders are used to update the positions of the particles. The group of leaders is usually kept in an external archive to be used as leaders when the positions of the particles of the swarm are required to be updated. These leaders are compared in order to choose one leader for each particle of the swarm. After the flight of particles, each of them is assessed and its p_{best} is updated. When the particle is dominated, a new particle replaces its best position, which is x_{pbesti}. After updating all of the positions, the group of leaders is also updated. In the algorithm, swarms of particles move to find the optimal position and the qualities of groups are measured and this process continues to the last pre-defined number of iterations [97, 100]. It should be noted that the set of optimal solutions varies by changing the number of iterations and the quality of assessment of swarms. This may change in each run. Therefore, some runs should be made to achieve the optimal solution by the engineer or designer. Similar to other optimisation algorithms, the disadvantages of PSO are the need for proper setting the

algorithm-dependent parameters and a large number of iterations. Additionally, PSO may fall into local optima in high-dimensional space. However, the number of variables in this study is quite limited. As a multi-objective optimisation algorithm, the main strength of PSO over other algorithms is its fast convergence. This algorithm can be programmed to facilitate its use. Here, MATLAB (2018b) was used to run the written PSO algorithm and the mathematical optimisation procedure.

```
Begin
  Initialize swarm
  Initialize leaders in an external archive Quality (leaders)
g = 0
While g  < gmax
        For each particle
            Select leader
            Update Position (Flight)
            Mutation
            Evaluation
            Update pbest
       EndFor
       Update leaders in the external archive
       Quality(leaders)
       g++
    EndWhile
Report results in the external archive
End
```

Fig. 4.5 A general Pseudocode for MOPSO algorithm [97].

4-5 **Application example**

For this practical application, two hypothetical cases were considered. In the first case, it was assumed that mine production engineer is looking for best design parameters C and SD, to produce CPB having UCS ranging from 750-800 kPa in 28 days, yield stress between 500-800 Pa and maximum cement cost of 17 AUD/m^3. The presented hybrid approach produced different batches, combination of key design parameters, C and SD (Table 4.7). All of the suggested batches satisfy the desired strength, yield stress and cost. The first batch (C1_1) provides the maximum strength and minimum yield stress, whereas the ninth batch (C1_9) provides minimum cost with strength and yield stress values in desired ranges.

Table 4.7 The suggested key design parameters for the hypothetical case #1.

Batch No	C, %	SD, %	28 days UCS, kPa	Yield stress, Pa	Cement cost, AUD/m^3
C1_1	7.01	78.57	799.54	505.18	16.54
C1_2	6.70	78.91	796.16	552.68	15.89
C1_3	6.83	78.68	788.29	522.56	16.16
C1_4	6.53	78.99	780.64	565.18	15.51
C1_5	6.69	78.78	778.49	536.78	15.831
C1_6	6.81	78.62	777.83	514.97	16.10
C1_7	6.46	78.99	771.92	566.95	15.36
C1_8	6.71	78.64	765.55	519.64	15.86
C1_9	6.32	79.00	751.83	570.22	15.01

For the second case, it was assumed that to increase production rate, high early strength and lower yield stress are the main concerns. The production engineer is looking for the same UCS (750-800 kPa) in 14 days and yield stress ranging from 400 to 500 Pa. In this case, the algorithm yields different batches with different C and SD (Table 4.8). Just like the former case, all the batches satisfy the desired UCS and yield stress values. The fourth batch (C2_4) provides mid UCS and yield stress with minimum cost, whereas the seventh batch (C2_7) yields in maximum UCS in the desired range.

Table 4.8 The suggested key design parameters for the hypothetical case #2.

Batch No	C, %	SD, %	28 days UCS, kPa	Yield stress, Pa	Cement cost, AUD/m^3
C2_1	7.43	78.03	754.44	431.48	17.42
C2_2	7.48	78.25	788.11	459.02	17.58
C2_3	7.66	77.92	774.61	414.51	17.94
C2_4	7.34	78.21	763.66	456.17	17.25
C2_5	7.43	78.03	754.93	432.07	17.42
C2_6	7.73	77.87	778.15	406.88	18.08
C2_7	7.84	77.90	798.60	409.84	18.35

As can be seen from the presented case, the proposed algorithm provides production engineers with different alternatives satisfying the desired values, then the design engineer chooses the best one considering current conditions and priorities. It should be noted that it is also possible to reduce the number of alternative batches or get a unique batch by narrowing the ranges for the desired outputs.

4-6 **Conclusion**

In practise, CPB is designed by using predictive models for indirect assessment of the key design parameters, or classical trial-and-error method requiring extensive laboratory studies time and thereby cost.

In this paper, a hybrid algorithm combining conventional regression modelling and particle swarm optimisation techniques is presented. An experimental program was completed for the construction of a database to generate predictive regression equations. The constructed models were used as objective functions in the optimisation algorithm designed for maximising strength, minimising yield stress and cost. For the desired strength, yield stress and cement cost values, the algorithm suggests a range of design parameters such as cement percentage and solid content providing specified strength, yield stress and cost. Such range provides an opportunity for the designer to select the best design parameters for the case under consideration. If the main concern is strength, then the designer can choose the design parameters yielding the highest strength with acceptable yield stress and possible minimum cost. For the cases where the main concern is the cost then the designer choses the relevant design parameters giving the lowest cost with acceptable strength and yield stress.

Using the presented approach, it is possible to find out the best combination of design parameters for the desired outputs (i.e. strength, yield stress, and cost), whichever is important to the production engineer.

The presented algorithm can provide instant batch designs, in terms of key design parameters C and SD, for each particular stope without a time consuming trial and error procedure. Acknowledging that each mine has specific requirements, the algorithm can be modified easily, even improved by adding mine specific variables or conditions.

Chapter 5 Multi-objective mixture design and optimisation of steel fiber reinforced UHPC using machine learning algorithms and metaheuristics

5-1 Abstract

Ultra-high-performance concrete (UHPC) is a recent class of concrete with improved rheological, mechanical and durability properties compared to normal concrete. The production cost of UHPC is considerably high due to the large amount of cement used, and the high price of other required constituents such as quartz powder, silica fume, fibres and superplasticisers. In order to achieve specific requirements such as desired production cost, strength and flowability, the proportions of UHPC's constituents must be well adjusted. The traditional mixture design of concrete requires cumbersome, costly and extensive experimental program. Therefore, mathematical optimisation, design of experiments (DOE) and statistical mixture design (SMD) methods have been used in recent years, particularly for meeting multiple objectives. In traditional methods, simple regression models such as multiple linear regression models are used as objective functions according to the requirements. Once the model is constructed, mathematical programming and simplex algorithms are usually used to find optimal solutions. However, a more flexible procedure enabling the use of high accuracy nonlinear models and defining different scenarios for multi-objective mixture design is required, particularly when it comes to data which are not well structured to fit simple regression models such as multiple linear regression. This paper aims to demonstrate a procedure integrating machine learning (ML) algorithms such as Artificial Neural Networks (ANNs) and Gaussian Process Regression (GPR) to develop high-accuracy models, and a metaheuristic optimisation algorithm called Particle Swarm Optimisation (PSO) algorithm for multi-objective mixture design and optimisation of UHPC reinforced with steel fibers. A reliable experimental dataset is used to develop the models and to justify the final results. The comparison of the obtained results with the experimental results validates the capability of the proposed procedure for multi-objective mixture design and optimisation of steel fiber reinforced UHPC. The proposed procedure not only reduces the efforts in experimental design of UHPC but also leads to the optimal mixtures when the designer faces strength-flowability-cost paradoxes.

Keywords Ultra-high-performance concrete; Mixture design; Multi-objective optimisation; Machine learning algorithms; Particle swarm optimisation.

5-2 Introduction

Ultra-high-performance concrete (UHPC) is a recent class of concrete presenting much better mechanical, rheological and durability properties than traditional concretes [101, 102]. UHPC can provide very high compressive strength i.e. over 150 (MPa), which can be used to construct high-capacity structures and to reduce the cross-sections of structural elements. Such improved properties certainly lead to more flexibility in architectural design, and lower installation and labour cost [20, 103].

The range and type of materials commonly used to produce UHPC are 27 to 40% cement, 6-12% silica fume, 7-14% quartz powder or flour, 35 to 45% sand, 4-10% water, 0.5 to 3% superplasticizer and also 0-8% steel fibers where required [104, 105]. Each material and proportion proposed have a distinctive influence on the properties of UHPC. Cement in UHPC plays the role of main binder considering its high hydraulic reactivity. Silica fume is a reactive pozzolan and an ultrafine powder used to fill the voids and react with free water in cement matrix [103]. Quartz flour or powder in UHPC reduces the initial porosity of the mixture and thereby improves the final strength. Obviously, superplasticizers are used to improve the workability and flowability of UHPC. Considering the characteristics of steel fibers, this material is utilized in UHPC in order to improve its ductility, tensile and flexural strength. It also increases the compressive strength through increasing the cohesion between the UHPC matrix components. However, excessive use of steel fibers yield in interlock with each other which in turn reduces the workability and also strength performance steel fiber reinforced UHPC [106]. Similarly, excessive use of sand or other components results in deteriorating key properties such as strength or flowability. The production cost of UHPC is considerably high due to the large amount of cement used and the high price of other required materials.

In order to achieve specific requirements such as desired production cost, strength and flowability, proportions of UHPC's constituents must be well adjusted. Conventionally, the mixture design of concrete requires comprehensive trial and error experiments, which are cumbersome and costly. The requirements and objectives of a project may contradict making the experimental design much more complicated and costly. For example, increasing the amount of cement to some extent increases the strength, while it increases the cost and decreases slump. Moreover, such conventional methods may not yield in the best results especially when multiple objectives are aimed. Therefore, mathematical optimisation, design of experiment (DOE) and statistical mixture design (SMD) methods have been recommended by researchers [20, 21, 107]. In these methods, the properties of concrete are objectives to be optimised to find the optimal amounts of the mixture constituents.

DOE and SMD methods commonly use response surface methodology (RSM). Traditional RSM requires the calibration of a polynomial regression model such as linear regression model, first-order linear regression with interaction terms, or second-order quadratic with or without interaction terms, on produced experimental data [108]. Experimental design applies to a specific set of pre-planned tests in which different level combinations of the mixture design treatments (variables or factors) are studied. In terms of concrete mixture design, different mixture design variables such as the amount of cement, water to cement ratio, or proportions of constituents are mixed and specimens are produced. Based on different conditions and requirements, these specimens are tested to study the properties of concrete such as strength, durability and flowability tests. In order to use RSM, it is essential to choose an experimental design to calibrate and evaluate the accuracy of the fitted model. In this regard, various design of experiment (DOE) plans can be utilized which can be also be used for SMD purposes such as k-factorial, central composite design, D-optimal method and so forth [109-111].

Although regression analyses and the models used in RSM or classical SMD methods mainly depend on statistical significance levels, influential terms might not be included in the model [108]. There might be nonlinear and complex relationship between the variables in dataset. Furthermore, in the polynomial model regression model, the number of terms is constrained to the number of experimental design results, and the selection of the proper polynomial equation can be considerably challenging due to the fact that each response needs its own polynomial equation [108]. This also leads to a complicated SMD process. Machine learning (ML) and artificial intelligence (AI) algorithms offer an alternative to the traditional regression method as a modeling tool. These data-driven algorithms are developed and inspired form the biological system of creatures and behavior of natural systems or phenomena. There exist several ML algorithms used for modeling and regression such as artificial neural networks (ANNs), genetic programming (GP), Gaussian regression process (GPR), support vector machines (SVMs) and so forth each of which may include different variants such as radial basis function (RBF) and multilayer perceptron (MLP) neural networks (NNs) based on their structure and parameters. In terms of modeling and regression analyses, ML algorithms are more useful when there is no sufficient information about the structure of the model or relationships between input and output variables [31, 100, 112-114]. For instance, ANNs are the most popular AI and ML algorithms used for different regression or classification purposes. ANNs are inspired from the structure of the neural system of an animal or a human brain. It was represented that ANNs are more powerful compared to RSMs in terms of modeling of complex relationships, particularly non-linear ones which cannot be developed using fitting nonlinear models and without complicated equations [30, 108]. It is also demonstrated that these models might have better predictive performance than regression models [113, 114].

Generally, optimisation means enhancing the performance of a process, a system or a product in order to take benefit from it. From the mathematical viewpoint, optimisation aims to find the best values of independent or input variables when some objectives, i.e. mathematical models, are met. There are nature-inspired algorithms or subset of AI algorithms called metaheuristics such as particle swarm optimisation (PSO), evolutionary algorithms (EAs) or ant colony optimisation (ACO) algorithms used to find all approximate solutions from a space of solutions to a problem [115-117]. In terms of design and mathematical optimisation of concrete mixtures, regression models such as RSMs are commonly used as objective functions. Ghafari and his colleagues [21] prepared different specimens of steel fiber reinforced UHPC with different proportions of constituents pre-designed based on D-optimal experimental design in order to produce data for modeling. A D-optimal design is generated by an iterative search algorithm and seeks to minimise the covariance of the parameter estimates for a specified model. This is equivalent to maximising the determinant $D = |X^T X|$, where X is the design matrix of model terms evaluated at specific run in the design space. The model is commonly a polynomial regression model with or without interaction terms. In D-optimal design, applicable terms such as X, X_1X_2 or X^2 for one or more variables may be chosen depending the problem. Therefore, unlike traditional DOE designs, D-optimal designs do not necessarily require all terms in polynomial

model. However, the selection of the model itself is challenging and is dependent on a great knowledge about the influence of material on the output. One may choose the linear regression model while the other may choose a quadratic model for a specific problem. Ghafari and his colleagues [21] considered compressive strength at two conditions and slump as objective functions. After doing relevant experiments and compiling a database, they applied classical SMD methods for developing the models and used them as objective functions for mathematical optimisation [21]. They used the linear regression model with interaction terms, so-called canonical model, to develop objective functions. They chose some of the interaction terms to investigate the response surfaces and doing experiments. Mathematical programming and simplex algorithms are usually used to find optimal solutions when RSM or polynomial models are used. They also developed a model using ANNs and proposed that the model found by ANNs is of higher accuracy than classical regression methods [21]. Although they developed ANN-based predictive models, they could not use them in a mathematical optimisation procedure due to inflexibility of the procedure they used. This inflexibility is one of the common deficiencies of the existing procedures. A lot of research exists in the literature proposing the applicability of mixture design of various cementitious mixtures using different experimental or mathematical approaches such as [100, 102, 118].

In order to provide a more accurate and flexible procedure, there are some other important issues required to be addressed. The mixture design variables must be interpretable and useful for both experimental design and development of objective functions. Especially, when experimental data are limited as in the case of many design studies, the accuracy of the models used as objective functions is of paramount significance. ML models can develop higher accuracy models without the need to assume the equation or terms of a regression model in advance. It is proposed that ML algorithms can be useful to generate a higher accuracy nonlinear model compared to classical regression models used in DOE, particularly when the data are irregularly obtained and are not well-structured. Another issue is that optimisation algorithms such as simplex are mainly developed for linear regression models. Therefore, ML algorithms might not be optimised using such traditional optimisation algorithms. This paper presents a procedure considering the aforementioned issues. As a first step to develop high accuracy cost, strength and slump models conventional MLR and machine learning (ML) algorithms such as artificial neural networks (ANNs) and Gaussian process regression (GPR) are used. The statistical performance of the models are analysed and the best models are selected. As a final step, the Particle Swarm Optimisation (PSO), a heuristic optimisation algorithm, for the multi-objective mixture design and optimisation of UHPC reinforced with steel fibers was employed. A reliable experimental dataset from published resources is used to develop the proposed models and to justify the final results. It was proved that the proposed procedure not only reduces the cost and increases the pertinence of UHPC characterisation but also leads to optimal mixtures when the designer faces strength-flowability-cost paradoxes.

5-3 Conceptual framework

The procedure proposed to mathematically optimise steel reinforced UHPC mixtures consists of four main stages, namely descriptive statement, experimental data streamlining, objective functions identification, and mathematical optimisation. The first stage covers the construction of descriptive statement based on the project specific requirements. It states the objectives and requirements of the project such as the UHPC mix design with at least 150 MPa compressive strength and 200 mm slump and minimum production cost. The second stage consists of identifying the mixture design variables and producing the relevant data using tests or observations. The third stage includes the development of objective functions that reflects the desired qualities and attributes of UHPC. This stage is particularly important as it can take into account the engineer's know-how. The fourth and last stage uses the optimisation algorithms to find values of the on the design variables once some objective functions are maximised or minimised subject to constraints. This procedure is briefly illustrated in Fig. 5.1.

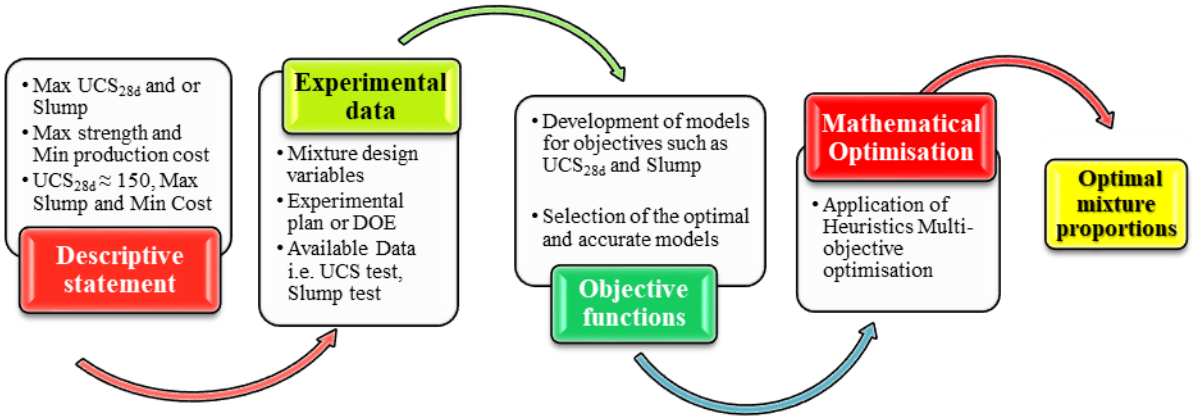


Fig. 5.1 Main stages of suggested procedure.

5-4 Design variables and experimental data

In experimental design of mixtures, one or more factors such as proportion or the amount of cement and the amount of water or water to cement ratio are changed in each specimen. These are the mixture design variables and called independent or input variables when it comes to modelling. These variables are changed to measure and study their influence on properties of concrete using tests and experiments such as UCS or slump tests. The acquired experimental data can be then analysed to come up with valid and objective conclusions. The output is the objective, which can be mathematically represented as a function of mixture design variables.

Ghafari et al. [21] conducted comprehensive experimental program on different mixtures of UHPC composed of Portland cement Type I 52.5 R (C), silica fume (SF), quartz flour (QF), siliceous sand (SA), water (W), polycarboxylate ether-based superplasticizers (SP), and steel microfibers. They used a D-optimal technique to produce the required data. A total of 53 experiments were conducted to measure the slump in fresh state in alignment with ASTM C1437, and unconfined compressive strength

at 28 day age using cubic specimens cured in water at 20°C, and the rest were cured at 90°C and 95% relative humidity according to with EN 196-1. The experiments were designed using D-optimal method [21]. They used 43 out of 53 data points for fitting the model, 5 for the lack of fit estimation, and 5 points to replicate the tests. In this paper, the unconfined compressive strength at 28 day age using cubic specimens cured in water at 20°C (UCS_{28d}) and slump values are considered as objectives. Note that the average of repetitive test results are considered here and therefore the number of data was reduced to 41 data points. It is recommended to prepare more than one specimen, suitable if 3 to 5, for each mix and consider the average value of the test results as the output so as to reflect uncertainty and experiments error in models.

Considering the fact that Ghafari et al. [21] used the same type of material to prepare their mixtures, the only component that varies is the amount of the material in the mixture. Variables of the mixture may include the amount of cement (kg/m^3) or in (kg), cement to the binder ratio, water to cement ratio or any other combinations. Ghafari et al. [21] used the ratio of each material weight to the total as the variable for model development and mixture design based on the D-optimal methodology. On the other hand, codes, specifications or other published resources often make recommendations for the water to cement ratio (W/C) or fine aggregate or sand to cement ratio (S/C) as cement is the main component. In this study, it is proposed that the amount of cement should be considered as the base and the amount of other constituents are measured as the ratio divided by cement amount. Assuming that X (kg/m^3, kg and so forth) is a mixture variable such as water, X/C is a scalar, normalized and proportional variable which is not sensitive to the unit of measurement. The independent variables considered here are sand, silica fume, quartz flour, water, superplasticizer and steel fiber to cement ratios which are respectively represented as S/C, SF/C, QF/C, W/C, SP/C, STF/C variables. By considering a value for C, other variables can be calculated for the mixture design. It is also assumed that these variables are changed to prepare the mixtures and specimens and therefore the employed database is changed. Obviously, the following equations can be used to calculate the amount of material in kg per cubic meter (kg/m^3).

$$\frac{C}{G_{s-c}}+\frac{S}{G_{s-\mathrm{S}}}+\frac{SF}{G_{s-\mathrm{SF}}}+\frac{QF}{G_{s-\mathrm{QF}}}+\frac{W}{G_{s-\mathrm{W}}}+\frac{SP}{G_{s-\mathrm{SP}}}+\frac{STF}{G_{s-\mathrm{STF}}}+V_{air}=1000 \tag{5-1}$$

In the equation above, G_s is the specific gravity (kg/m^3) of the material and V_{air} is the volume of air is commonly considered as 1% to 3%. The descriptive statistics of the considered independent variables are represented in Table 5.1. Fig. 5.2 illustrates the box plot of the amount of variables changes in each mix and the measured output.

Table 5.1 Descriptive Statistical indices of the variables in dataset.

Parameter	Independent/Design Variables (Input)	Dependent variables (Output)

	S/C	*SF/C*	*QF/C*	*W/C*	*SP/C*	*STF/C*	*UCS_{28d} (MPa)*	*Slump (mm)*
Ave	1.4	0.3	0.3	0.8	0.1	0.0	140.2	192.6
StD	0.26	0.07	0.11	0.06	0.01	0.03	27.05	14.77
Range	1.20	0.32	0.48	0.31	0.07	0.08	110.00	63.00
Min	1.00	0.18	0.00	0.59	0.09	0.00	90.00	171.00
Max	2.20	0.50	0.48	0.90	0.15	0.08	200.00	234.00
Sum	56.16	12.82	13.07	31.40	5.03	0.65	5748.00	7898.50
Count	41	41	41	41	41	41	41	41

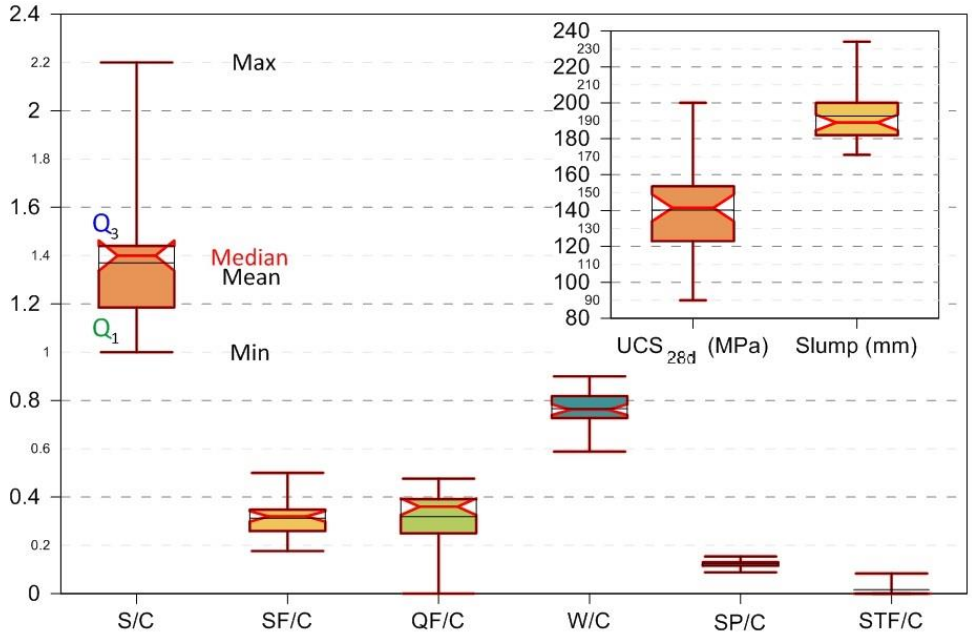


Fig. 5.2 The box plot of the variables in considered dataset.

5-5 **Model construction and selection**

Slump and UCS_{28d} can be represented as functions of the considered mixture design variables as follows:

$$\mathrm{UCS_{28d}} = f_u\left(\frac{\mathrm{S}}{\mathrm{C}}, \frac{\mathrm{SF}}{\mathrm{C}}, \frac{\mathrm{QF}}{\mathrm{C}}, \frac{\mathrm{W}}{\mathrm{C}}, \frac{\mathrm{SP}}{\mathrm{C}}, \frac{\mathrm{STF}}{\mathrm{C}}\right) \tag{5-2}$$

$$\mathrm{Slump} = f_s\left(\frac{\mathrm{S}}{\mathrm{C}}, \frac{\mathrm{SF}}{\mathrm{C}}, \frac{\mathrm{QF}}{\mathrm{C}}, \frac{\mathrm{W}}{\mathrm{C}}, \frac{\mathrm{SP}}{\mathrm{C}}, \frac{\mathrm{STF}}{\mathrm{C}}\right) \tag{5-3}$$

In this study, production cost is also considered as an objective. The production cost is the sum of the amount of constituents in the mixture each of which multiplied by their unit price. In the present paper, approximate unit prices per kg of materials ($/kg) are merely assumed to consider the production cost as an objective model after reviewing and calculating the average of prices for a specific location.

However, the real price of material must be considered depending on the location and the amount of order. The assumed values are given in Table 5.2.

Table 5.2 The assumed unit price of the SFR-UHPC constituents.

Item	C	S	SF	QF	W	SP	STF
Price ($/kg)	0.1	0.01	0.5	2	0.003	0.4	3.4

Accordingly, the normal production cost as an objective function is:

$$Cost\ (\$) = 0.1C + 0.01S + 0.5SF + 2QF + 0.003W + 0.4SP + 3.4STF \tag{5-4}$$

As the variables here are S/C, SF/C, QF/C, W/C, SP/C, STF/C and must be fixed in the optimisation process, the cost function for optimisation is as follows:

$$\frac{Cost}{C}\left(\frac{\$}{\text{kg of Cement}}\right) = 0.1 + 0.01\frac{\text{S}}{\text{C}} + 0.5\frac{\text{SF}}{\text{C}} + 2\frac{\text{QF}}{\text{C}} + 0.003\frac{\text{W}}{\text{C}} + 0.4\frac{\text{SP}}{\text{C}} + 3.4\frac{\text{STF}}{\text{C}} \tag{5-5}$$

Once a well-designed plan of experiment along with sufficient data are available, it can be expected that the models would exhibit acceptable performance. However, it is not always possible to fit models on data using traditional regression methods. Generation of model using conventional linear modelling is pretty straightforward given that the relationship is quite clear. However, the relations between input variables and output variables such as strength and cost are far more complex. Therefore in addition to conventional multiple linear regression two robust AI algorithms Multi-Layer Perceptron Neural Network (MLPNN) and Gaussian Processing Regression (GPR) are used to develop Slump and UCS_{28d} models. After evaluating and comparing the accuracy of the models, the best performing models are chosen for multi-objective mix design and optimisation of UHPC.

5-5-1 **MLR**

MLR aims to fit a linear model on data. The general form of the model is as follows:

$$y = \sum_{i=1}^{n} \alpha_i x_i + \varepsilon \tag{5-6}$$

y is the output, x_i is the i^{th} independent variable, n is the number of input or independent variables and α_i is the coefficient typically calculated while fitting the model using least squares method. This coefficient mainly indicates the mean change in the output variable for each one unit change in one input variable when all other input variables are held constant.

5-5-2 **MLPNN**

Generally, ANNs contain a series of nodes arranged in sequential layers mainly including the input and

output layers and one or more hidden layer(s) between them. The input-processing-output procedure starts at the first layer which is the input layer. Thereafter, inputs are sent to hidden layer(s) applying weights based on different architectures [30]. After adding a bias term, the network produces new outputs, which would be close to the observed output data with a high accuracy. MLPNN is a commonly used architecture of ANNs with a multi-layer feed-forward structure which is typically trained with a backward propagation of errors algorithm for learning [119]. Except for the input nodes which are fixed, other nodes are processing elements applying a nonlinear activation function, usually sigmoid functions. The major distinction of MLPNN is the fact that each neuron in a layer is interconnected by weights to all of next layer's neurons where independent calculations might be carried out on input data.

Generally, in each processing node, the input (x_i) is multiplied by a weight coefficient (w_i) that exists between two layers. The output (y) is acquired through summing the products of the inputs and weights and finally transformed by an activation function. This process can be summarized and presented by the following equation:

$$y = f(\sum_{i=1}^{n} w_i x_i + Bias_i) \tag{5-7}$$

After completing the feed-forward iteration, the process is reverted. In this phase, a backward propagation of errors algorithm would be employed to alter the network weights until the desirable fitness is reached and the relevant criterion is satisfied [31]. It is noteworthy that that one epoch refers to performing a complete feed-forward and a backward iteration. Fig. 5.3 shows a schematic architecture of MLPNN for two input and one output variables, with two hidden layers.

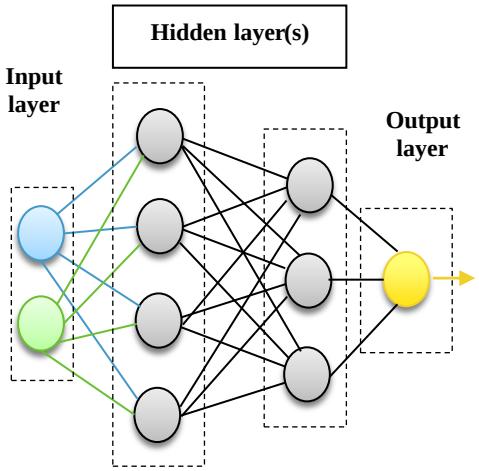


Fig. 5.3 A typical MLP-NN architecture.

5-5-3 **GPR**

Gaussian process regression (GPR) is a nonparametric kernel-based probabilistic regression approach which can also be categorized as a supervised machine learning algorithm [43, 120]. Given the training

dataset $D = \{(x_i, y_i) | i=1, \ldots, n\}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}^n$ are the input and output vectors, taken from an unknown distribution. GPR considers y as the following function [43, 121]. :

$$y = f(x) + \varepsilon \tag{5-8}$$

where $\varepsilon \sim N(0, \sigma^2_n)$. ε follows the Gaussian distribution with a mean of 0 and error variance of σ^2_n considering the output values or y_i.

If $\{f(x), x \in \mathbb{R}^d\}$ is a Gaussian process (GP), then for a given i observations $x_1, x_2, \ldots, x_i$, Gaussian transformed function is a joint distribution of the random variables $f(x_1), f(x_2), \ldots, f(x_i)$. Given two x_i and x_j, A GP can be defined by $m(x)$ and $k(x_i, x_j)$ which are the mean and covariance kernel functions respectively as represented as follows:

$$m(x) = E(f(x)) \tag{5-9}$$

$$k(x_i, x_j) = Cov(f(x_i), f(x_j)) = E[\{f(x_i) - m(x_i)\}\{f(x_j) - m(x_j)\}] \tag{5-10}$$

where $k(x_i, x_j)$ is usually parameterized by a set of kernel parameters or hyperparameters. Therefore:

$$f(x) \sim GP(m(x), k(x_i, x_j)) \tag{5-11}$$

and accordingly:

$$y \sim GP(m(x), k(x_i, x_j) + \sigma^2_n (x_i - x_j)) \tag{5-12}$$

$m(x)$ or the mean function is typically a constant which can be zero or the mean of the training dataset. Some common kernel functions $k(x_i, x_j)$ are constant, linear, radial basis, squared exponential, rational quadratic and Matern 5/2 kernel functions. Kernel functions may consist of a combination of multiple kernels such as linear and radial basis functions [43]. After the process of training and fitting different GPR models on the employed data, Squared Exponential GPR (SEGPR) model outperformed other GPR models. Therefore, it was selected as the best GPR model to proceed. The exponential kernel function used in SEGPR is as follows:

$$k(x_i, x_j) = \sigma_f^2 \, exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right) \tag{5-13}$$

where the maximum allowable covariance is defined as σ^2_f and l is the length scale of the kernel function. GPR is a capable data-driven approach due to the presence of hybrid kernel and Gaussian functions. The parameters or hyperparameters, such as σ^2_n, l or σ^2_f are updated during the training process by Bayesian inference until convergence is reached or a stop criterion is met. These parameters

are calculated by optimising the log likelihood function using a Bayesian optimisation and gradient-based optimisation algorithms. More details can be seen in [122-124]. Finally, the model can be used for simulation and prediction aims.

5-5-4 **Constructed models**

MLR fits a linear model on data. This means that the relationships between input and output variables are assumed to be linear. The coefficients in MLR models in Table 5.3 explains the variance in the dependent variable while input variables are changed. However, nonlinear relationships may exist. For this aim, two robust AI algorithms, SEGPR and MLPNN, are developed which are black-boxes rather than equations. Such models have a fixed structure where some parameters which are changing in order to improve the performance. Note that such models can be easily run using a computer.

Table 5.3 Models of UCS_{28d} and Slump developed by MLR.

Model	MLR
UCS_{28d} (MPa)	$268.55 + 4.4\frac{S}{C} + 141.46\frac{SF}{C} + 101.12\frac{QF}{C} - 330.34\frac{W}{C} + 253.93\frac{SP}{C} + 693.07\frac{STF}{C}$
Slump (mm)	$93.74 - 22.68\frac{S}{C} - 144.16\frac{SF}{C} - 58.08\frac{QF}{C} + 247.31\frac{W}{C} + 50.5\frac{SP}{C} - 127.19\frac{STF}{C}$

In terms of ML and AI prediction algorithms, underfitting or overfitting may occur. Underfitting is the weak performance of the model or the incapability of the model to perform estimation. Overfitting is the weakness of the model in predicting unseen data. In order to avoid overfitting, some recommendations have been made by several researchers [30, 112]. It is recommended to split the data into at least two different groups i.e. training and testing groups. The model is developed using train data and the generalization performance of the model is examined on test data, which are unused. Another way is to use cross-validation techniques. Cross-validation aims to check the model's performance on new data [125]. This leads to reducing the bias and overfitting. In this paper, a k-fold cross validation where k is equal to 5 is used. In this approach, the data is firstly divided to k equal-sized folds or groups of data. Then, the algorithm develops the model using k-1 folds of data and hold out one fold in each iteration. This process is repeated for all folds and ultimately the average of the models will be considered as the best model as it produces the mean error on all folds. Note that the number of experimental data is limited and the main objective is to find the best model that fits them as there are no more data for validation or testing the model. This is the reason behind using a cross-validation rather than splitting the data into train and test. As the cross-validation is used for development of ML models, it can be considered that the risk of overfitting is reduced and the model is generalizable. Here, MATLAB was used to write, program and develop MLPNN and GPR models for Slump and UCS_{28d}.

In MLPNN structure, many parameters need customization such as the number of hidden layers, the number of nodes in each hidden layer, learning rate, number of epochs, and activation function type between different layers [31]. The number of hidden layers and consisting nodes or neurons directly influence the model performance. However, there is no specific method to find their optimal numbers as it depends on the nature of data [126]. The optimal number of hidden layers and nodes can be calculated using a trial and error process [127]. Researchers recommended that a single hidden layer network would be sufficient to develop nonlinear models [31, 128]. After a process of trial and error using Levenberg–Marquardt (trainlm) and back-propagation of error algorithms, an architecture with one hidden layer consisting of 10 neurons achieved good results. Furthermore, the transfer function between the input and hidden layer was considered to be a hyperbolic tangent or tan-sigmoid function which is of the form $(e^{x}-e^{-x})/(e^{x}+e^{-x})$. Besides, a pure line, which is a linear function is utilized as the transfer function between the hidden layer and output layer. During the process of training, weights and biases values may vary in each run until the fittest model is obtained. The learning rate of 0.05 was chosen and the model was trained for 500 epochs and iterations for both Slump and UCS_{28d}.

As already mentioned, the parameters required to be identified in SEGPR model are σ^2_n, and kernel parameters l and σ^2_f. A subset data used to train GPR model is commonly stored as an active set vector. $K(X_{new}, A)*\alpha$ computes the predictions in each iteration where K is the matrix of kernel products between X_{new} and active set vector A and α is a vector of weights. Here, all training data were used as A. The log likelihood function is optimised using Quasi-Newton algorithm in MATLAB. Finally, σ_n, l and σ_f are obtained equal to 0.27, 14.27 and 47.2 for UCS_{28d} and 0.148, 1.41 and 15.33 for Slump models respectively.

5-5-5 **Model selection**

In order to examine the performance of the models, the coefficient of correlation (R), mean absolute error (MAE) and the root mean square error (RMSE) between the predicted and experimental values of output, i.e. Slump and UCS_{28d}, are used. The greater the R value and lower the RMSE and MAE values, the better the model. The performance of different Slump and UCS_{28d} models are given in Table 5.4.

Table 5.4 Prediction performance of different model on dataset

Model	Slump (mm)			UCS_{28d} (MPa)		
	MLR	MLPNN	SEGPR	MLR	MLPNN	SEGPR
R	0.764	0.987	0.999	0.925	0.973	0.999
R^2	0.583	0.974	0.998	0.856	0.948	0.998
RMSE	9.4	2.3	0.002	10.1	6.2	0.005
MAE	7.9	1.4	0.002	8.4	4.4	0.004

It can be seen in Table 5.4 that SEGPR outperformed the two other models in terms of accuracy for indirect estimation of both Slump and UCS_{28d}. In order to better reflect the performance and accuracy of different models, Fig. 5.4 represents scatter plots of the predicted versus experimental values of Slump and UCS_{28d} obtained in each case.

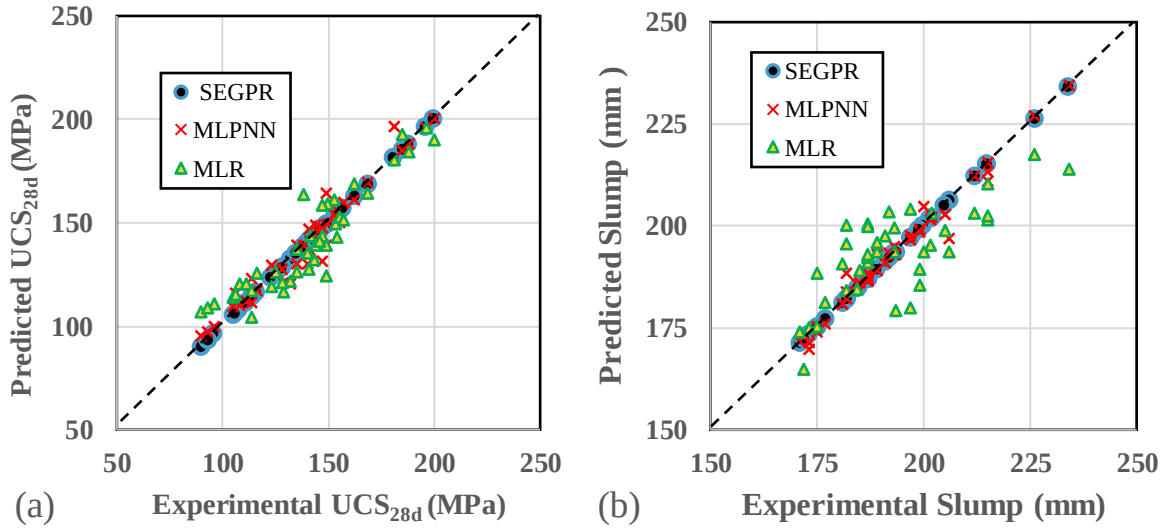


Fig. 5.4 Scatter plots of the predicted versus experimental values of (a) UCS_{28d} and (b) Slump obtained by different models.

The residual error (RE), which is equal to the difference between predicted and experimental values made by each model, can be seen in Fig. 5.5. According to Fig. 5.5 (a), the error made by SEGPR, MLPNN and MLR UCS_{28d} models for test number 5 is around 0, 15 and 10 MPa respectively. When it comes to Slump models, Fig. 5.5(b) indicates that the least error is also made by SEGPR model. For example in case number 3 in Fig. 5.5(b), SEGPR, MLPNN and MLR made 0, 6 and 14 mm error in estimating the slump of UHPC. It can be realised that the least error is made by SEGPR Slump and UCS_{28d} models which is around 0 for all cases. These indicate the robustness and high accuracy of SEGPR as a machine learning tool for predictive modeling of the properties of UHPC, particularly when a limited number of experimental data is available. Additionally, the reliability of optimisation results depends on performance and accuracy of objective models where machine learning algorithms such as GPR are useful. Also, note that MLPNN and SEGPR models are developed using a process of cross-validation, which means that the risk of overfitting is lower than that of MLR model. Due to the higher accuracy of Slump and UCS_{28d} models found by SEGPR, they are chosen here as objective functions for the multi-objective design and optimisation of UHPC mixture, in addition to the cost function.

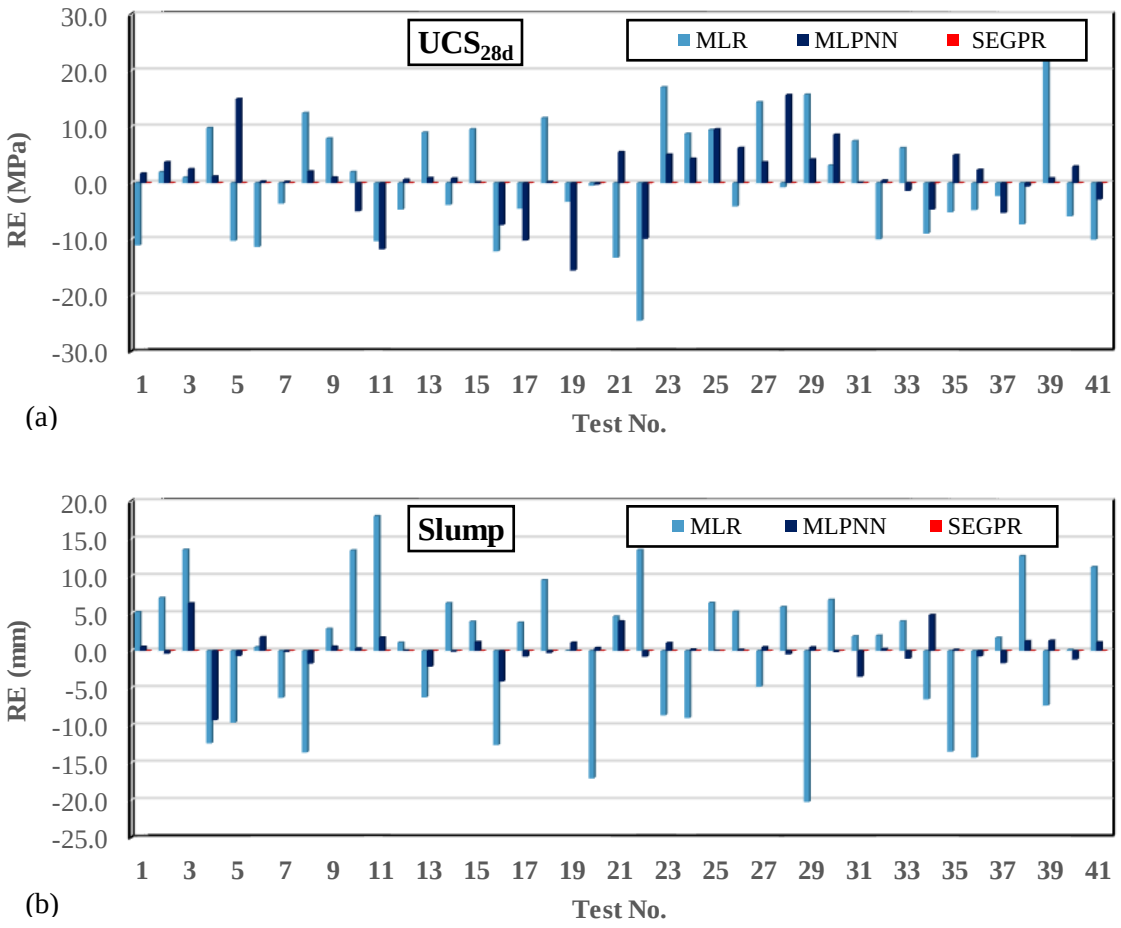


Fig. 5.5 Column plots of the RE obtained by each model for estimation of (a) UCS_{28d} and (b) Slump.

5-6 Particle swarm optimisation

After defining the statement of the project, collecting the relevant experimental data and developing the objective functions, the project can be converted to a mathematical optimisation problem. A typical engineering constrained optimisation problem can be defined as follows [97, 100, 115]:

$$\begin{aligned} &\text{Objective(s): Minimise or Maximise } f_m(X) && m = 1, 2, 3, \ldots, M \\ &\text{Subject to: } a_i \le x_i \le b_i && i = 1, 2, 3, \ldots, n \end{aligned} \tag{5-14}$$

where $f_m(X)$ is the m^{th} objective function such as Cost, Slump or UCS_{28d}; a solution $X \in R^n$ is a vector of design variables $X = [x_1, x_2, x_3, \ldots, x_i, \ldots, x_n]^T$, a_i and b_i are lower and upper bound of x_i, n is the number of design variables and xi is the design variables, which are here the UHPC mixture proportions. In this paper various scenarios are considered in order to represent the single-objective or multi-objective mix design and optimisation of UHPC using a metaheuristic called particle swarm optimisation algorithm. Generally, optimisation algorithms aim to find the optimal values of design variables from within a range when one or more objectives are being maximised or minimised [115, 129]. In most practical problems, multiple objectives commonly exist which may conflict, e.g. improving one objective such as UCS_{28d} increasing the production costs. Optimisation algorithms have been proposed to solve a system of equation(s) where it is not simple or possible to find solutions

mathematically, e.g. when there are nonlinear equations and the number of variables are more than that of equations. The purpose of multi-objective optimisation (MOO) algorithms is to explore a set of solutions, each of which meets multiple objectives at an acceptable level without being dominated by any other solution [115, 129].

PSO is a stochastic population-based search algorithm which is inspired from the social behaviour of birds or fish within a flock seeking food [97, 98, 115, 130]. This algorithm was firstly introduced by Kennedy and Eberhart [98]to solve single-objective optimisation (SOO) problems. Since then, many researchers have developed different variants and approaches for solving SOO or multi-objective optimisation (MOO) problems such as [63, 130-133]. In PSO, possible solutions are called particles which are flown in a hyper-dimensional search space with an initial position and velocity. A swarm consists of a set of particles and leaders which are used to update the best position and velocity of the swarm. In order to find solutions which meet the objectives, any particle in a swarm changes its position and velocity within the search space according to its own experience and the social psychological tendency and success of other particles and leaders. This process continues until the optimal solutions are found. Fig. 5.6 represents a typical flow chart for the PSO algorithm.

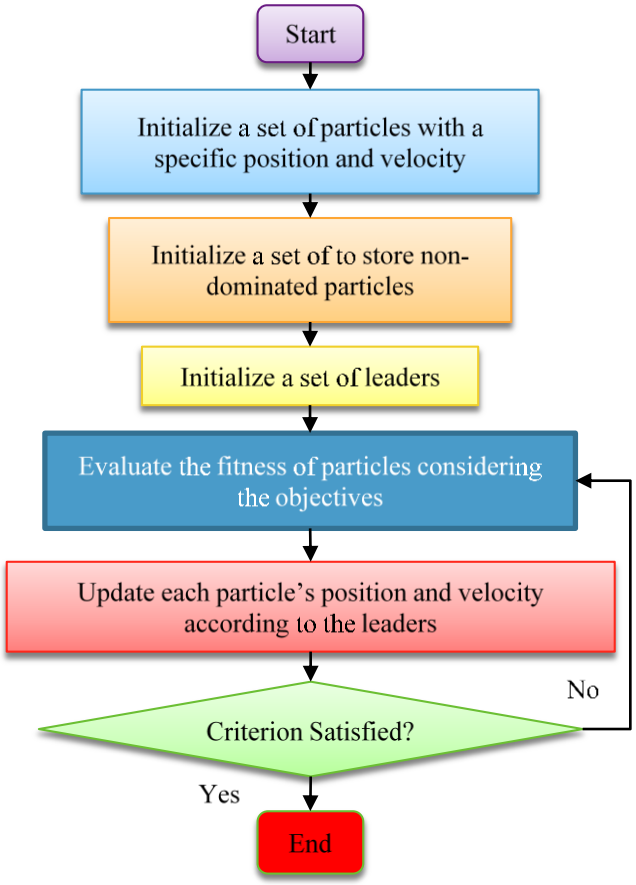


Fig. 5.6 A typical flow chart of multi-objective particle swarm optimisation algorithm.

In this paper, the PSO variant proposed by [131, 134, 135] and programmed as a code in MATLAB was used as the optimisation algorithm for SOO or MOO scenarios. This algorithm based on Pareto dominance and using a factor to filter out the list of available leaders in a swarm and different mutation

operators. Generally, in SOO problems, the purpose is to find solutions satisfying merely one objective. When there is more than one objective, i.e. MOO, the optimum solution cannot only meet one objective alone, when there exist other important objectives to be satisfied. In this situation, selecting only one solution that satisfies one objective leads to a sacrifice or ignorance of one or more other objectives. This trade-off feature between the non-dominated solutions makes it reasonable to find a variety of non-dominated solutions, which is called Pareto-optimal solutions, before making a decision [133]. In order to consider these, multi-objective optimisation aims to find a set of solutions, Pareto-optimal front, which are diverse enough to meet all the objectives. Fig. 5.7 represents a Pareto-optimal front in which solution A can be selected when maximum UCS is of priority or C can be the best solution when cost is more important. B can be selected when both objectives are important and accordingly other solutions considering the priority of the objective. In this way, the selection of solutions depends on the designer or decision-maker. Note that the number of solutions found by the algorithm depends on the population size of the particles which are possible solutions and the number of generations considered in the PSO algorithm customization.

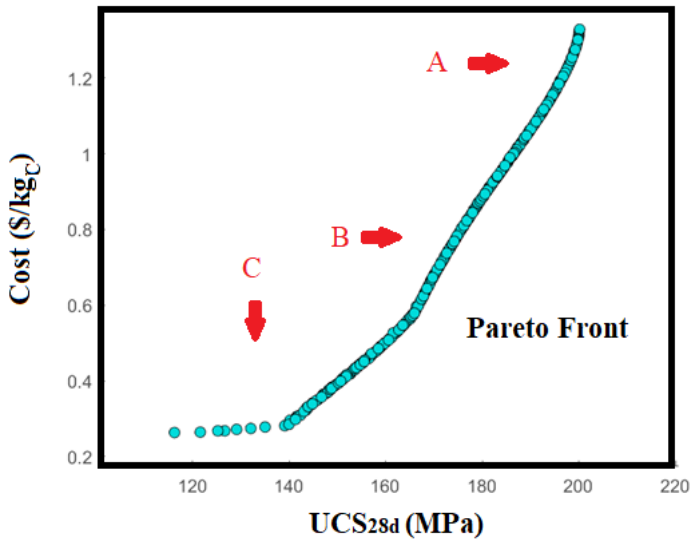


Fig. 5.7 A Pareto-optimal front of solutions when two UCS_{28d} and Cost are objective functions to be optimised.

In the present study, different scenarios are considered to represent the broad applicability of the mathematical optimisation procedure for multi-objective mix design of UHPC. In each scenario, UHPC is designed to meet specific objectives such as production cost, maximum or desired UCS_{28d} or Slump and so forth. Here, solutions and Pareto fronts are given numerically. The results are summarized in Table 5.5. Some of the experimental data are also given in Table 5.5 in order to compare and justify the results found by the proposed procedure. For instance, first scenario is looking for a mixture with maximum strength and other objectives are not prioritised, while in fourth scenario is a mixture with a UCS_{28d} of around 150 (MPa), maximum Slump and minimum Cost are desired. Some experimental data are used in scenarios 1-3 in Table 5.5 to validate the results of the optimisation procedure. After confirming the robustness of the proposed procedure, scenarios 4 and 5 are considered to represent the

feasibility of defining different scenarios or more objectives which needs several experiments and the proposed procedure can give results without doing new experiments. As can be seen, the suggested procedure provides the designer the best alternatives satisfying all the project specific requirements and enables them to choose the optimal mix design amongst a number of alternatives. It is clear that the number of solutions meeting the objectives are larger than conventional mixture design alternatives. More importantly, it is always possible to increase the number of objectives for multi-objective mixture design of steel fiber reinforced UHPC.

Table 5.5 The results of the proposed procedure based on different scenarios.

Scenario				**Method**	**Mixture design variable**						**Objective**		
No.	UCS_{28d} (MPa)	Slump (mm)	Cost ($/kg-C)		S/C	SF/C	QF/C	W/C	SP/C	STF/C	UCS_{28d} (MPa)	Slump (mm)	Cost ($/kg-C)
1	Max	-	-	Proposed Method	1.437	0.266	0.383	0.720	0.133	0.080	200.230	187.350	1.341
				Validation (Experimental)	1.458	0.250	0.375	0.750	0.125	0.083	200.000	182.000	1.325
2	-	Max	-	Proposed Method	1.409	0.248	0.423	0.832	0.130	0.000	92.530	232.660	1.138
				Validation (Experimental)	1.417	0.250	0.417	0.833	0.125	0.000	93.000	234.000	1.125
3	150	Max	Min	Proposed Method	1.318	0.276	0.342	0.702	0.118	0.000	150.157	179.485	0.984
					1.503	0.289	0.290	0.717	0.137	0.000	149.214	187.831	0.897
					1.088	0.307	0.219	0.802	0.101	0.061	150.000	193.489	0.953
					1.874	0.483	0.176	0.631	0.130	0.000	148.508	190.064	0.765
				Validation (Experimental)	1.036	0.321	0.286	0.714	0.107	0.000	149.000	199.000	0.888
					1.240	0.320	0.360	0.800	0.120	0.040	153.500	191.000	1.179
					1.400	0.240	0.360	0.760	0.120	0.000	149.000	187.000	1.004
4	Max	-	Min	Proposed Method	1.402	0.280	0.356	0.708	0.136	0.080	198.990	188.675	1.293
					1.017	0.219	0.008	0.764	0.130	0.003	115.390	198.600	0.301
					1.406	0.259	0.063	0.628	0.147	0.078	170.258	191.717	0.694
					1.425	0.239	0.002	0.607	0.145	0.039	151.158	192.853	0.430
					1.389	0.263	0.120	0.648	0.145	0.078	175.738	191.037	0.809
					1.370	0.241	0.009	0.646	0.146	0.072	162.053	192.433	0.556
					1.446	0.261	0.385	0.714	0.133	0.080	200.008	186.847	1.342
					1.437	0.258	0.086	0.615	0.147	0.071	170.439	191.496	0.718
5	Max	Max	Min	Proposed Method	1.001	0.180	0.000	0.592	0.130	0.000	130.976	194.677	0.254
					1.317	0.280	0.234	0.702	0.135	0.072	184.427	188.341	1.020
					1.446	0.261	0.385	0.714	0.133	0.080	200.008	186.847	1.342
					1.434	0.306	0.294	0.702	0.128	0.079	192.688	182.442	1.176
					1.261	0.326	0.425	0.836	0.129	0.078	168.460	195.663	1.444
					1.001	0.180	0.000	0.592	0.130	0.000	130.977	194.677	0.254
					1.260	0.246	0.400	0.846	0.132	0.008	101.102	217.240	1.119

5-7 **Conclusion**

Concrete is undoubtedly the mostly used construction and building material. Considering the production costs and the abundance of mixture designs, the optimisation of recent categories of concrete such as UHPC is a very important issue which is commonly tackled by conducting experiments. In order to reduce the efforts in experimental design approach, various methodologies have been proposed for mixture design and optimisation of concrete such as RSM. In any mathematical procedure, the accuracy of the models used as objective functions is of paramount significance. Although a plan for doing experiments helps to collect suitable data, one must be able to produce high-accuracy models even if the data are informally obtained where ML algorithms were proposed to be useful. This paper demonstrated that ML algorithms are more robust than conventional regression approaches such as MLR and RSMs, particularly when it comes to nonlinear relationships and not well-structured data as in the case of UHPC design optimisation. It was shown that the design variables can be considered in a meaningful way for experimental design, modelling, and optimisation. In this paper, a variant of GPR, i.e. SEGPR, was confirmed to be robust for the development of high accuracy models that predict the key properties of steel fiber reinforced UHPC. A metaheuristic algorithm, namely PSO, was used to find the optimal mixture design variables to produce UHPC based on different pre-defined scenarios. The obtained results confirmed the robustness of the proposed procedure. The major advantage of the proposed procedure lies in its flexibility. In the suggested procedure, any nonlinear and black-box models can be used as an objective function due to the optimisation algorithm used compared to traditional methodologies which are based on RSM and simplex or mathematical programming algorithms. For example, if MLPNN had a better performance than SEGPR for Slump, it can be used as the Slump objective function in addition to UCS_{28d} model obtained by SEGPR. Due to the higher accuracy of Slump and UCS_{28d} models found by SEGPR, they are chosen here as objective functions in the mathematical optimisation procedure for the design of UHPC mixture under study. Finally, it is noteworthy that the accuracy of the results obtained by using the proposed procedure is dependent on the quality of experimental data, the accuracy of the models and the capability of the algorithms used for optimisation.

Chapter 6 Multi-objective mixture design optimisation of composite materials using an engineered AI framework: Sustainable SCC

6-1 Abstract

Composite materials such as self-compacting concrete (SCC) are composed of precise proportions of materials with specific properties mixed based on determined conditions. Optimum design of such materials using trial-and-error experiments is challenging and costly due to multiple influential parameters and nonlinear relationships. Furthermore, designers are required to consider multiple objectives and constraints pertaining to emphasis on sustainability issues, and make the best possible decisions at a faster pace while decreasing costs. Recently, researchers have demonstrated exciting applications of mathematical methods such as statistical mixture design and artificial intelligence (AI) techniques for prediction of properties and mixture design of materials like SCC. However, proposed methods commonly consider the proportions of constituents as variables only and ignore unique features of materials or other influential factors. Several issues are still required to be addressed with respect to conceptualisation, dealing with influential factors, and discovery of robust mathematical tools. This paper engineers the concept and propose an innovative framework for multi-objective mixture design optimisation (MOMDO) of sustainable SCC using AI algorithms. Various assessments and design examples are considered to confirm the applicability of the proposed approach. Gaussian process regression is found to be robust for predictive modeling of properties of SCC, and non-dominated sorting genetic algorithm called NSGA-II for optimisation and finding mixture designs. Using this approach as a universal tool certainly enables a more sustainable production and consumption and allows for making the best design decision fast while reducing the costs and efforts compared to traditional design methods.

6-2 Introduction

Concrete is the major composite construction material due to its valuable and controllable properties in fresh or hardened states under different conditions including flowability, workability, compressive strength, fire resistance, thermal and acoustic insulation properties and so forth. Self-compacting concrete (SCC) is designed to be flowable under its own weight, pass through reinforcement and fill the required space completely without segregation and mechanical vibrating compaction [136, 137]. These beneficial properties of SCC allow to construct more durable concrete structures without the need for skilled workers. This would result in improving the quality and speed of construction while reducing the overall costs and risks [138]. SCC is more expensive than conventional concrete due to higher use of binder and other admixtures such as superplasticizer (SP). In terms of structural applications, the unconfined compressive strength (UCS) is obviously the key mechanical property of hardened concrete. UCS is typically measured through lab tests on specimens after 28 days of curing (UCS_{28d}).

Furthermore, various experiments and criteria are recommended by standards or specifications to design and measure the workability of fresh SCC, such as EFNARC [139], as are given in Table 6.1.

Table 6.1 Experiments for measuring workability properties of SCC according to EFNARC.

Property	Test Method	Unit	Recommended range
Flowability / Filling ability	Slump-flow by Abrams cone	mm	650 - 850
	T_{50cm} slump flow	sec	2 - 5
	V-funnel	sec	6 - 12
	Orimet	sec	0 - 5
Passing ability	J-ring	mm	0 - 10
	U-box (h2-h1)	mm	0 - 30
	L-box (h2/h1)	-	0.8 - 1.0
	Fill-box	%	90 - 100
Segregation resistance	V-funnel at T_{5min}	sec	0 - 3
	GTM screen stability test	%	0 - 15

Cement production requires a rigorous process resulting in considerable carbon dioxide (CO_2) emission, which is not eco-friendly. This CO_2 emission accounts for at least 5-7% of the total greenhouse gas emissions each year [2, 3]. Constituents of concrete are basically acquired from natural resources and might not be recyclable after use. Emphasis on sustainability issues with respect to excessive use of natural resources and environmental concerns has risen to importance in recent projects. Considering these issues, concrete requires to be designed to satisfy multiple objectives to achieve a more sustainable and cleaner concrete production in today's projects as categorised Fig. 6.1.

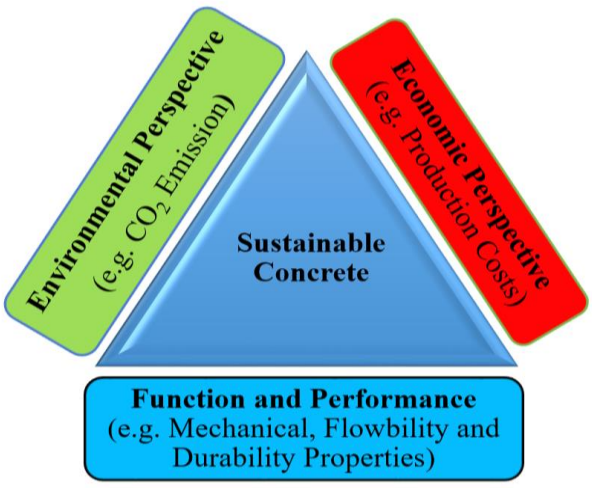


Fig. 6.1 General requirements for the design of sustainable concrete here.

Recently, various waste and recycled materials or by-products have been proposed to be used as the main material or partial replacement to reduce the use of costly materials such as cement and natural resources and strive for a higher level of sustainability such as granulated blast-furnace slag (GGBS or GGBFS), silica fume (SF), fly ash (FA) and meta-kaolin (MK) [137, 140]. Here, such materials are

named partial cement replacement (PCR) materials or supplementary cementitious materials (SCMs). Such materials can enhance the properties of concrete depending on their chemical compositions and physical attributes. These materials contain different amounts of silicon oxide or silica (SiO_2), both amorphous and crystalline, calcium oxide (CaO), ferric oxide (Fe_2O_3), aluminium oxide or alumina (AL_2O_3), magnesium oxide (MgO), and other minerals or oxides which are responsible for the cementitious and strength-gain properties. If an optimal amount is used, some PCRs may represent self-cementing or hydraulic, and pozzolanic properties in the presence of water during and after the hydration process of cement. Hydraulic reactivity means that the material itself reacts with water with or without the presence of Portland cement, while pozzolanic reactivity requires the hydrated cement paste in order to form more strength phases [141]. PCRs with high CaO and reactive SiO_2 commonly represent hydraulic reactivity where pozzolanic materials have high content of SiO_2 and other main oxides and are also low in CaO. Regardless of reactivity, they may also be used as fillers due to their powdery size and shape of their particles in order to reduce the costs or improve the properties such as flowability, viscosity and density. In addition to improving the properties, use of waste or by-products such as FA and GGBS, particularly as replacement of reactive or unreacted cement, leads to reduction of the use of natural resources, a cleaner production, and a profitable waste management strategy which makes concrete more sustainable.

Aggregates in concrete are generally considered as inert fillers to provide bulk, stability, resistance to erosion and shrinkage, thermal resistance, and other desired properties in concrete. Aggregates are typically divided into fine and coarse grades. Those with the maximum particle size and less than 4.75 mm are commonly considered as fine aggregate (FAgg), and those with particle size above 4.75 mm are called coarse aggregate (CAgg). Additionally, solid particles with a size of less than 200 μm are commonly named as powders and fillers which are used to provide density and a more homogenous texture. Properties of aggregates such as maximum size, particle size distribution (PSD) and gradation, unit weight and water absorption significantly influence both fresh and hardened properties of concrete. Fig. 6.2 illustrates common PSD of some powders, fine and coarse aggregates used by different researchers and specifications in the literature for producing SCC as reported by [142-148]. It can be seen that cement, limestone powder (LP), FA, SF, MK and GGBS are in form of powders regarding their particle size.

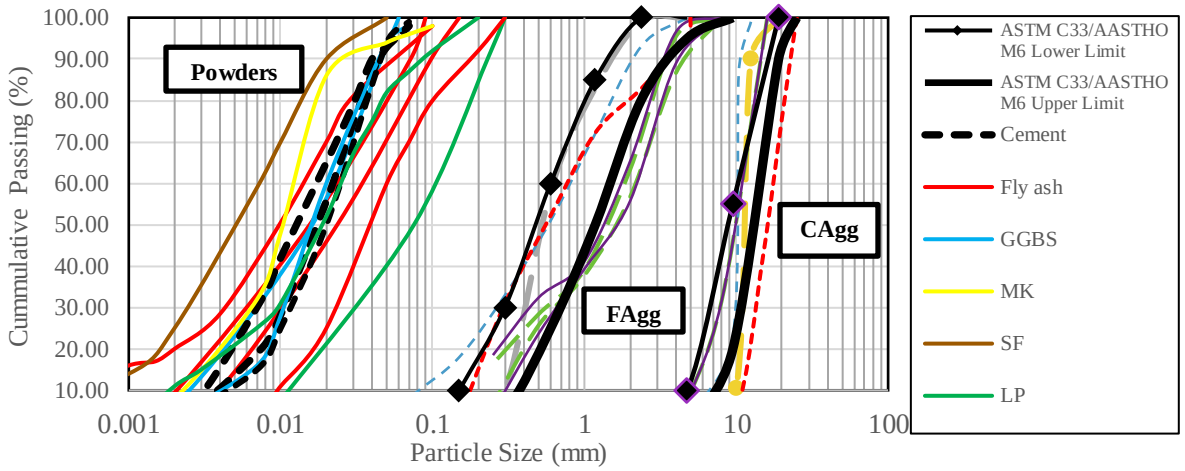


Fig. 6.2 PSD examples of some powders and aggregates used to produce SCC.

In order to have a comparative vision regarding the particle size and distribution, shape of powdery materials and SCMs, images taken using secondary electron (SE) scanning electron microscope (SEM) are represented in Fig. 6.3. It can be seen that particles of cement (OPC) and GGBS have angular shape, FAF and SF can be regarded as sphere-like shape and MK has flake shape. Particle's shape, size and sized distribution, morphological characteristics and texture and so forth produce influence on properties of SCC.

Optimal mixture design of composite materials such as SCC is challenging due to the large number of influential factors, objectives and constraints. Traditionally, concrete is designed through time–consuming and expensive trial-and-error experiments, past experience and recommendations of standards and specifications. However, a large number of tests may still be required, particularly to meet multiple objectives and constraints such as cost, slump and compressive strength which may contradict. Experimental design would be very costly or even may not lead to the optimal mixture due to the complexity of the system and large number of influential factors and desired objectives. In fact, the available experimental results are impressively valuable to analyse and make data-driven solutions and decisions. Notably, specifications and standards are also developed based experimental data. Furthermore, there is an increasing demand for designers to consider multiple requirements with respect to emphasis on sustainability issues and make the best possible decisions at a faster pace while decreasing costs. A smart strategy is to use such data to develop mathematical methods for predictive modelling and optimisation aims. In this regard, researchers have been attempting to find proper mathematical methods to enhance the process and reduce the costs and efforts of mixture design of composite materials such as SCC. Various approaches have been proposed such as statistical mixture design (SMD), design of experiment (DOE), response surface methodology (RSM), artificial intelligence (AI) algorithms, machine learning (ML), and optimisation algorithms [19, 100, 115].

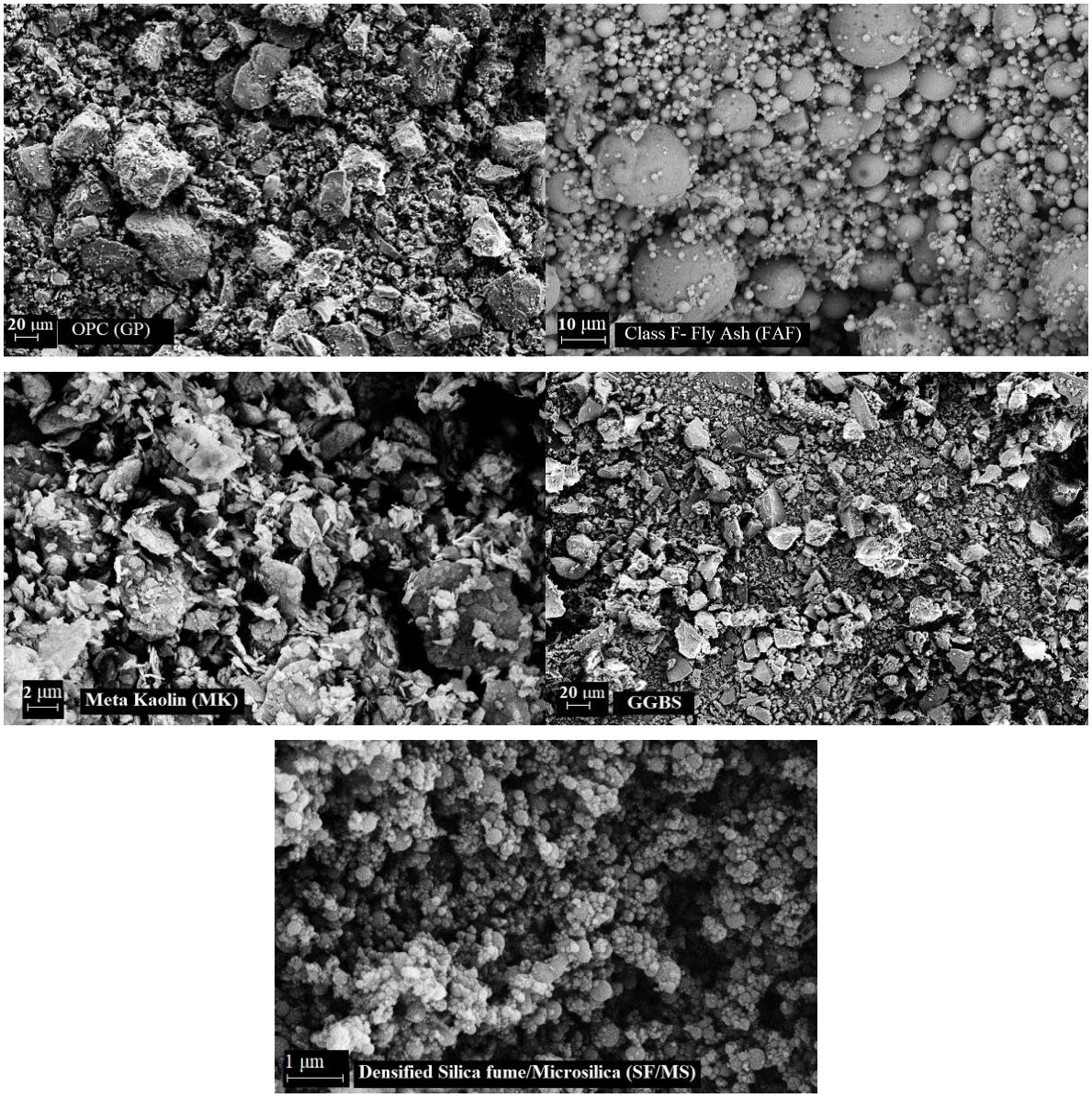


Fig. 6.3 SEM images of some powdery materials and SCMs used in concrete.

Generally, SMD methods, DOE and RSM aim to provide a strategy for doing experiments to produce efficient data for calibrating pre-determined statistical models [109, 149-151]. Researchers have applied various DOE, RSM and classical SMD methods such as central composite design, factorial design, D-optimal for optimisation and mixture proportioning of types of concrete [21, 107, 152-155]. In such approaches, the structure of the model must be known, e.g. a multiple linear, polynomial regression with different terms or a nonlinear model. Such models are commonly simple and calibrated based on analysis of variances in the output when changing the amount of one or more influential factors which are input or independent variables of models. In fact, some influential terms and relationships might not be known and therefore may not be considered in such models which is a remarkable weakness [108]. Considering the number and variability of influential factors and multiple requirements in mixture design of concrete, determination of an appropriate experimental plan or DOE and the model is also very difficult. Therefore, such approaches are more useful when dealing with

determined types of materials and small number of influential variables and or factors so that a perfect DOE is available for calibrating the considered model [100].

Researchers have demonstrated the capability of some AI algorithms for data analysis and modeling of various properties of different composite materials and cementitious mixtures such as concrete [112, 156-159]. These algorithms are inspired from intelligence in learning, natural phenomena such as the behaviour of animals or their neural system and so forth. They enable computers solve different classification, regression, and optimisation problems in supervised or unsupervised forms. In terms of regression and predictive modeling, several ML algorithms have been proposed such as Genetic Programming (GP), Support Vector Regression (SVR), Ensemble Trees (EN), Artificial Neural Network (ANN), and Gaussian Process Regression (GPR) [112, 126, 160, 161]. The concept of developing a model by each algorithm is different. For example, ANN is motivated from the neural system of the body of animals while GP is inspired and computerised based on the theory of evolution. ML algorithms are data-driven and self-adaptive which means that their parameters are iteratively tuned during a training process on a given dataset until the required fitness is obtained. They can learn from data to generate high accuracy models particularly when a large number of variables, irregular data and nonlinear relationships exist, each of which makes the problem sophisticated to analyse using other approaches.

Mixture design of composite materials such as SCC using mathematical optimisation methods which can cumulatively provide the designer with better ideas about the optimal mix design which results in reducing the number of trial tests [19]. More importantly, they are useful to consider multiple objectives and constraints where experimental design approaches are costly or impossible [19, 115, 162]. In a mathematical optimisation procedure, the requirements of the project such as production cost, mechanical and rheological properties, or environmental concerns can be considered as objectives or constraints. Such requirements can be modelled using regression and ML methods based on the considered variables, factors and experimental data. Finally, an appropriate optimisation algorithm can be used to discover the optimum values of design variables from a pool of solution, once one or more objectives are maximised or minimised and some constraints are also met. In terms of mathematical optimisation, finding optimal solutions, even when dealing with a single objective, would be a difficult task due to the high variability, stochastic, randomness, and high nonlinearity of relationships and parameters of problems. Importantly, objectives and constraints are commonly conflicting when it comes to multi-objective multi-constraint optimisation. In this regard, metaheuristic algorithms such as simulated annealing (SA), particle swarm optimisation (PSO), and genetic algorithm (GA) have become valuable choices for solving complex optimisation problems. Such nature-based algorithms are designed to generate and search a space of solutions to find the optimal one iteratively rather than using a deterministic approach [163, 164]. The mixture design of composite materials and cementitious mixtures such as SCC can be usefully converted to a mathematical optimisation problem which has

been demonstrated by researchers through various methods [19, 20, 154, 165]. It is noteworthy that any relationship or issue can be considered in this procedure which makes it flexible and efficient. However, reviewing the literature reveals that several gaps and defects still exist which are open questions to be addressed in to develop a comprehensive mathematical optimisation framework for MOMDO of composite materials.

The most important step is to correctly and accurately convert the project description to a mathematical optimisation form [166]. In terms of SCC experimental mixture design, the designer commonly mixes different amounts or proportions of materials and prepare some specimens to measure the properties such as UCS or slump based on specific methods and conditions. For this scenario, it can be assumed that all factors of the system are constant except for the amount or proportions of materials. To convert this scenario into a mathematical optimisation problem, the controllable parameters or variables are only the proportions of materials. Therefore, only the amount of materials are the input or independent variables of developing models to use as objectives or constraints such as strength, slump or cost. Correspondingly, the final outputs of the mathematical optimisation framework and model variables are the amount or proportions of constituents, which meet the considered objectives and requirements. This scenario has been represented by many researchers for mixture design of composite materials such as concrete using different mathematical optimisation frameworks [20, 107, 152, 154, 162]. How about a more comprehensive insight?

When dealing with different types of materials and experimental methods or conditions, there are other influential factors in addition to the material fractions in a mixture. Available experimental datasets commonly include results of different tests such as UCS conducted on different specimens with different materials. Assume that an experimental dataset contains UCS test results conducted on 10 mm cubic specimens of SCC with Portland cement type I, while another dataset contains results of 15 mm cubic specimens with Portland cement type V. It is incorrect to combine these two datasets and develop predictive UCS models based on the amount of cement only without considering the influence of cement type and specimen size. For this simple case, chemical and physical properties of cement can be considered as influential factors to be added to the model as variables. This way enables to reflect the type of cement. Similarly, a new variable can be considered for specimen size effect. Accordingly, it can be proposed that more comprehensive models can be developed by collecting different experimental datasets containing different types of materials and methods. Although some factors can be considered as systematic error, a comprehensive model must be based on all possible factors as variables. In fact, such models can be used to develop a more comprehensive framework to use worldwide for MOMDO of composite materials. This is the key idea of this paper where several questions still remain open, and the literature fails to represent a correct feasibility.

This paper proposes the feasibility of developing a comprehensive platform for MOMDO of composite materials and cementitious mixtures such as SCC. To do so, a dataset consisting of various variables

such as chemical and physical features of constituents and their proportions in different SCC mixtures and corresponding experimental results and conditions is compiled through a precise survey of the literature. During the data collection, several uncertainties and error might exist, which should be decided to ignore, remove or consider as a new factor. To reduce the uncertainties, data collection is narrowed to these experimental results on which slump flow tests were conducted based on EFNARC and UCS test on cubic specimens of different SCC mixtures. It is also strived to have a worthy number of data for modeling. In order to have dimensionless variables and reduce the dimensionality, some indices and ratios are also considered. The applicability and performance of various regression and ML algorithms are examined to develop models of expected objectives or constraints. The best models are chosen to use after a process of validation. The reason is that the prediction performance of objective models are of paramount significance. Different scenarios considered and the assumed mixture as design projects are converted to mathematical optimisation problems. A variant of metaheuristic algorithms, namely non-dominated sorting genetic algorithm called NSGA-II is used to find the optimal mix designs of SCC meeting the project requirements. The results confirm that the proposed procedure is robust and the fact that it can be used as a universal tool or online platform for multiple objective mixture design and optimisation of cementitious mixtures like SCC. In fact, use of this method reduces the costs and efforts of experimental mixture design of widely-used cementitious mixtures to find optimal and multi-objective designed mixtures which leads to a more sustainable industry.

6-3 **Conceptual framework**

A general flowchart of the conceptual framework is illustrated in Fig. 6.4.

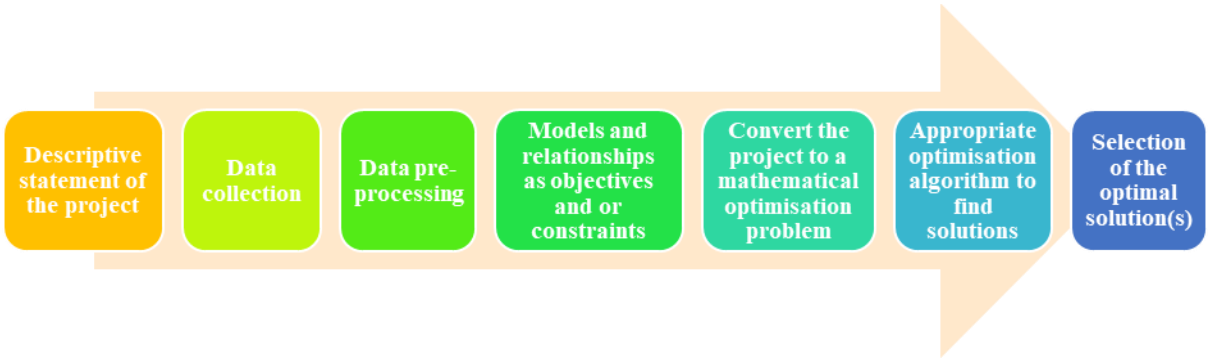


Fig. 6.4 A flowchart of conceptual framework for mathematical optimisation and mixture design of SCC.

A typical constrained multi-objective optimisation problem can be considered as follows:

$$\begin{aligned} &\text{Objective(s): Maximise or Minimise } f_m(X) && m = 1, 2, 3, \ldots, M \\ &\text{Constrained to: } a_i \le x_i \le b_i && i = 1, 2, 3, \ldots, n \\ &h_k(X) \le 0 && k = 1,2, \ldots, K \\ &g_j(X) = 0 && j = 1,2, \ldots, J \end{aligned} \tag{6-1}$$

where a solution $X \in R^n$ is a vector of variables $X = [x_1, x_2, x_3, \ldots, x_i, \ldots, x_n]c^T$, f_m are the objectives which are functions or mathematical models of X to be optimised, n is the number of design variables, b_i and a_i are upper and lower threshold of x_i, k is considered as the number of inequality and j for equality constraints each of which can be considered as functions of X or relationships between a number of x_i considered by g_j and h_k respectively.

Any design optimisation method aims to convert the design problem into a mathematical optimisation problem to find and select the optimal design among many alternatives mathematically. Descriptive statement of the project is the most important step in any problem as well as mathematical optimisation procedure. An appropriate statement describes required factors, design variables, objectives and constraints of the project which must be met. Generally, in a mathematical optimisation procedure, objectives and constraints can be models or relationships between variables based on the considered scenario. In terms of mixture design of SCC, objectives and constraints (f_m, g_j and h_k) can be mechanical, rheological and durability properties, production costs, environmental issues or other existing relationships between variables such as those will be used and discussed in this paper. Constraints can also be any combination of design variables expressed as equalities or inequalities that must be satisfied. Models here are developed using various regression, ML and AI algorithms. If one objective, i.e. m equals 1, only one f(X) is optimised, the problem is called single objective optimisation (SOO). When m>1 and multiple objectives exist, i.e. $f_1(X)$, $f_2(X)$, …, $f_M(X)$, the problem is named multi-objective optimisation (MOO). Optimisation algorithms search a pool of solutions, which can meet multiple objectives and constraints with a level of accuracy without being dominated by any other solution [129, 162, 163]. By means of constraints, the pool of solutions is limited to those meeting the constraints. Thus, the algorithm is focused to the fittest solutions and results faster and more accurate. Based on the project priorities, this procedure enables the designer to pick the optimal without the need for doing new tests.

6-4 **Data collection and variables**

A dataset consisting of 326 test results conducted various SCC specimens with different mixture proportion and type of constituents is collected through a careful survey of the literature. The database includes data reported by [142-144, 146, 147, 154, 167, 168]. The dataset contains slump flow, L-box, segregation resistance, V-funnel and UCS_{28d} test results. In order to separate materials in terms of type, reported features of materials are considered as variables, i.e. main oxides in chemical composition of cementitious material (SiO_2, CaO, Fe_2O_3, Al_2O_3 and MgO), physical properties of constituents such as relative specific gravity (RSG), maximum size of coarse aggregates (D_{max}), water absorption (WA) of aggregates. The dataset contains different types of Portland cement (C), fly ash type F (FAF), GGBS, SF, MK and LP, water (W), fine aggregates (FAgg) and coarse aggregates (CAgg) with different types such as natural or crushed aggregates, and various superplasticisers (SP).

In the present paper, instead of considering the kg/m^3 variables, the ratios of the weight of i^{th} constituent to the weight of binder (B) ratio (Wt_i/Wt_B) are proposed as model variables to reflect the influence of material proportion in the mixture. The binder weight is considered as the sum of weights of C, FA, GGBS, SF and MK of SCC mixes as reported. Wt_i/Wt_B such as Wt_W/Wt_B is a dimensionless, more interpretable variable in terms of regression and predictive modeling. This way also converts the experimental design in practice to a mathematical form as cement is replaced with other materials in the binder. The control mix is the one where Wt_B equals Wt_C. Therefore, when cement is replaced with PCR materials a binary, ternary or composite binder is produced. Although LP might be considered to use as a PCR material, it is considered as an admixture to indicate the feasibility of considering different variables in models.

Relative specific gravity (RSG) is a dimensionless variable defined as the ratio of the specific density of constituents to that of water. Considering the available data, the maximum size of fine aggregates was considered to be 4.75 (mm) and the ratio of maximum size of coarse to fine aggregate, namely maximum size index (MSI) is considered as a variable to account for the properties of aggregates and somehow packing density in models. WA of fine and coarse aggregates are also considered as model input variables. In case more information such as PSD or moisture content and other factors were available, more variables could be considered to reflect the characteristics of aggregates. However, such issues are recommended to consider for future works.

LP and W was assumed to be constant in terms of type. Therefore, there is no need to add another variable to separate their type. It must be noted that if the type of LP or water in the dataset is different, some variables must be considered to account for their type. However, such characteristics were not reported in references. Vital information such as chemical composition of the admixtures including SP, wide range of water reducing admixture, and viscosity modifying agents (VMA) are not commonly published in the literature and accordingly in this database. As different types of SP have very different influence, the type of SP was labelled by numbers to reflect in model as a variable to tackle this issue. Note that type of material can be seen as a categorical variable such as SP type in present paper. In terms of modeling, categorical variables can be converted to quantitative form, e.g. by labelling them. In the present paper, it is proposed to consider chemical and physical properties of materials instead of labelling them. This way enables to also consider the effect of physical and chemical properties of materials. However, a large number of data is required for this aim. Self-labelling the material type has some deficiencies. The problem appears when it comes to using the model for prediction for new materials. If a new material exists, it must be labelled with another new number which is not defined for the model. Therefore, the model is not useful and may not recognize the type and relate it to an accurate proportion. To tackle this, a characteristic such as chemical composition of SP might be useful which was not available in references. The same process is done in image recognition datasets using ML techniques but there is a predefined system.

The more the number of influential variables and datasets are compiled the more descriptive and comprehensive the ML model. It is strived to reduce the uncertainties which results in better data analysis conclusions and the ML model. For example, the dataset may contain slump flow test results in which different experimental methods and standards are used such as with different cones having different dimensions. Therefore, such test results are not comparable and suitable for model development unless a variable is defined to correlate their difference. All slump flow tests in this dataset were conducted according to EFNARC [139]. Therefore, all slump flow test results are comparable and suitable for model development. It is indicated that the UCS of cylindrical specimens with different aspect ratio is different from the cubic ones [169]. To reduce the uncertainties, the UCS of cubic specimens are merely considered in this database rather than cylindrical or other shapes as a small number of suitable dataset with desired variables were available in the literature. Smaller cubic specimens commonly give higher UCS due to the Poisson's effect and the collected database contains UCS_{28d} test results conducted on 100 and 150 mm cubic specimens. In order to reflect this in the UCS_{28d} model, the normalised width ratio (NWR) is defined as follows:

$$NWR = \frac{Width\ of\ specimen\ (mm)}{10\ (mm)} \tag{6-2}$$

This paper considers 49 variables for development of models as given in the collected database attached as a supplementary document. The weight of constituents to binder ratio in the mixture including Wt_C/Wt_B, Wt_{FA}/Wt_B, Wt_{GGBS}/Wt_B, Wt_{SF}/Wt_B, Wt_{MK}/Wt_B, Wt_{LP}/Wt_B, Wt_{CAgg}/Wt_B, Wt_{FAgg}/Wt_B, Wt_W/Wt_B, and Wt_{SP}/Wt_B. These 10 variables are considered as controllable parameters of mixture design to prepare mixtures. According to the available data, variables reflected type of constituents include SP type, MSI, WA_{CAgg} and WA_{FAgg} (%), 9 variables for RSG of C, FA, GGBS, SF, MK, LP, CAgg, FAgg and SP, 25 variable to reflect chemical composition of C, FA, GGBS, SF and MK using their SiO_2, Al_2O_3, Fe_2O_3, CaO and MgO content (%). These variables in addition to NWR which is considered to reflect the specimen size effect in UCS_{28d} data and SP type varies from 1 to 9 are considered as uncontrollable or fixed in this paper. The descriptive statistics of variable are given in Table 6.2 and 3. Fig. 6.5 shows a ternary plot of cementitious materials in the collected database. Note that SiO_2, Al_2O_3, Fe_2O_3, CaO and MgO content (%) are considered as separate variables to develop models. This was due to the better performance and predictability of developed models using such separate variables compared to SiO_2, Al_2O_3+Fe_2O_3, CaO+MgO, or other variables after a trial-and-error process of development S_f and UCS_{28d} models. The description of materials, standard methods used for tests, experimental results, specimens, type and proportion of materials and variables and so forth are provided in attached online supplementary document (Supplementary Document 1) for more details and future works.

Table 6.2 Descriptive statistics of the weight of constituents to binder ratio (Wt_i/Wt_B) and RSG_i of i^{th} material (i indicates the material) in collected database.

Var.	Index	C	A	GGBS	SF	MK	LP	CAgg	FAgg	W	SP
Wt_i/Wt_B	**Mean**	0.708	0.147	0.108	0.026	0.011	0.136	1.495	1.888	0.419	0.017
	StD	0.168	0.137	0.201	0.040	0.039	0.134	0.411	0.376	0.057	0.008
	Min.	0	0	0	0	0	0	1	1.245	0.320	0.002
	Max.	1.000	0.600	1.000	0.250	0.300	0.429	2.359	2.389	0.550	0.034
RSG_i	**Mean**	3.077	1.383	0.795	0.861	0.271	1.716	2.678	2.579	1	1.090
	StD	0.187	1.086	1.279	1.079	0.782	1.338	0.159	0.024	0	0.041
	Min.	0	0	0	0	0	0	1.680	2.530	1	1.040
	Max.	3.150	2.380	2.920	2.220	2.580	2.800	2.780	2.633	1	1.220

Table 6.3 Statistical indices of considered variables of properties of aggregates, NWR, Sf and UCS28d of SCC specimens in database.

Item	CAgg		FAgg	NWR	S_f (mm)	UCS_{28d}
	MSI	WA (%)	WA (%)			(MPa)
Mean	2.921	0.940	0.957	1.12	735.37	57.84
StD	0.463	0.366	0.328	0.21	84.28	16.57
Min.	2.632	0.220	0.655	1	400	20.9
Max.	5.263	1.200	2.010	1.5	925	98.6

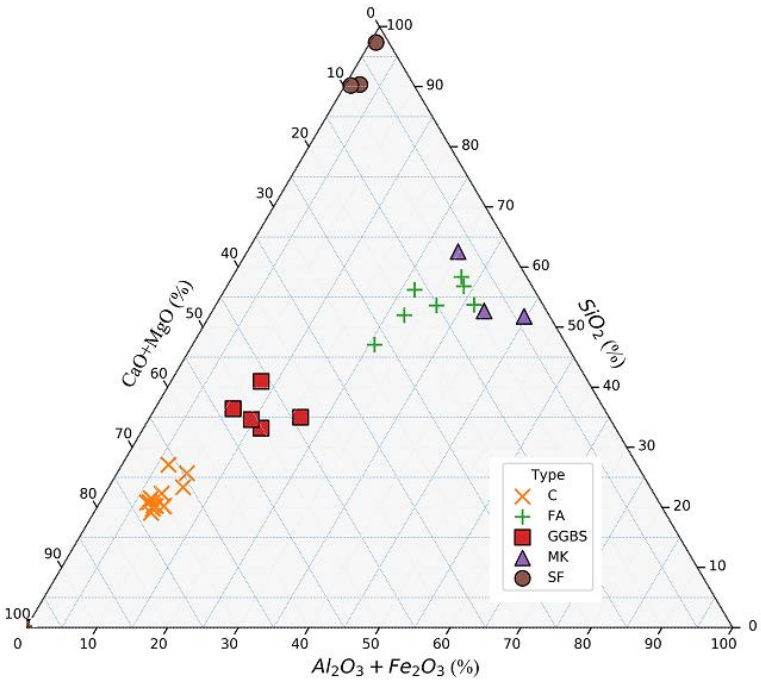


Fig. 6.5 Ternary plot of chemical composition of cementitious materials in the collected database.

As already noted, published datasets often lack uniformity and project information, and specific characteristics or properties of materials are sometimes overlooked which leads to reducing the number of data and variables. Additionally, different papers may use different experimental methods or standards that must be considered during the data collection. There might exist several other factors and uncertainties such as the systematic or random error in experimental results due to method, expertise of the individual or the high variability of properties of materials. Such factors and variabilities must be considered during the data collection to compile reliable database for model development and analyses. This paper tries to mainly point out such issues and agrees that the models here are merely useful for the range and type of materials or methods considered in the collected dataset.

6-5 **Objectives and constraints**

The collected database contains all results of S_f and UCS_{28d} experiments for all 326 mixture designs. However, the experimental results of other properties such as L-Box or J-ring were not fully reported in all references. Therefore, this paper merely considers S_f and UCS_{28d} as objective properties of SCC to represent the applicability of the key idea and objectives of this paper. In order to produce a more sustainable SCC, the production costs and the equivalent CO_2 emission ($CO_{2\text{-}eq}$) of the mixture are important in today's projects. Note that other cost-related issues such as transportation cost, and environmental concerns may also be modelled as objective functions for MOMDO of SCC. This paper merely considers UCS_{28d} and S_f, Production Cost (P_c) and $CO_{2\text{-}eq}$ as objectives or constraints for the optimisation procedure. Note that other properties such as V-funnel, segregation resistance, J-ring, tensile and flexural strength and so forth may be considered as objective functions for mixture design of SCC. However, sufficient data were not available to develop other models of properties as already mentioned.

6-5-1 **UCS_{28d} and S_f models**

Finding high accuracy models of properties of SCC such as UCS_{28d} and S_f is complicated due to the high nonlinearity and large number of influential variables. Here, some robust regression and ML algorithms are used to develop UCS_{28d} and S_f models. The best models are chosen after a process of validation and comparing their performance. Considering the available variables (x_i) in the dataset, UCS_{28d} and S_f are considered as a function of variables as follows:

$$UCS_{28d} = f_u(x_i) \qquad \textit{for i=1 to 49} \tag{6-3}$$

$$S_f = f_s(x_i) \qquad \textit{for i=1 to 48} \tag{6-4}$$

Note that UCS_{28d} has one more variable compared to S_f which is the NWR of cubic specimens. All S_f tests in database were carried out according to EFNRAC. Therefore, there is no need to add more variable to separate the type of test. Although the curing conditions for UCS test results are different, it

is assumed that there is a minor difference in experimental results which can be considered as systematic experimental error. Here, the purpose was to indicate the fact that variables must not be cautiously chosen and not necessarily the same for both models. Different variables can be considered to develop models and accordingly in the mathematical optimisation.

6-5-1-1 Methods

In regression approaches, the structure of the linear or nonlinear model might be clear in advance such as, in the simplest form, multiple linear regression (MLR) model. The coefficients in such structured regression models can be calculated based on analysis of variances using ordinary least squares (OLS) method which minimises the sum of squares of error between predicted and actual values of output. Additionally, several pre-structured models can be used to fit a dataset and the accuracy of each model is different. However, generation of a high-accuracy model using traditional methods such as OLS is a very difficult task due to the nonlinearity of relationships and large number of influential factors, unknown correlations and multicollinearity and other problems. Hence, traditional models such as MLR or polynomial models which are based on OLS and analysis of variances are not useful. To tackle these, ML and AI algorithms have been proposed. ML methods commonly use regularization techniques to add a penalty to model parameters, except for constants or intercepts, to make them robust against problems such as multicollinearity [170]. These algorithms are iteratively improved and programmed in several steps to solve regression problems without prior assumption or detailed information. Therefore, these models are more useful for prediction purposes and reduce the chance of overfitting. In present paper, the capability of some robust ML algorithms, i.e. SVR, GPR and EN approaches are examined to develop UCS_{28d} and S_f models. The structure and input-output process in these algorithms are briefly explained here and more details can be found in references cited.

Given a set of data for training, SVR transfers the original space of data to a higher dimensional feature space using a function such as a linear or a kernel function. The optimal model in the feature space is found by minimising the deviations of the dimensional feature space, subject to a set of constraints [171-173]. GPR is a nonparametric Bayesian regression approach which uses a kernel-based probabilistic function. GPR computes the probability distribution over all functions fitting the data used, i.e. training data, instead of merely the coefficients or parameters. GPR takes advantage of kernel functions to a prior on the Gaussian process distribution over the target relationships and uses a training dataset to define a likelihood function. The parameters of model can be found by optimisation of the log likelihood function though Bayesian optimisation and gradient-based optimisation algorithms [43, 121].

Kernelisation of ML algorithms such as SVR or GPR helps to find nonlinear relationships in a higher dimensional feature space. Generally, kernel functions convert inputs in the original dimensional space using a transformation function into a higher dimensional space in which they become separable. In a higher dimensional space, it is more likely to find a hyper-plane that can relate a set of input variables

to the output. Kernel functions are nonparametric or instance-based rather than having some fixed parameters corresponding to the features of their inputs [174]. Given that there is a set of data (x_i, x_j) in, kernel function recalls the weight of x_i and uses it for the prediction of x_j. Some common Kernel functions such as those provided in Table 6.4. Considering the number of variables in database collected in this paper, it can be expected that kernel functions are useful to find dominant nonlinear relationships between a set of variables for prediction of UCS_{28d} and S_f. Therefore, various GPR and SVR models with different kernel functions are examined on the dataset to develop models.

Table 6.4 Common kernel functions commonly used in and SCR GPR models.

Type	Equation	
Squared Exponential (Fine Gaussian)	$k(x_i, x_j) = \sigma_f^2 \, exp\left(-\frac{r^2}{2\sigma_l^2}\right)$	
Rational Quadratic	$k(x_i, x_j) = \sigma_f^2 \, (1 + \frac{r^2}{2\alpha\sigma_l^2})^{-\alpha}$	where,
Matern 5/2	$k(x_i, x_j) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right) \exp\left(\frac{\sqrt{5}r}{\sigma_l}\right)$	$r = \sqrt{(x_i - x_j)^T (x_i - x_j)}$
Exponential	$k(x_i, x_j) = \sigma_f^2 \, exp\left(-\frac{r}{\sigma_l}\right)$	

σ^2_f is the maximum allowable covariance; σ_l is the kernel function's length scale, and α is a non-negative parameter of covariance.

Decision tree (DT) generates predictive models in the form of a tree. After breaking down the data to smaller subsets, a DT is developed for each subset which consists of decision and leaf nodes. A decision node contains a value for a condition that is considered by the decision or root node. A DT model starts at the topmost decision node and grows down to the bottom most from left to right. DTs can handle both categorical and numerical data. However, DT regression models commonly represent high variance and overfitting for new datasets. To tackle this, DTs are combined with other algorithms to reduce the variance and overfitting in predictions. Ensemble methods combine multiple models which are called weak learners to generate a better predictive model compared to when each of constituent model can perform alone. An ensemble tree (ET) is a model composed of a weighted combination of multiple DTs. In present paper, bagged ensemble trees (BaggedET) is used. Bagging generates models which are called weak learners, trains them on boostrapped samples independently and finally averages all predictions made by those weak learners using the weighted average approach to reduce the variance of the final model.

6-5-1-2 Data pre-processing

Data pre-processing is the process that may require data cleaning, scaling, normalization, transformation, and so forth which results in a polished dataset for training the ML and regression models. Creation of new variables such as MSI and NWR as used in this paper can also be considered as a data pre-processing step. In present paper, all influential factors are considered, either as variables

or during the data collection to reduce the uncertainties and error. Therefore, the database is reliable and all variables are used to reflect a factor into the model. In regression analysis, the range of independent variables in the dataset may vary widely which influences the parameters of the models. Additionally, ML techniques assume that variables are normally distributed. However, variables in collected datasets may not follow this rule of thumb. To tackle these, standardisation is a necessary step in data pre-processing to develop ML and regression models. Note that the range of data may only be scaled in data scaling and normalisation while standardisation also tends to transform the distribution of data closer to normal distribution. Here, the variables are standardised using the z-score normalization method.

6-5-1-3 Accuracy indicators

In the present paper, the overall performance of UCS_{28d} and S_f models are examined using four statistical metrics, correlation coefficient (R), mean bias error (MBE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). R indicates how strong the relationship is between the predicted and experimental values. MAE and RMSE represent the average model prediction errors regardless of their direction, i.e. negative or positive error. RMSE is a quadratic scoring rule and therefore it is more sensitive to large errors. If the absolute value is not taken in MAE, the average error becomes the MBE. MBE intends to measure the average model bias. Positive MBE indicates that the predicted values are overestimated in average. Note that MBE should be interpreted carefully as negative and positive errors remove each other. MAPE is another accuracy indicator which measures the relative error as a percentage. Note that each of these indicators can be used as the fitness function in regression and model development using ML and AI algorithms. Here, RMSE was selected as an initial performance indicator and the fitness function to develop models. Note that the model with lower MAE, RMSE and MAPE values and R closer to 1 gives better prediction where MBE indicates the average direction of the model errors [115]. These indicators can be obtained using the equations in Table 6.5.

6-5-1-4 Model validation

A reliable ML model must neither underfit nor overfit. The degree of the reliability is directly associate with the estimation error made by the model. Underfitting occurs when the model has not accurately captured the relationships between input and output variables. Therefore, it performs inaccurate on those data used to develop model. Overfitting occurs when the model fails to perform well on new data that have not been used for model development. In other words, the performance of model drops considerably when it comes to new data. To consider these, the dataset was firstly separated into two subsets of data, 70% i.e. 243 rows of dataset to develop or train the ML models, namely Train Data, and 83 rows, i.e. 30% remainder, for testing the predictability on unseen data, i.e. Test Data. Train and Test data are determined and labelled in a column in the supplementary document attached. ML models

must perform well on both Train and Test Data to state that they neither underfit nor overfit respectively. In order to reduce the chance of overfitting, a k-fold cross validation is also useful. In this process, the training data, i.e. Train Data here, is initially divided to k equal-sized sub-datasets and the algorithm trains the model using k-1 folds of data while holding out one fold in each iteration, named as validation fold. The parameters and hyper-parameters of the model are updated after each iteration. This process is reiterated for each fold. The k-fold validation method enables updating the parameters of the model k times on different k folds of data which results in reducing overfitting. In this paper, the best k was obtained after a process of trial and error to be 5 after running the algorithms for different k values.

Table 6.5 Accuracy indicators used.

$$R = \frac{\sum_{i=1}^{n}(E_i - \bar{E})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(E_i - \bar{E})^2}\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|E_i - P_i|$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(E_i - P_i)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_i - P_i)^2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{E_i - P_i}{E_i}\right| \times 100$$

n is the number of data, E_i and P_i are respectively the experimental and predicted value of the i^{th} data point.

6-5-1-5 Performance assessment of ML models

Several GPR, SVR and BaggedET models are developed after a trial-and-error process. Many parameters should be tuned in the model including weights or coefficients of input variables and hyper-parameters such as learning rate, regularisation parameter in these algorithms or even the number of k in k-fold cross-validation. The initial values of hyper-parameters are commonly set by the user and best values are obtained using a trial-and-error process or they can be obtained using an optimisation algorithm within the algorithm. For instance, optimal number of leaf nodes and weak learners in BaggaedET are found to be 5 and 30 respectively after a process of trial and error. Considering the number of variables and parameters of algorithms, the generated models are very complicated to interpret or analyse. Such ML models are commonly used as black-box models. They can be recalled to use or retrain by a computer rather than representing by an equation or mathematical expression. Note that the major objective is to find a high accuracy model to use as objective functions where ML

algorithms are beneficial. It is assumed that a model with higher accuracy has better found the existing relationships and therefore is more useful to use as an objective function. The statistical indictors in Table 6.6 are used to evaluate the overall performance of different UCS_{28d} and S_f models on Train and Test datasets for evaluation of the performance of models.

Table 6.6 Accuracy indicators for performance assessment of UCS_{28d} models on Train and Test Data.

Item		UCS_{28d} (MPa)			S_f (mm)		
Index	Data type	EGPPR	BaggedET	FGSVR	EGPPR	BaggedET	FGSVR
R	Train Data	0.999	0.94	0.98	0.999	0.89	0.98
MAE		0.416	4.57	2.29	1.613	30.34	12.74
MBE		0.000	0.00	0.13	-0.017	0.36	0.57
RMSE		0.651	5.92	3.31	2.842	44.00	19.72
MAPE		0.827	8.76	4.35	0.232	4.35	1.81
R	Test Data	0.974	0.95	0.91	0.931	0.75	0.93
MAE		3.006	4.70	4.98	18.716	30.52	19.13
MBE		0.152	0.41	-0.68	0.997	-4.91	1.85
RMSE		3.961	5.80	7.44	25.141	45.57	27.05
MAPE		5.526	9.60	9.02	2.527	4.11	2.57

According to initial accuracy assessment results in Table 6.6, GPR models with exponential kernel function (EGPR) outperformed SVR and BaggedET on both Train and Test data. Note that all models are 5-fold cross-validated on Train data. For UCS_{28d}, EGPR ranked first followed by fine Gaussian kernel-based SVR (FGSVR), and BaggedET respectively. EGPR S_f model also performs better than other models to estimate S_f. Interestingly, the performance of BaggedET trained UCS_{28d} model became better on Test Data. However, when it comes to S_f, these models overfits on Test Data. This shows the deficiency of this model. According to these results EGPR models are of high accuracy in prediction and therefore acceptable. Fig. 6.6 (a) and (b) represent the scatter plots of predicted values of UCS_{28d} and S_f found by EGPR versus experimental observations in order to examine its performance on Train and Test Data. It can be seen that EGPR models can estimate the UCS_{28d} and S_f of SCC with less than 10% error for most of mixture designs considering the type of constituents.

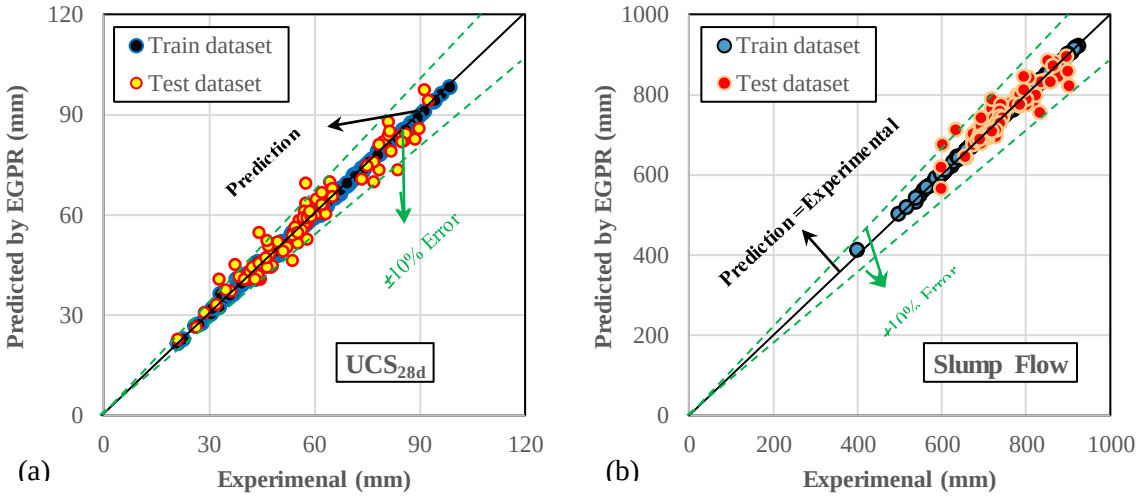


Fig. 6.6 Predicted versus experimental values of (a) UCS_{28d} and (b) S_f found by EGPR for different SCC mixture designs considering the type of constituents.

In order to have a more detailed observation of error, the residual error (RE=Predicted-Experimental) or bias made by best model by each approach for Test Data are represented in Fig. 6.7 (a) and (b). These figures indicate the prediction error amount and range produced by different models.

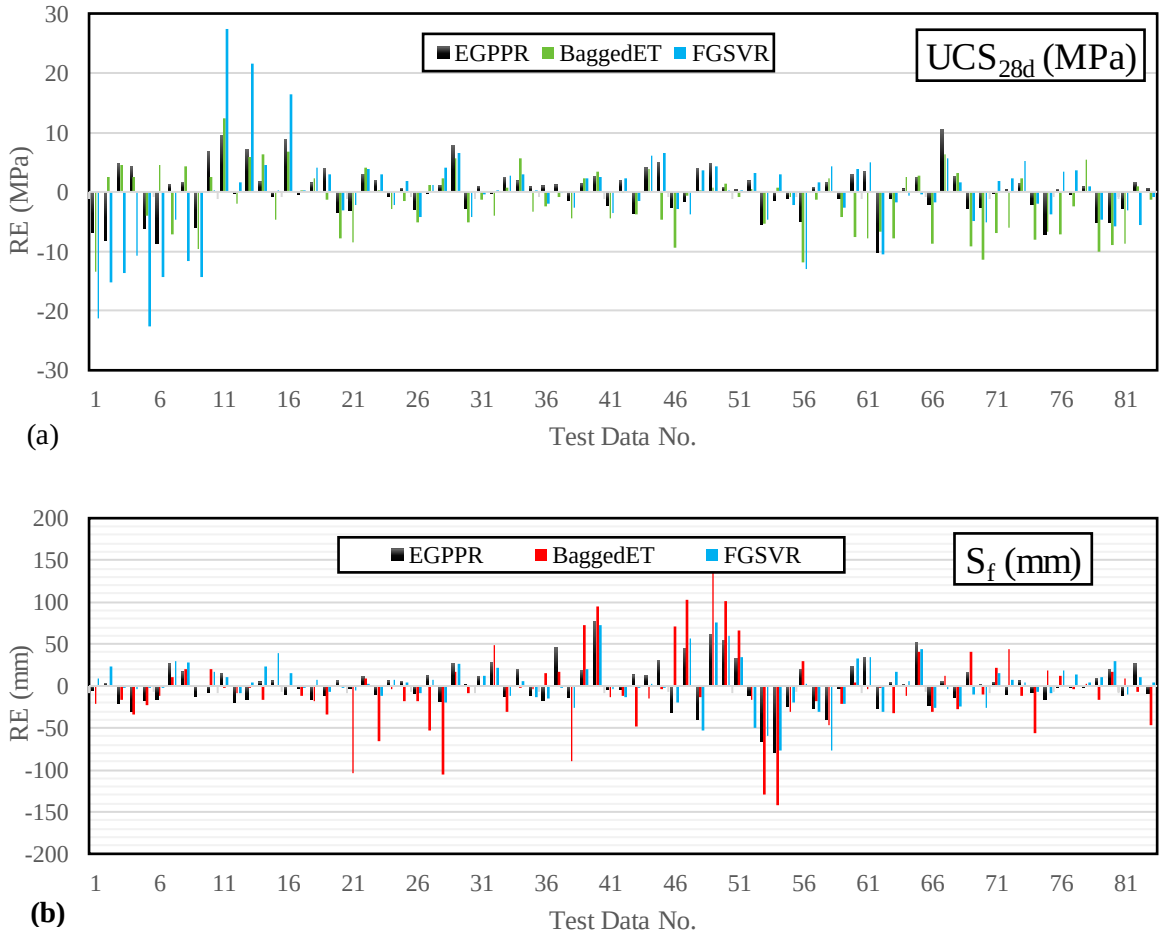


Fig. 6.7 RE values made by different models on Test Data for prediction of (a) UCS_{28d} (MPa) and (b) S_f of SCC.

According to Fig 6.7, it can firstly be realised that EGPR model performs better than other models both for UCS_{28d} and S_f prediction. More importantly, these figures give an idea of the range of prediction errors for different mixture designs and types of material provided by researchers which must be considered when using the model. For example, it can be seen that UCS_{28d} model found by EGPR works well for Test data range between 16 – 41 which are those experimental data provided by [142, 154], where the prediction error is almost less than 4 MPa. However, when it comes to Test data No. 1-16 provided by [143], the model has higher error. The performance of models is different for S_f. It can be seen that EGPR S_f model gives an error almost less than 10 mm for mixture design considered as Test data, 1 to 41, provided by [142, 143, 154], while it performs less accurate for mixture design provided by [146].

The prediction error on Test or Train data might be caused by the complicated relationships between variables, outliers or missing data, accuracy of data, systematic experimental error, other missed influential variables and so forth. Nonetheless, the performance of EGPR models for prediction of S_f and UCS_{28d} of SCC specimens was better than other models and acceptable. In addition to the fact that these results indicate the robustness of GPR, these models are used as objectives or constraints in the proposed mathematical optimisation procedure.

5.2. Production Cost and $CO_{2\text{-}eq}$ models

Development of Production Cost (P_c) and $CO_{2\text{-}eq}$ models is straightforward using conventional multiple linear regression (MLR) method. Here, the production cost is only assumed to be sum of the price of all constituents of the mix and $CO_{2\text{-}eq}$ is the sum of the equivalent CO_2 emissions of the constituents of SCC. The proportion variables considered here are the weight of i^{th} constituent (Wt_i) divided by the unit weight of binder (Wt_B) which can be kg or kg/m^3. Importantly, the proportion variables must be the same in all models used as objective functions for mathematical optimisation. Therefore, the P_c and $CO_{2\text{-}eq}$ can be considered as follows:

$$Production\ Cost\ \left(\frac{\$}{unit\ weight\ of\ binder}\right) = \sum_{i=1}^{n} \frac{Wt_i}{Wt_B} \times P_i \tag{6-5}$$

$$CO_{2-eq}\ \left(\frac{\text{kg}\ CO_{2-eq}}{unit\ weight\ of\ binder}\right) = \sum_{i=1}^{n} \frac{Wt_i}{Wt_B} \times (CO_{2-eq})_i \tag{6-6}$$

where, W_t is the weight (kg), and P_i and $CO_{2\text{-}eq}$ are respectively the price ($/kg) and $(CO_{2\text{-}eq})_i$ (kg $CO_{2\text{-}eq}$/kg) of the i^{th} material in SCC mix and n is the number of materials in the mix which is equal to 10 here. The price and $CO_{2\text{-}eq}$ of each material is represented in the Table 6.7. These values are taken from [162, 175] or provided values by sellers and assumed here to merely develop the P_c and $CO_{2\text{-}eq}$ objective functions. In terms of costs, real value for each material type must be considered. Life cycle analysis (LCA) must also be done to calculate the real $CO_{2\text{-}eq}$ per kg or other environmental factors of the materials used in the mix.

Table 6.7 The assumed unit price and $CO_{2\text{-eq}}$ of the SCC constituents.

Component	Binder						W	CAgg	FAgg	SP
	C	FA	GGBS	SF	MK	LP				
Price ($/kg)	0.1	0.03	0.05	0.5	0.6	0.2	0.001	0.007	0.008	0.4
$Co_{2\text{-eq}}$ (kg $Co_{2\text{-eq}}$/kg)	0.84	0.21	0.13	0.3	0.6	0.2	0.0004	0.008	0.003	0.51

6-5-2 **Constraints**

The constraints might be any linear or nonlinear, inequality or equality relationships between variables and or the range, i.e. lower and upper bound, of variables. In the proposed mathematical optimisation procedure, each model can be used as an equality or inequality constraint based on the project statement. For example, the EGPR model of UCS_{28d} can be used either as objective to be maximised or as a constraint such as UCS_{28d}>45 MPa, S_f=700 mm and or Cost< 50 and so forth. The optimisation algorithm narrows down the space of solutions meeting these constraints while optimising some objectives. Additionally, the range of variables can be constrained to a lower and upper bound or can be fixed at a specific amount when aiming at optimising for a specific material. Most importantly, when the material is not available in the mixture all related variables, both type and proportion, must be constrained to zero. For example, in a ternary binder consisting of C, FA and SF, all related variables to GGBS, MK and LP must be constrained to zero. When dealing with a specific material, all type-related variables such as CaO and RSG must be considered constant at their specific values. Henceforth, the designer may change the lower and upper bounds for proportion-related variables to find the optimal mix. These are important issues of converting the descriptive statement to the mathematical optimisation problem. This paper mainly assumes that the designer tends to find the optimal proportions of some fixed types of material. If a more comprehensive database is available, other scenarios may also be considered. Furthermore, some other relationships may be considered as constraints based on the designer's decision (e.g. $Wt_C/Wt_B+Wt_{MK}/Wt_B$>0.5) which shows the flexibility of the approach. The following relationship is considered here as a constraint which reflects the replacement of cement with SCMs in the mixture design:

$$\frac{Wt_c}{Wt_B}+\frac{Wt_A}{Wt_B}+\frac{Wt_S}{Wt_B}+\frac{Wt_{SF}}{Wt_B}+\frac{Wt_{MK}}{Wt_B}+\frac{Wt_{LP}}{Wt_B}=1 \tag{6-7}$$

6-6 **MOMDO of sustainable SCC**

As already noted, metaheuristic algorithms such as genetic algorithm (GAs) can be applied to convert the mixture design of SCC project to a mathematical optimisation problem once required models and relationships are established to use as objectives or constraints. Note that UCS_{28d} and S_f models considered all available proportions and type related variables. Variables representing the proportion of

mixture constituents such as Wt_C/Wt_B and Wt_{FAgg}/Wt_B are assumed to be the controllable parameters and those related to the type of material such as SiO_2, RSG and WA_{FAgg} and other factors such as NWR can be considered fixed at constant values when dealing with specific types of materials or other factors. Accordingly, the final solutions or outputs of the mathematical optimisation procedure are the proportion variables such as Wt_C/Wt_B. In this regard, various scenarios are considered to represent the applicability and flexibility of the proposed approach.

6-6-1 **Metaheuristic optimisation using NSGA-II**

GA was firstly suggested by Holland with the main idea of ‘‘survival of the fittest’’ and theory of evolution and was then popularized [129, 176]. In GA, a solution vector of X is named “chromosome” consisting of genes which are binary digits. A chromosome is a unique solution which requires a mapping mechanism known as encoding. The process of finding optimal solutions in GA starts with searching the space with an initial population of approximate solutions or chromosomes created randomly within a range, i.e. lower and upper bound on each variable. The initialization procedure involves a random creation of solutions. After evaluating the rank of each solution, the solutions are ranked and a tournament begins [176, 177]. Two solutions are selected as parents randomly or based on their ranks. In evolutionary optimisation algorithms, new population of solutions are generated using genetic and stochastic operators such as selection, crossover, mutation and elite-preservation and some other stochastic selection operators existing in the literature [178]. Crossover, also called recombination, combines the genetic information of two parents by exchanging different parts of them to generate new offspring or solutions as is represented in Fig. 6.8. Crossover uses a preservation strategy that increases the number of solutions to find better ones by making a competition amongst the elitist solutions in the newly-formed generation. Mutation is another useful genetic operator which alters one or more part of a solution to generate a new one. Such operators try to generate new solutions and grow the space of solutions to avoid trapping in a local optima [163]. This iterative process continues until the termination criteria are met which can be a time limit, the fitness, or any other user-defined criteria.

Typically, SOO problems are about satisfying merely one objective. In terms of multi-objective optimisation, the optimum solution cannot only satisfy one objective, as there are more objectives to be met. Thus, only one solution that meets one objective may result in ignorance or sacrifice of other objectives. This trade-off makes results in finding a set of different non-dominated solutions, so-called Pareto-optimal solutions, prior to final decision-making based on priorities. In the present paper, the controlled elitist GA, a variant of non-dominated sorting GA known as NSGA-II is used for both MOO and SOO scenarios [178, 179]. An elitist GA always select the individuals with higher rank and better fitness, while the controlled elitist GA also considers the solutions increasing the population diversity even with a lower fitness value to better search the solution space. The crowding distance in the NSGA-II is based on the cardinality of the solution sets and their distance to the solution boundaries or

constraints. Another feature of this algorithm is that constraints can also be met in the design procedure. This algorithm mainly tries to search Pareto-optimal solutions through an elitist principle and a clear diversity preservative mechanism while giving emphasis to non-dominated solutions [178, 179]. After sorting the solutions, the set with the highest rank and lower crowding distance are selected as optimal front of solutions. This process is represented in Fig. 6.8.

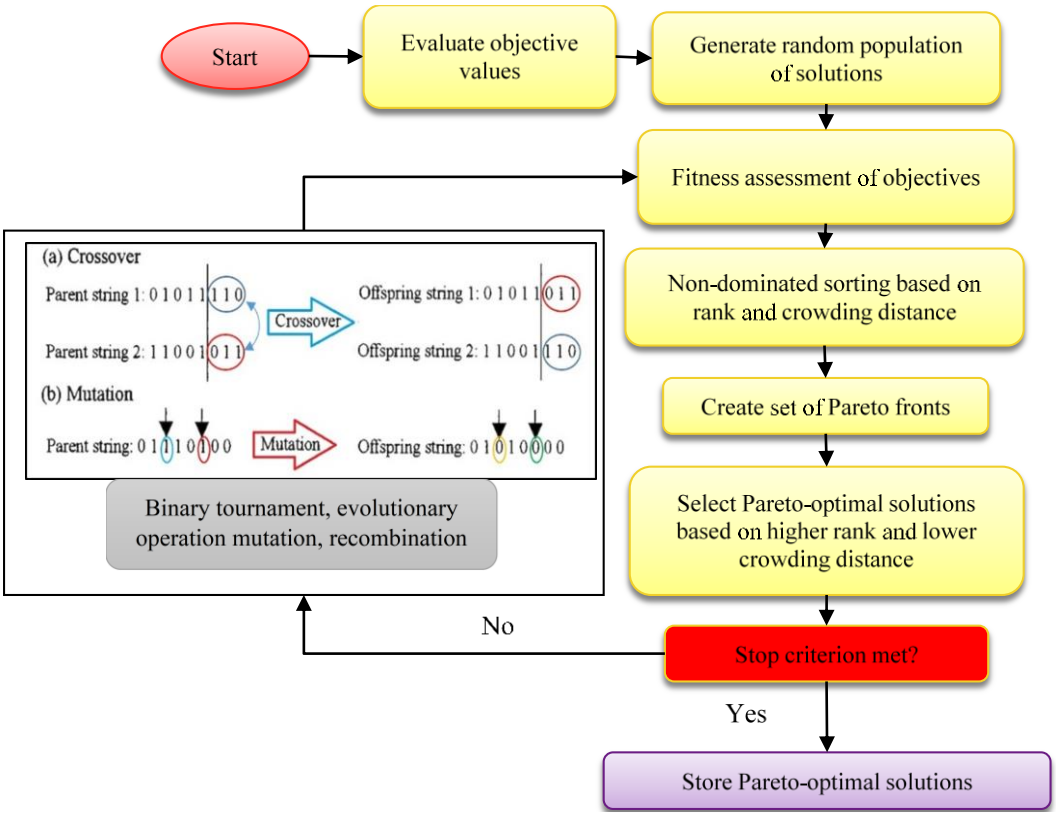


Fig. 6.8 The procedure of finding Pareto-optimal solutions by NSGA-II.

Similar to other AI algorithms, the accuracy of the solutions found by the optimisation algorithm requires adjusting some parameters of the algorithm in advance such as creation, selection, crossover and mutation functions, initial population size, maximum number of generations, and so forth. The more the number of generations and the more the initial population size, the greater the solutions space and the more accurate solution. This way enables the algorithm the chance to better produce and search the space of solutions.

6-6-2 **Mixture design examples and results**

Kannan and Ganesan [144] did some experiments to measure different properties of SCC such as UCS_{28d} and S_f. They designed a mixture composed of only cement based on trial-and-error experiments with the help of specifications, named as the control mix. To optimise that SCC mixture design, they partially replaced cement with FA and or MK and measured the properties and examined the changes. They stated that some mixture designs containing FA and MK had some better properties compared to

the control mix. However, the optimal mixture design may not be found using experiments due to the large number of influential factors and costly experiments, particularly when it comes to meeting multiple objectives. For example, results of Experiment 2 are far from the optimal one in Fig… and more tests need to be carried out to find the optimal mixture design. MOMDO using metaheuristics such as NSGA makes it possible to do several experiments using computers at a much faster pace and find the optimal solutions without the need to do tests. NSGA-II first generates an initial population of solution and finds the optimal solutions as already presented. Finally, the optimal solution which is the optimal mixture design can be selected based on priorities form a set of Pareto-optimal solutions. This comparison is illustrated in Fig…

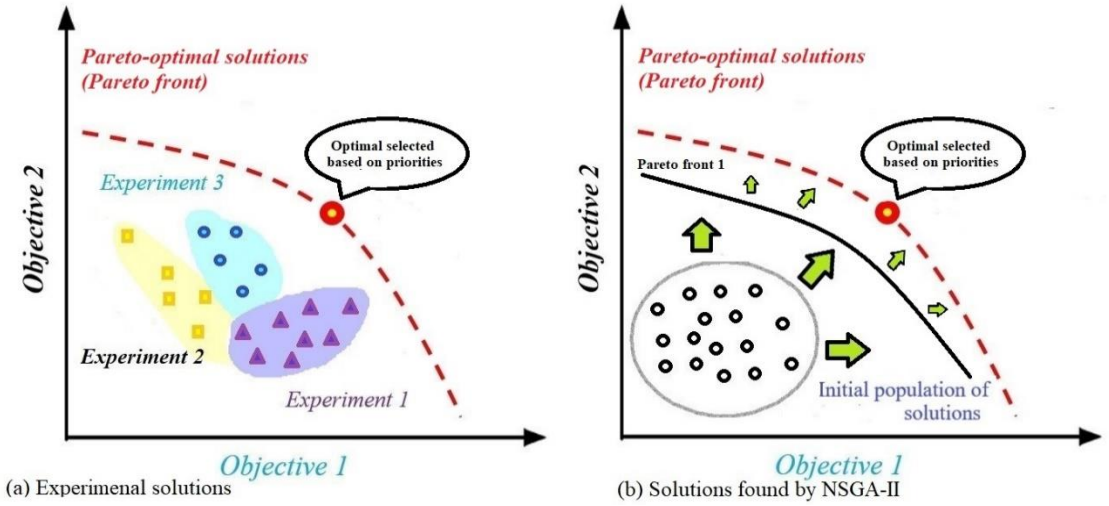


Fig. 6.9 Comparison of how solutions or mixture designs are found by (a) experiments and (b) optimisation algorithm using computers based on considered requirements.

Two scenarios are considered here to represent the applicability of the suggested methodology for MOMDO of SCC mixtures provided by [144]. Both scenarios look for optimal mixture designs based on different statements assuming that the available type of materials are the same as those used by [144]. Those variables representing the type of materials, i.e. physical and chemical properties, are considered as constant variables. Therefore, those variables representing the type of materials are constrained to constant values. The details of experimental data and materials used by [144] to produce SCC specimens and variables considered here are given in attached Supplementary Document 2.

First scenario (Sc#1) looks for the optimal mixture designs of SCC considering two project statements (Sc#1-1 and Sc#1-2) based on the maximum (Max) and minimum (Min) values of proportions in collected dataset. Therefore, the lower bound (LB) and upper bound (UB) of Wt_i/Wt_B in optimisation algorithm are constrained to Max and Min values as given in Table 6.8.

Table 6.8 LB and UB considered for mixture proportions in Sc#1.

Constraints	Controllable/design variables						
	Wt_C/Wt_B	Wt_{FA}/Wt_B	Wt_{MK}/Wt_B	Wt_{CAgg}/Wt_B	Wt_{FAgg}/Wt_B	Wt_W/Wt_B	Wt_{SP}/Wt_B
Min.	0	0	0	1	1.245	0.32	0.002
Max.	1	0.6	0.3	2.359	2.389	0.55	0.034

This scenario is helpful to find the first batch for experimental programs. Obviously, finding the first mixture design is difficult and needs the knowledge and expertise of designer about the properties of materials and their effects, specifications and standards. In addition to considering multiple objectives, the optimal mixture found by the proposed methodology can help the designer to reduce the number of trial tests. The proposed mixture proportions found by the proposed framework are summarised in Table 6.8. Sc#1-1 aims to find the optimal mixture with the Max. UCS_{28d} subject to S_f > 700 mm which can be considered as an SOO problem as there is only one objective, i.e. UCS_{28d}, to be maximised. In this scenario, it is assumed that only cement is available as the binder. In Sc#1-2, an optimal ternary binder consisting of C, FA and MK is looked for and multiple objectives are optimised which leads to many solutions each of which dominates for each objective. For example, the first proposed mixture design in the first row of Sc#1-2 in Table 6.9 has predicted values of UCS_{28d}, S_f, P_c and $CO_{2\text{-}eq}$ of 56.5 (MPa), 687 (mm), 0.25 ($/Unit weight of Binder) and 0.60 (kg CO_2/Unit weight of Binder) respectively, whereas those of second row are 49.6, 714, 0.16 and 0.80. Obviously, the first mixture design offers a higher UCS_{28d} and lower $CO_{2\text{-}eq}$ compared to the second one which might be more preferable. However, the second mixture suggests a lower P_c and higher s_f which might be required in another project. It can be realised that the proposed mixture design method enables the designer to choose the desirable one amongst several solutions.

Table 6.9 Some of The results provided by the proposed procedure for Sc#1.

Scenario	Description / Statement	Other conditions	Optimisation type	Mixture Proportions (Optimisation results)							Predicted values			
				WtC/WtB	WtFA/ WtB	WtMK/ WtB	WtCAgg/ WtB	WtFAgg/ WtB	WtW / WtB	WtSP/WtB	**UCS28d**	**Sf**	**Pc**	**Co2-eq**
Sc#1-1	Max. UCS28d Subject to Sf > 700	Only Cement; SP type 5, LB and UB for SP=0.02.	SOO	1	0	0	1	1.24	0.32	0.02	68.1	710	0.12	0.86
Sc#1-2	Max. UCS28d , Max Sf, Min. Pc and Min. CO2-eq	Ternary binder (C+FA+MK); SP type 5, (LB=0.002 UB=0.03).	MOO	0.4	0.3	0.3	1.34	1.38	0.383	0.024	56.5	687	0.25	0.6
				0.9	0.05	0.05	1.7	2.04	0.498	0.027	53.8	718	0.15	0.83
				0.76	0.12	0.12	1.31	1.47	0.366	0.024	62.3	703	0.18	0.76
				0.95	0.03	0.02	2.04	2.21	0.541	0.022	48.6	721	0.14	0.85
				0.64	0.18	0.18	1.23	1.34	0.359	0.024	61.4	696	0.2	0.71
				0.48	0.26	0.26	1.38	1.4	0.382	0.024	57.6	686	0.24	0.64
				0.77	0.11	0.11	1.24	1.38	0.362	0.023	62.9	705	0.17	0.77
				0.96	0.02	0.02	1.75	2.21	0.532	0.03	52.6	723	0.14	0.86
				0.83	0.08	0.08	1.19	1.34	0.344	0.027	66.2	710	0.16	0.79

The proposed framework also enables the designer to do experiments mathematically. In second scenario (Sc#2) considered here, it is assumed that the designer aims to merely optimise the binder in SCC, similar to what Kanan and Gansen [144] did experimentally. To show this, the LB and UB of the proportions of binder constituents are constrained to Max and Min values considered by [144] as are given in Table 6.10. It can be seen that other variables, i.e. Wt_{CAgg}/ Wt_B, Wt_{FAgg}/ Wt_B, Wt_W/Wt_B, Wt_{SP}/ Wt_B are kept fixed at their considered specific amounts.

Table 6.10 LB and UB considered for mixture proportions in Sc#2.

Constraints	Controllable/Design variables						
	Wt_C/Wt_B	Wt_{FA}/ Wt_B	Wt_{MK}/ Wt_B	Wt_{CAgg}/ Wt_B	Wt_{FAgg}/ Wt_B	Wt_W/ Wt_B	Wt_{SP}/ Wt_B
Min.	0.6	0	0	2	2.2	0.55	0.02
Max.	1	0.3	0.3				

Various project statements are considered for Sc#2 in Table 6.11 to represent the flexibility of the suggested approach for MOMDO of composite materials such as SCC. For validation, the results of the optimisation algorithm in Table 6.10 can be compared to those provided experimentally by [144] which are given in the attached Supplementary Document 2.

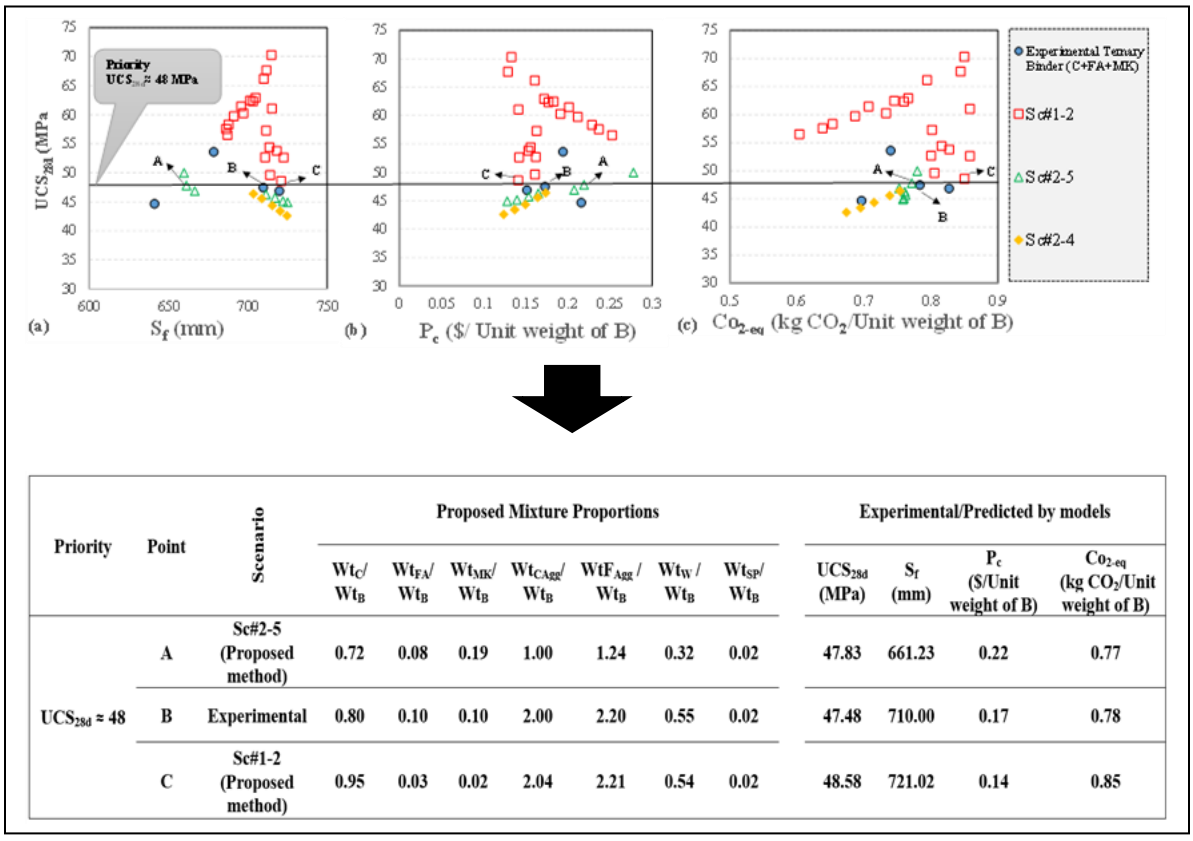
Table 6.11 The results provided by the proposed procedure for Scenario #2

Scenario	Description/ Statement	Other conditions	Type	Mixture Proportions (Optimisation results)							Predicted by models			
				Wt_C/Wt_B	Wt_{F_A}/Wt_B	Wt_{MK}/Wt_B	Wt_{C_Agg}/Wt_B	Wt_{F_Agg}/Wt_B	Wt_W/Wt_B	Wt_{SP}/Wt_B	UCS_{28d}	S_f	P_c	$CO_{2\text{-}eq}$
Sc#2-1	Max. UCS_{28d} Subject to S_f> 700	Binary Binder (C+FA)	SOO	0.85	0.15	0.00					48.2	767.1	0.12	0.78
Sc#2-2	Min. Pc subject to S_f>700 and UCS_{28d} >45	Ternary binder C+FA+MK	SOO	0.81	0.17	0.02					44.9	725.1	0.13	0.76
			SOO	0.87	0.12	0.01					45.9	724.9	0.13	0.79
Sc#2-3	Min. $CO_{2\text{-}eq}$ constrained to Sf>700 and UCS>45	Binary Binder C+FA	SOO	0.81	0.19	0.00					45.0	775.4	0.12	0.75
Sc#2-4	Min. P_c, Max UCS_{28d} subject to Sf> 700	Ternary Binder C+FA+MK	SOO	0.67	0.30	0.03	2	2.2	0.55	0.02	42.6	724.7	0.12	0.67
			SOO	0.71	0.22	0.07					44.4	715.4	0.15	0.72
			SOO	0.73	0.17	0.09					45.6	708.8	0.16	0.74
			SOO	0.69	0.26	0.05					43.4	720.4	0.14	0.69
			SOO	0.75	0.14	0.11					46.4	703.4	0.17	0.75
Sc#2-5	Max. UCS_{28d}, Max. S_f Min. P_c and Min. $CO_{2\text{-}eq}$	Ternary Binder C+FA+MK	MOO	0.67	0.03	0.30					50.0	659.8	0.28	0.78
			MOO	0.81	0.17	0.02					44.9	725.1	0.13	0.76
			MOO	0.72	0.08	0.19					47.8	661.2	0.22	0.77
			MOO	0.77	0.14	0.09					46.3	711.3	0.16	0.76
			MOO	0.71	0.12	0.17					46.9	666.5	0.21	0.75
			MOO	0.79	0.15	0.07					45.7	717.1	0.15	0.76
			MOO	0.80	0.16	0.04					45.1	722.0	0.14	0.76

According to the results in Table 6.11, it can be realised that considering constraints such as S_f>700 or UCS_{28d}> 45 MPa or LB and UB for controllable variables reduces the number of possible solutions. In Sc#1, the algorithm is given the chance to search a wider space of solutions or mixture designs according to the collected database in this paper compared to Sc#2. For example, in Sc#1 the LB and UB for W/B is 0.32 and 0.55, while they are fixed at 0.55 in Sc#2. The designer may use different LB and UB for controllable variables based on the available knowledge and expertise of the designer, recommendations made by various researchers and specifications. In this paper, the same idea is reflected in defining the scenarios.

It is demonstrated that MOO commonly provides the designer with more than one mixture designs while SOO gives merely one solution in each run. When there are several solutions, the designer is still able to narrow down the space and select the optimal mixture design based on significant priorities as is represented in Fig. 6.10. Actually, it is not possible to analyse a four-dimensional space, i.e. UCS_{28d}, S_f, P_c and $CO_{2\text{-}eq}$. To tackle this, scatter plots can be used to depict and easier choose the solutions in

two dimensions of objectives. For example, Fig. 6.10 (a) represents solutions based on different scenarios in a scatter plot of UCS_{28d} versus S_f, while Fig. 6.10 (b) is the scatter plot of solutions for UCS and Pc. It clarifies that there certainly exist mixture designs that may not be easily obtained using experiments, mainly due to the costs of experiments where the proposed method can find easier.



Priority	Point	Scenario	Proposed Mixture Proportions							Experimental/Predicted by models			
			Wt_C/Wt_B	Wt_{FA}/Wt_B	Wt_{MK}/Wt_B	Wt_{CAgg}/Wt_B	WtF_{Agg}/Wt_B	Wt_W/Wt_B	Wt_{SP}/Wt_B	UCS_{28d} (MPa)	S_f (mm)	P_c ($/Unit weight of B)	$Co_{2\text{-}eq}$ (kg CO_2/Unit weight of B)
$UCS_{28d} \approx 48$	A	Sc#2-5 (Proposed method)	0.72	0.08	0.19	1.00	1.24	0.32	0.02	47.83	661.23	0.22	0.77
	B	Experimental	0.80	0.10	0.10	2.00	2.20	0.55	0.02	47.48	710.00	0.17	0.78
	C	Sc#1-2 (Proposed method)	0.95	0.03	0.02	2.04	2.21	0.54	0.02	48.58	721.02	0.14	0.85

Fig. 6.10 The optimal mixture designs considered by the designer based on a priority ($UCS_{28d} \approx 48$ MPa).

In this paper, the mixture proportion variables are considered in form of Wt_i/Wt_B due to the aforementioned reasons. The values of these dimensionless variables can be converted to quantities in unit of kg/m^3 to calculate the required volume of concrete to fill a space. It is noteworthy that the following relationship exist between the constituents of the mixture:

$$\sum_{i=1}^{n} V_i + V_{air} = 1000 \ litre\ (lr) or\ 1\ m^3 \tag{6-8}$$

where n is the number and V_i is the volume of available constituent in the mixture. Considering the variables in this study, the above equation can be converted to the following:

$$\sum_{i=1}^{n} \frac{{Wt_i}/{Wt_B}}{RSG_i} + V_{air} = {1000}/{Wt_B} \tag{6-9}$$

The specific density of water is assumed to be 1000 kg/m^3 and V_{air} can be assumed between 1.5 to 2.5 lr/m^3 of concrete. Finally, Wt_B can be calculated using this equation which is in unit of kg/m^3 and accordingly other Wt_i variables to calculate the required amount of concrete to fill a volume.

6-7 **Conclusion**

The present paper aimed to propose and design a comprehensive AI framework for MOMDO of composite materials such as SCC considering several influential factors and uncertainties such as proportion and properties, experimental methods and conditions. Type of materials and experimental methods can be considered as categorical or qualitative variables which can be converted to quantities such as type of SP. However, considering such variables is problematic as was discussed. It was proposed that quantitative properties of materials can be considered instead of categorical variables to reflect the type of materials. Some indices such as MSI or NWR were also considered to reduce the number of variables. Different variables may also be considered to develop models such as dimensionless variable Wt_c/Wt_B instead of Wt_C (kg/m^3). Anyway, the accuracy of models and relationships, and correct use of them are of priority in a mathematical optimisation framework.

As a mixture design tool, the framework was designed here to find optimal mixture proportions assuming that the designer has some specific types of materials in hand and looks for the best proportions to prepare an optimal mixture meeting different objectives which were considered in different scenarios. In the proposed framework, all influential factors were used to develop models which are then used as objectives or constraints. Variables were then categorised to two controllable and fixed with respect to the mathematical design optimisation concept. The proposed approach was designed to get the available information and factors such as physical and chemical properties of constituents or specimen size as fixed variables, and output the optimal mixture proportions as considered controllable variables. Other uncertainties and differences such as experimental methods were considered during the data collection.

This paper validated that GPR is a useful ML algorithm for predictive modeling of properties of SCC and therefore proposed it for future works. The proposed framework is flexible and any project can be defined given that efficient models are available which is dependent on data and methods. Importantly, the problem must be correctly defined and converted from the start point. Obviously, specifications and standards such as ASTM or ENFRAC are based on experimental data. It can be expected that such specifications can be converted to such tools using artificial intelligence algorithms. In case sufficient data is available, more accurate results can be expected from the proposed method.

Depending on the project, optimisation algorithms such as NSGA-II may find several solutions. The space of solutions can be narrowed using constraints. Therefore, constraints are imperative to find accurate solutions or mixture designs. After finding optimal mixture designs based on the considered objectives and scenarios in a mathematical optimisation framework, the designer can still pick the best amongst Pareto-front solutions based on the priorities and expertise. It is obvious that a more expert designer can get more accurate results.

This way enables to find and use optimum material amounts and resources which leads to a cleaner production and a more sustainable industry. Indeed, the idea of this paper can be extended for MOMDO

of any composite materials and mixtures. Obviously, the greater the number of international experimental data and the richer the provided information, the more accurate and comprehensive the framework's results. Models can be retrained whenever a new dataset is available, and this leads to a more comprehensive model and more accurate optimisation results until even no new experiment is required for all types of materials and mixture designs. The proposed concept and methodology here can be used for all mixture design, optimisation and discovery of composite materials.

Chapter 7 An engineered ML model for prediction of the compressive strength of Eco-SCC based on type and proportions of materials

7-1 Abstract

Recently, various waste materials and industrial by-products such as supplementary cementitious materials (SCMs) have been proposed to improve the properties of self-compacting concrete (SCC). This profitable waste management strategy results in lowering the costs and carbon emission, and a more sustainable, cleaner and eco-friendly production of SCC (Eco-SCC). The properties of such a complex material are commonly measured through costly experiments. Researchers also proposed experimental data analysis and predictive modeling methods such as machine learning (ML) algorithms for prediction of the properties of concrete. However, proposed models commonly relate the properties to the proportion of constituents only and ignore the effect of their type and properties, and other influential factors. This paper aims to engineer the concept and develop a more efficient ML model for prediction of the 28-day uniaxial compressive strength (UCS_{28d}) of SCC containing SCMs. A comprehensive dataset is collected through a precise literature survey. Some dimensionless ratios are proposed to reduce the dimensionality of variables and reflect the effects of considered influential factors in different ML models. Two separate datasets are considered to test the predictability of models where one has new proportions of materials only and the other contains new type of material with new properties. After validation and comparison between various ML models, Gaussian process regression (GPR) model proved to perform well on both considered Test datasets with R^2, RMSE and MAE of around 0.96, 3.66 and 2.49 respectively. Sensitivity analysis results confirm the contribution and importance of considering type and properties of materials as model variables. This paper demonstrates and highlights that all influential factors must be considered to develop engineered ML models to use as universal tools for indirect estimation of properties of composite materials such as Eco-SCC.

7-2 Introduction

Self-Compacting Concrete (SCC), also named as self-consolidating concrete, is a class of concrete that is designed to flow and pass through reinforcement under its own weight without the need for getting externally vibrated after casting [136]. It offers many advantages over conventional concrete such as ease of placement resulting in reducing labour and overall cost, better segregation resistance, producing a denser and more homogenous concrete [136]. SCC is composed of aggregates and a binder paste consisting of water and cementitious material, typically Portland cement. The mixture design and choosing the type of constituents are dependent upon the requirements. Traditional SCC often requires higher amount of cement which increases its production cost and risk of thermal cracks [180]. Cement production is very energy-intensive and leads to emitting a large volume of carbon dioxide (CO_2) that is not environmentally-friendly. CO_2 is a reason of global warming and the reduction of CO_2 emission

has become increasingly crucial nowadays. Researchers have recently proposed partial replacement of cement with some waste material and by-products, known as supplementary cementitious materials (SCMs). This profitable waste management strategy results in lowering the production costs and carbon emission and a more sustainable and eco-friendly SCC (Eco-SCC). SCMs may enhance the properties of a cementitious mixture like SCC through their pozzolanic and or hydraulic activity, or they can be used as filler to provide an improved mixture [181]. SCMs include a wide range of materials and by-products such as ground granulated blast-furnace slag (GGBS or GGBFS), fly ash (FA), meta-kaolin (MK), silica fume (SF) lime powder (LP), and waste glass powder (WGP). FA and GGBS are by-products of coal combustion and steel manufacturing respectively while MK is a product of thermal treatment and calcination of kaolin clay in temperatures between 500°C and 900°C.

Specifications classify cement differently. In terms of compressive strength, EN 197-1 [58] classifies cement into three classes, namely 32.5, 42.5 and 52.5. According to ASTM C150 / C150M-20, cement is classified into ten types including type I, IA, II, IIA, II(MH), II(MH)A, III, IIIA, IV and V with different properties [1]. The chemical composition of cement in this specification must include: aluminum oxide or alumina (AL_2O_3), iron oxide (Fe_2O_3), magnesium oxide (MgO), sulfur trioxide (SO_3), dicalcium silicate (Ca_2SiO_4 or C_2S), tricalcium aluminate ($Ca_3Al_2O_6$ or C_3A), tricalcium silicate (Ca_3SiO_5 or C_3S), and tetracalcium aluminofernite ($4CaO.Al_2O_3Fe_2O_3$ or C_4AF). During the hydration process and reactions between ions in cementitious material, various compounds such as calcium silicate hydrates (CSH), calcium aluminate hydrates (CAH) and calcium aluminium silicate hydrates (CASH) may form and strengthen the final product [182, 183]. Calcium silicate hydrate ($CaO\text{-}SiO_2\text{-}H_2O$ or CSH) which is the product of the reaction of C_3S and C_2S in Portland cement with water is the main compound responsible for gaining strength, hydraulic and self-cementing properties in presence of water [182]. Calcium hydroxide ($Ca(OH)_2$ or CH) is another product of the hydration process; it may crystallize with CSH in an appropriate pH and produce a form of crystals interlocked with other constituents such as aggregates. Although carbonation may heighten the strength of concrete, the pH decreases and causes corrosion in steel reinforcements such as rebars.

SCMs contain different amounts of calcium oxide (CaO), silicon oxide or silica (SiO_2), both amorphous and crystalline, AL_2O_3, Fe_2O_3, MgO and other oxides in form of different minerals or compounds which are responsible for the cementitious properties. SCMs are usually in form of powder and as fine as cement with a considerable specific surface area. Pozzolanic SCMs such as SF, volcanic FA, low-calcium FA and MK mainly contain silica or aluminosilicate (Al_2SiO_5) components. They do not have self-cementing properties per se but they are able to react with CH in presence of water [181]. Therefore, they commonly need a cementing agent such as Portland cement to harden. SCMs such as GGBS or high-calcium fly ash can exhibit hydraulic and self-cementing properties that develop the strength of concrete [184, 185]. The hydration of hydraulic SCMs is different from PC. For example, GGBS may contain minerals which can produce CSH in the presence of water, or consume CH to form additional CSH which is responsible for self-cementing [184, 186]. Pozzolanic SCMs may also react with CH to

reduce carbonation and use it to produce other compounds which can improve the performance of concrete. Some SCMs such as SF may assist in gaining faster strength at early age, while some like GGBS may cause an increase in later ages [185, 187]. Based on the physical and chemical properties and proportion of SCMs in the mixture, strength may increase or decrease [168, 188]. They may also enhance the workability, pumpability, durability of concrete such as SCC [187, 189]. Importantly, by partially replacing cement with one or more SCMs, a binder that is more sustainable and eco-friendlier but with comparable or better properties can be produced as opposed to the mixture containing only cement as the binder [3, 175, 190].

The chemical composition of cementitious materials has the main role in the formation of reaction and properties of concrete, assuming that physical properties such as specific surface area and particle size distribution are appropriate. Reviewing the literature indicates that five oxides of CaO, SiO_2, MgO, Fe_2O_3 and Al_2O_3 are commonly considered by researchers to evaluate the chemical composition of cementitious materials. Researchers and specifications proposed different ratios or conditions based on chemical composition to classify and evaluate cementitious materials used in concrete. Some of the proposed indices and terms for classification of ashes, raw or natural pozzolans and slags as material with pozzolanic and hydraulic properties are summarised in Table 7.1 [58, 184, 191, 192].

Table 7.1 Classification of slags and fly ash and natural pozzolans based on their chemical composition for use in concrete.

Material	Ratio/Index	Requirement		
Coal fly ash and raw or calcined natural pozzolan	$SiO_2 + Al_2O_3 + Fe_2O_3$ (%)	Class N≥70	Class F ≥70	Class C≥50
	Max. Sulfur trioxide (SO_3) (%)	Class N≤4	Class F≤5	Class C≤5
	Max. $Na_2O + 0.658\ K_2O$	Class N≤1.5	Class F≤1.5	Class C≤1.5
	Max. loss on ignition	Class N≤10	Class F≤6	Class C≤6
Hydraulicity of slags	CaO/SiO_2	1.3–1.4		
	$(CaO + MgO)/SiO_2$	>1.0		
	$(CaO + MgO + Al_2O_3)/SiO_2$	>1.0 In Japan, ≥ 1.4 In South Korea, ≥ 1.6		
	$(CaO + 0.56Al_2O_3 + 1.4MgO)/SiO_2$	≥1.65		

The proportion of fine and coarse aggregates in concrete must be designed in a way to reduce voids and achieve the optimal particle packing. Regardless of the mineralogy or morphology of aggregates, SCC mixes containing aggregates with higher packing density show higher UCS, while there is no significant change when aggregates have the same packing density [193]. Several other raw or waste materials and by-products such as recycled glass, limestone powder, marble powder, pumice powder, quartz powder and so forth have been proposed to use as SCMs or as inert filler in Eco-SCC to improve properties such as density and strength [154, 180, 194]. LP can be used as an inert filler to increase the paste content or viscosity of the mixture or it can be used as a replacement of unreacted cement [154, 195, 196]. In general, inert fillers are used to enhance particle packing and fresh state properties with no significant influence on the UCS and hydration process [195, 196]. In this paper, LP is considered to be an inert filler used as an admixture in Eco-SCC mixture.

Moreover, superplasticiser (SP) is a necessary chemical admixture in Eco-SCC to get the required properties, particularly in fresh state such as flowability, viscosity and workability. Polycarboxylate SP is one of the widely used SPs consisting of water and polycarboxylate polymers, mainly polyethylene glycols [197]. They typically benefit the properties of concrete through reducing the water to cement/binder ratio, e.g. by 40%, without negatively affecting the workability but increasing the strength in the same proportion of constituents [198]. Moreover, there are several other admixtures such as those provided in ASTM C494 / C494M-19 [199] that can be used in concrete. According to [200], there are significant variations in the fresh properties of Eco-SCC such as V-funnel, slump flow, U-Box, L-Box and J-ring results [201]. Such variations can be due to many factors such as mixture

proportions, quality of materials but above all is the admixture content and their chemical compositions. Vital information such as chemical composition of the admixtures including SP, wide range of water reducing admixture, and viscosity modifying agents (VMA) are not commonly published in the literature [192, 200]. Fresh state properties of Eco-SCC can be improved using some well-known admixtures such as SP with no significant change of the hardened state properties. But, it would be difficult to do the same for properties at hardened state such as compressive strength as it takes time to prepare specimens and test them.

Compressive strength is the major mechanical property of concrete which reflects its quality in different applications, particularly for structural concrete. Compressive strength of Eco-SCC and other types of concrete is typically measured through lab uniaxial compressive strength (UCS) tests on specimens, typically after curing for 28 days (UCS_{28d}). In order to reduce the testing costs and logistics, researchers proposed predictive modeling, data analysis methods that enable prediction of the properties of concrete such as UCS_{28d}. In this regard, artificial intelligence (AI) and machine learning (ML) approaches have been applied for predictive modeling of different properties of SCC such as [159]. These algorithms are able to find high-accuracy nonlinear models relating the output to input variables in a set of data. Several ML algorithms such as support vector machines (SVMs), artificial neural networks (ANNs), genetic programming (GP), and so forth have been successfully utilized for solving engineering problems [100, 126, 202-204]. Although several efforts have been made to propose efficient models, whether and how complicated properties of concrete, such as the compressive strength can be modelled to use for prediction aims are still open research questions [19, 39].

Researchers have demonstrated the capability of different approaches for predictive modeling of properties of concrete [112, 205-209]. Proposed models often relate the properties of concrete such as UCS to only the amount of constituents of the mixture. This is merely correct in case the only influential parameter is the amount or proportion of materials. However, problem arises when it comes to use such models and make predictions. They cannot differentiate the type of material and the influence of their properties and output the same results e.g. the same UCS_{28d} for an Eco-SCC containing Cement type I and Type V with the same mixture proportions. There are several other uncertainties and influential factors which must be reduced and considered to develop reliable models. When dealing with different types of materials in collected datasets, particularly to develop predictive models, factors such as the type of constituents pertaining to their chemical or physical properties are influential. However, they have been overlooked in the existing literature of AI-based concrete modelling. This paper aims to better engineer the problem and propose a more efficient model which can consider the proportion and type of constituents of Eco-SCC to predict the UCS_{28d} using different ML and regression methods. For this aim, a database consisting of UCS_{28d} test results conducted by different researchers worldwide on cubic SCC specimens with various mixture proportions of different types of materials is collected. Some indices and dimensionless ratios are reasonably considered and proposed to reduce the dimensionality of variables and reflect the effects of considered factors in models. Different ML algorithms are used

and evaluated for predictive modeling of the UCS_{28d} of Eco-SCC. In order to develop reliable ML models, several steps must be implemented and the accuracy model must be validated. The most important step is to verify whether the model can perform accurately when it comes to unseen proportion and type of materials which can be called predictability. After a process of validation, the predictability of the efficient model is confirmed.

7-3 **Method**

Predictive modeling of the properties of concrete is challenging due to the large number of influential factors and nonlinear relationships. In this paper, quadratic regression with interaction terms (QRI), and two variants of powerful ML algorithms namely, Gaussian process regression (GPR) and support vector regression (SVR) algorithms are used to develop UCS_{28d} prediction models.

2.1 Quadratic regression with interaction terms

The general form of the quadratic regression with interaction terms (QRI) model used here is as follows:

$$y = \sum a_i x_i^2 + \sum b_i x_i + \sum c_i x_i x_j + \varepsilon \quad (7\text{-}1)$$

where y is called dependent or model output variable, x_i is the independent or design variables, and a_i, b_i and c_i are coefficients obtained using statistical analysis of variances after fitting the model on data. ε is the constant or bias term of model which is the mean for the response when all of the explanatory variables are kept equal to 0. These coefficients or parameters can be found using analysis of variance (ANOVA) using the ordinary least squares (OLS) approach through minimising the sum of the squares of the error between the measured and estimated value of dependent variable for a given dataset.

2.2 SVR

SVR is a data-driven and supervised ML algorithms which aims to find a model, f(x), with at most ε deviation from the target (y) which is the output variable in the database [171, 172]. Note that support vector machines (SVMs) are typically used for classification aims with support vector methods, and SVR is commonly used for regression. Consider that a linear relationship exists between vectors of $X=\{x_1, x_2, x_3, \ldots, x_i, \ldots, x_n\}$ as input and corresponding $Y=\{y_1, y_2, y_3, \ldots, y_i, \ldots, y_n\}$ as output variables. The general form of the SVR model can be represented as follows:

$$Y = f(X) = W^T X + b \quad (7\text{-}2)$$

W is the vector of coefficients and b is the constant. As there might be a nonlinear relationship, the vector X can be transformed using kernel function φ(X).

Therefore, the above equation can be changed to the following expression:

$$y = f(X) = W^T \varphi(X) + b \quad (7\text{-}3)$$

In SVR, the input space is mapped to a higher infinite dimensional feature space through a nonlinear kernel function, namely φ(X), to find a higher accuracy model. There are various types of kernel functions such as polynomial, Gaussian, and radial basis functions [210]. The general form of polynomial kernel function is as follows:

$$K(W,X) = (1 + W^T X)^d \tag{7-4}$$

where d is equal to 1, 2 and 3, the polynomial kernel functions is called linear, quadratic and cubic kernel function. In simple form, a least squares algorithm is used to minimise the existing error between f(X) and Y. In order to minimise deviations within the dimensional feature space and accordingly overfitting, Lagrange multipliers are used. Finally, the SVR problem is considered as a constrained mathematical optimisation problem using the structural risk minimisation principle. Solving the following risk optimisation problem yield in finding the coefficients w and b [211, 212]:

$$Minimise\ \frac{1}{2}\|W\|^2 + C_p \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{7-5}$$

subject to:

$$\begin{cases} y_i - Wx_i - b \le \varepsilon + \xi_i \\ Wx_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i\ , \xi_i^* \ge 0 \qquad i = 1,2,3,\dots,n \end{cases} \tag{7-6}$$

where $\xi_i + \xi_i^*$ are positive slack variables; ε is the dimension of the insensitive zone; and C_P is a penalty parameter and determines the trade-off for ε.

2.3 GPR

GPR is a nonparametric supervised machine learning algorithm for predictive modeling, regresion and data analysis [43, 121]. Given a linear function as follows:

$$y=wx+\varepsilon. \tag{7-7}$$

this approach specifies a prior distribution, p(w), on the parameters or coefficient matrix, *w*, and apply probabilities depending on data using Bayes' Rule:

$$P(w|y,X) = \frac{p(y|x,w)p(w)}{p(y|x)} \tag{7-8}$$

p(w|y,X) is named posterior distribution which gets data from both prior distributions of w and dataset, i.e. p(y|x,w) and p(y|x). Posterior probability is the conditional probability based on the Bayes theorem [162]. Therefore, GPR is considered as a Bayesian approach. In order to predict new x, i.e. x^*, with f^*

values, the prediction distribution is updated through weighting all possible x values using their obtained posterior distribution as follows:

$$p(f^*|x^*,y,x) = \int_w p(f^*|x^*,w)p(w|y,x)\,dw \tag{7-9}$$

By considering the prior distribution as a Gaussian distribution, a prediction can be made using the mean and variance values. The Gaussian processes term is due to the use of the Gaussian distribution for transforming the multiple-dimensional generalization of multiple-variable normal distributions [121, 162]. GPR calculates the probability distribution over all functions fitting the data used, i.e. training data, instead of merely the coefficients or parameters. This is why this approach is called non-parametric.

Similar to other regression methods, GPR can be used to find models existing between data. Given a dataset $D = \{(x_i, y_i)|\ i=1, \ldots, n\}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}^n$ are the input and output vectors, taken from an unknown distribution. GPR considers y as the following function [43]:

$$y = f(x) + \varepsilon \tag{7-10}$$

where $\varepsilon \sim N(0, \sigma^2_n)$. ε follows the Gaussian distribution with an average value of 0 and error variance of σ^2_n considering the output values or y_i.

If $\{f(x), x \in \mathbb{R}^d\}$ is a Gaussian process, then for a set of i observations $x_1, x_2, \ldots, x_i$, Gaussian transformed function is a combined distribution of variables $f(x_1), f(x_2), \ldots, f(x_i)$. Given two x_i and x_j, a GP can be defined by $m(x)$ and $k(x_i, x_j)$ which are the mean and covariance kernel functions respectively as represented as follows:

$$m(x)=E(f(x)) \tag{7-11}$$

$$k(x_i,x_j) = Cov(f(x_i),f(x_j))=E[\{f(x_i)-m(x_i)\}\{f(x_j)-m(x_j)\} \tag{7-12}$$

where $k(x_i,x_j)$ determines the covariance between x_i and x_j. Therefore:

$$f(x) \sim GP\,(m(x),\ k(x_i,x_j)) \tag{7-13}$$

and accordingly:

$$y \sim GP\,(m(x),\ k(x_i,x_j) + \sigma^2_n\,(x_i-x_j)) \tag{7-14}$$

$m(x)$ is the mean function is normally a constant and can be zero or the average of the training dataset. The general forms of some common kernel functions used in GPR is represented in Table 7.2.

Table 7.2 Common kernel functions commonly used in GPR models.

Type	Equation
Squared Exponential (Gaussian)	$k(x_i, x_j) = \sigma_f^2 \, exp\left(-\frac{r^2}{2\sigma_l^2}\right)$
Rational Quadratic	$k(x_i, x_j) = \sigma_f^2 \, (1 + \frac{r^2}{2\alpha\sigma_l^2})^{-\alpha}$
Matern 5/2	$k(x_i, x_j) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right) \exp\left(\frac{\sqrt{5}r}{\sigma_l}\right)$
Exponential	$k(x_i, x_j) = \sigma_f^2 \, exp\left(-\frac{r}{\sigma_l}\right)$

In the equations listed in Table 7.2, the maximum allowable covariance is considered as σ^2_f, σ_l is the length scale of the kernel function, α is a non-negative parameter of covariance, and r is defined as follows:

$$r = \sqrt{(x_i - x_j)^T (x_i - x_j)} \tag{7-15}$$

The parameters or hyper-parameters of the GPR model, such as σ^2_n, σ_l, α or σ^2_f are updated during the modeling process by Bayesian inference until convergence is reached or a stop criterion is met. These parameters are calculated by optimising the log likelihood function using a Bayesian optimisation and gradient-based optimisation algorithms [122, 123].

7-4 **Dataset and variable selection**

A comprehensive dataset is collected through a careful survey of the literature. The dataset contains test results conducted on cubic specimens of Eco-SCC with different mix designs to measure their UCS_{28d} and a wide range of different parameters of constituents. The database includes 456 data taken from 20 published papers in the literature and is attached as a supplementary document [140, 142-148, 154, 167, 168, 194, 213-220]. The database includes reported proportions of Eco-SCC's constituents, chemical compositions of cementitious material such as main oxides (CaO, SiO_2, Al_2O_3, Fe_2O_3, MgO), some physical properties of materials such as maximum size of coarse aggregates (D_{max}), water absorption (WA) of aggregates as reported by researchers. The description of materials and specimens are also given such as different types of cement (C) such as Portland type I, II or IV, different types of ash (A) such as rice husk ash (RHA), fly ash class F (FAF), pulverised fly ash (PFA), GGBS, SF, MK and LP, water (W), fine aggregates (FAgg) and coarse aggregates (CAgg) with different types such as natural or crushed aggregates, various superplasticisers (SP) and so forth. Finally, there are 38 independent variables in the collected database where the goal is to develop a model for predicting the UCS_{28d}.

The more data and variables, the more comprehensive the model. It was strived to collect as large number of data and influential factors as possible with the least uncertainties to develop ML models.

The UCS_{28d} of cylindrical specimens with different aspect ratio is different from the cubic ones [169]. Therefore, the UCS of cubic specimens are merely considered in this database rather than cylindrical or other shapes. Other factors and experimental procedure issues are compulsorily considered here as systematic experimental error in data. Note that such factors were not reported in some papers and therefore were not possible to be considered.

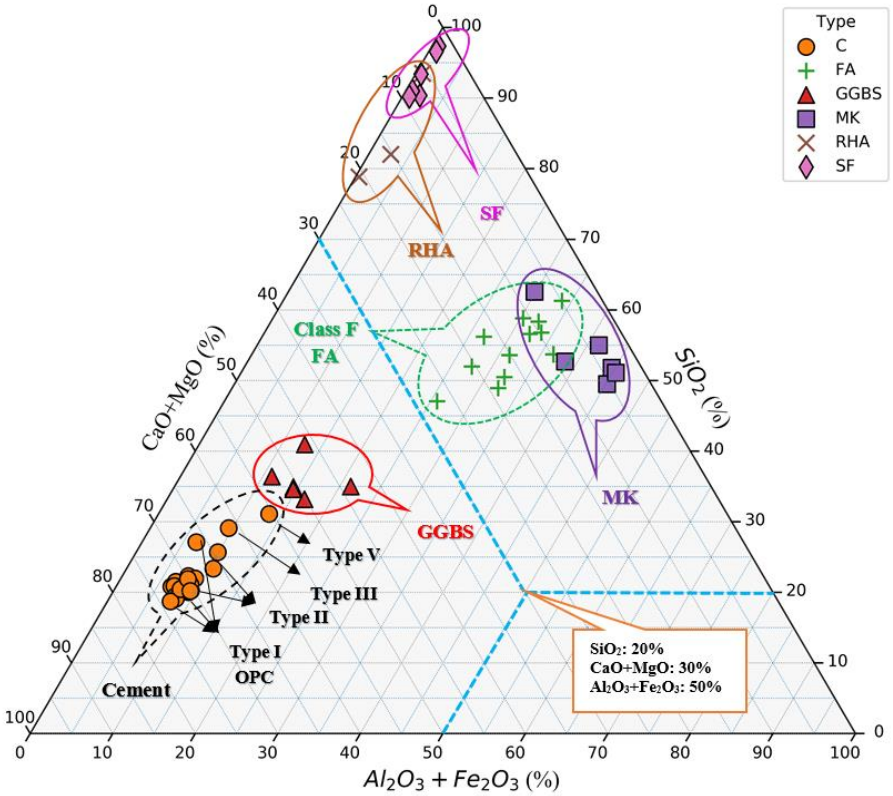


Fig. 7.1 A ternary plot of chemical composition of cementitious material in database.

Fig. 7.1 represents a ternary plot of chemical composition of available types of cement and SCMs in the collected database. It can be seen that SF and RHA have a high content of SiO_2 and MK and FA are low in CaO+MgO. GGBS and Cement are close but cement has certainly a higher content of CaO+MgO. It can be seen that cement types do not have the same chemical composition. Different chemical composition may produce different compounds. Obviously, each type of material with a different chemical composition have a different influence on properties of mixture such as UCS. Therefore, chemical composition of cementitious materials are important and must be considered as variables to develop predictive models which have been ignored.

In terms of modeling, each oxide such as CaO can be considered as an independent variable which leads to 25 variables for 5 types of cementitious materials, i.e. cement, ash, GGBS, SF and MK. It is possible to consider CaO+MgO, SiO_2 and Al_2O_3+Fe_2O_3 yielding in 15 variables. In order to reflect the chemical compositions and reduce the dimensionality, this paper proposes the following pozzolanic and hydraulic reactivity indices, PRI and HRI respectively, can be proposed as were also considered in Table 7.1:

$$PRI = (SiO_2 + Al_2O_3 + Fe_2O_3)/100 \quad (7\text{-}16)$$

$$HRI = (CaO + Al_2O_3 + MgO)/SiO_2 \quad (7\text{-}17)$$

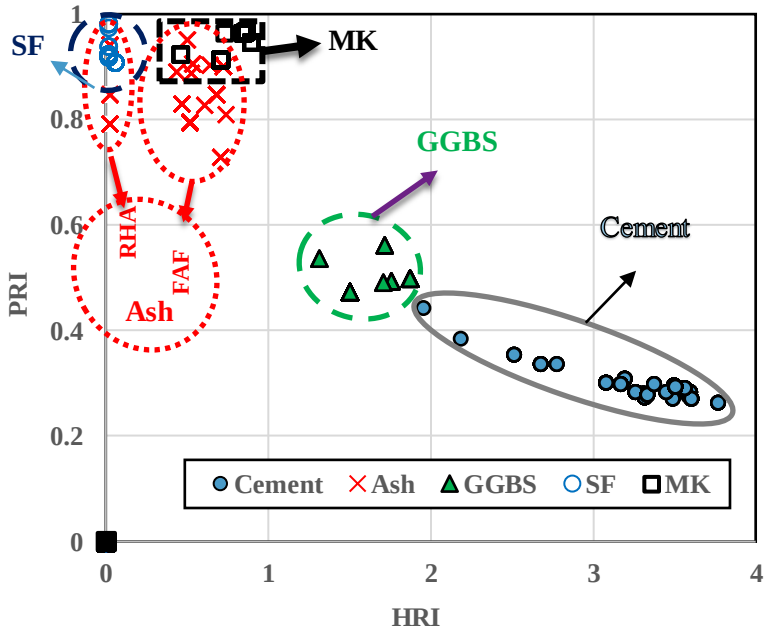


Fig. 7.2 A scatter plot to classify cementitious material in terms of hydraulic or pozzolanic reactivity.

Fig. 7.2 can be used for classification of cementitious materials in terms of pozzolanic or hydraulic reactivity only. HRI and PRI are considered to reflect the type of cementitious material in the database. The reason why PRI is divided by 100 is to scale it within the range of other variables as it affects the regression model performance. Considering HRI and PRI indices instead of chemical composition such as SiO_2 and CaO results in a meaningful dimensionality reduction of model variables, from 25 to 10 variables.

Typically, 4.75 mm is considered to classify aggregates in terms of size. Below this threshold, particles are considered as fine aggregates (FAgg) and above it they are considered as coarse aggregates (CAgg). The maximum size index (MSI) is the ratio of the maximum size of coarse to fine aggregates:

$$MSI = D_{max}/4.75 \quad (7\text{-}18)$$

Additionally, water absorption of aggregates (WA) is considered as a factor reflecting the nature of aggregates in the model. Fig. 7.3 illustrates the PSD of some fine and coarse aggregates used by different researchers in the literature for producing SCC [142, 143, 145-148, 194, 221]. When there is no idea about the optimal PSD of FAgg or CAgg, analysis of data can be helpful such as that provided in Fig. 7.3. Note that guidelines and specifications are also based on experimental data. In the present paper Fig. 7.3 and MSI are proposed to consider optimal particle size distribution of aggregates for mixture design. Packing density and specific gravity of constituents of the mixture are other important factors which can be considered in future studies where relevant data is avilable.

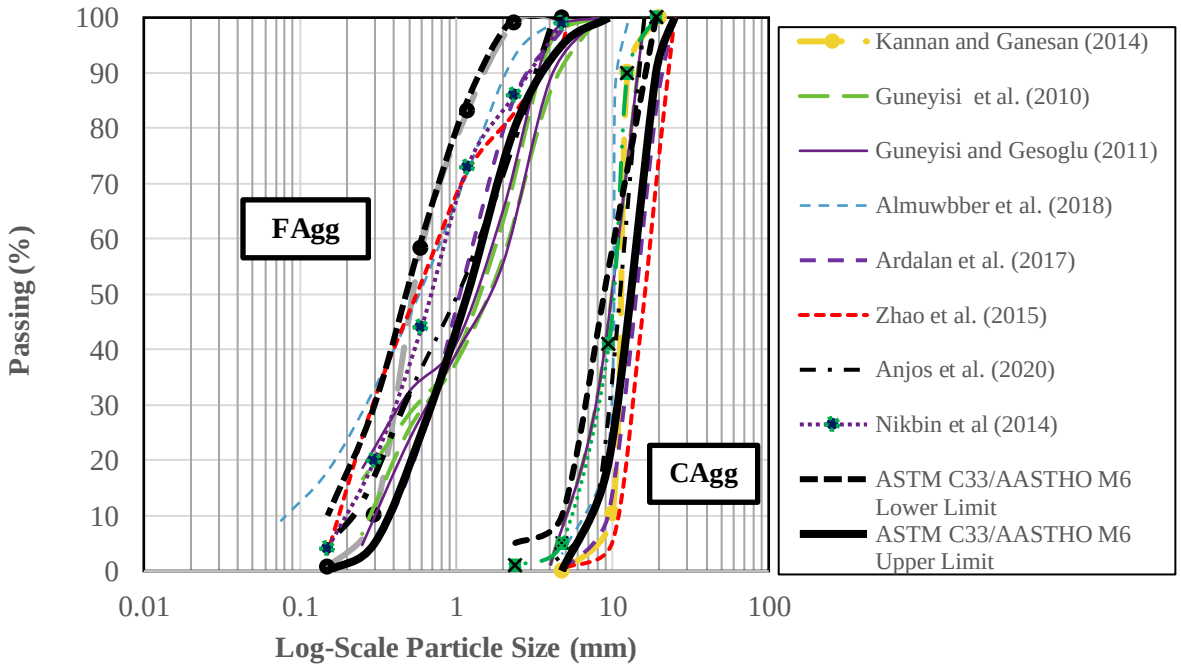


Fig. 7.3 Some PSD analysis results of aggregates in the database used to produce SCC.

Furthermore, the weight (Wt) of water to weight of cement or binder ratio (Wt_W/Wt_C or Wt_W/Wt_B) is a dimensionless and a more interpretable variable compared to weight of water in the mixture, e.g. with the unit of kg or kg/m^3. Here, the weight of binder (B) is considered as the sum of weights of C, A, GGBS, SF and MK of SCC mixes reported. The ratio of the weight of constituent i to the binder (Wt_i/Wt_B) is dimensionless which is considered instead of considering the amount of material (kg or kg/m^3) in the mix. This way, the control mix is the one where B equals C. Therefore, when cement is replaced with SCMs a binary, ternary or composite binder is produced.

Although the raw form of features can be considered as model input variables, some new indices and ratios are proposed in the present paper. The proposed variables are normalised and dimensionless. Additionally, the number of input variables is reduced from 38 to 23 input variables without ignoring any features. Obviously, the less the number of input variables, the less the chance of complexity and overfitting, particularly when there is a limited number of data. Although LP can be used as replacement of cement, it is considered as an admixture to indicate the feasibility of considering different variables. Due to the lack of reported characteristics, LP, water and SP are considered to be the same in terms of type. Therefore, only their proportions in the mixture were considered as variables. Finally, the UCS_{28d} can be represented as a function of the developed independent input variables as follows:

$$UCS_{28d} = f\left(\frac{Wt_C}{Wt_B}, \frac{Wt_A}{Wt_B}, \frac{Wt_{GGBS}}{Wt_B}, \frac{Wt_{SF}}{Wt_B}, \frac{Wt_{MK}}{Wt_B}, \frac{Wt_{LP}}{Wt_B}, \frac{Wt_{CAgg}}{Wt_B}, \frac{Wt_{FAgg}}{Wt_B}, \frac{Wt_W}{Wt_B}, \frac{Wt_{SP}}{Wt_B}, MSI, WA_{CAgg}, WA_{FAgg}, HRI_C, PRI_C, HRI_A, PRI_A, HRI_S, PRI_S, HRI_{SF}, PRI_{SF}, HRI_{MK}, PRI_{MK}\right) \quad (7\text{-}19)$$

Note that Wt_i/Wt_B are proportional variables which reflect the amount and MSI, WA_i, HRI_i and PRI_i are representatives of characteristics and type of i^{th} material in the model.

Values of collected raw data and model variables are attached as supplementary document for more details and considerations. Descriptive statistics of variables considered here are given in Table 3-5.

Table 7.3 Descriptive statistics of variables reflecting the proportions of materials in SCC.

Indicator	Wt_C/Wt_B	Wt_A/Wt_B	Wt_{GGBS}/Wt_B	Wt_{SF}/Wt_B	Wt_{MK}/Wt_B	Wt_{LP}/Wt_B	Wt_{CAgg}/Wt_B	$Wt_{FAgg}/Wt/_B$	Wt_W/Wt_B	Wt_{SP}/Wt_B
Mean	0.74	0.14	0.08	0.02	0.02	0.13	1.56	1.90	0.43	0.02
St. D.	0.18	0.14	0.18	0.04	0.05	0.17	0.47	0.38	0.07	0.01
Sample Var.	0.03	0.02	0.03	0.002	0.003	0.03	0.22	0.14	0.005	0.0001
Min.	0	0	0	0	0	0	0.893	0.84	0.26	0.0024
Max.	1	0.745	1	0.25	0.3	1	3.787	3.105	0.702	0.034

Table 7.4 Descriptive statistics of variables reflecting type and properties of materials in SCC.

Indicator	MSI	WA_{CAgg}	WA_{FAgg}	HRI_C	PRI_C	HRI_A	PRI_A	HRI_{GGBS}	PRI_{GGBS}	HRI_{SF}	PRI_{SF}	HRI_{MK}	PRI_{MK}
Mean	2.9	0.92	1.11	3.38	0.29	0.29	0.48	0.36	0.10	0.01	0.29	0.15	0.19
StD	0.53	0.44	0.48	0.25	0.02	0.29	0.42	0.69	0.20	0.02	0.43	0.31	0.37
Sample Var.	0.28	0.19	0.23	0.06	3e-4	0.09	0.17	0.48	0.04	5e-4	0.18	0.10	0.14
Min.	2.1	0.22	0.08	1.95	0.26	0	0	0	0	0	0	0	0
Max.	5.3	2.0	2.490	3.76	0.45	0.73	0.95	1.86	0.56	0.058	0.98	0.89	0.96

Table 7.5 Target or output variable.

Indicator	UCS_{28d} (MPa)
Mean	56.00
StD	16.92
Sample Var.	286.37
Min.	16
Max.	118.404

7-5 **Model development and results**

In present study, the experimental data provided by [144], including 17 UCS_{28d} test results, is considered as Test data to examine the generalization performance of UCS_{28} models on mixtures with new material type and proportions. The remainder dataset is separated into two groups, Train and Validation data.

70% of the data published in each research paper including 328 data points are used to develop the models, Train data, and 30% including 111 experimental results are considered as Validation data. Models are trained using only Train Data. This scenario and data split form is considered to check the performance of the model on some unseen mixture proportions only while the material type is known in trained model, i.e. Validation Data. Test data is chosen to assess models on a dataset from consisting of new type of constituents. The coefficient of correlation (R), coefficient of determination (R^2), the root mean square error (RMSE) and the mean absolute error (MAE) between the predicted and experimental values of UCS_{28d} as the output are used for the initial accuracy evaluation of the models. As already noted, the coefficients of variables and bias term in QRI model are obtained using OLS method and analysis of variances similar to linear regression but with quadratic and interaction terms. GPR and SVR are nonlinear models with a number of parameters and hyper-parameters which need to be adjusted based on data. ML techniques assumes that data are normally distributed. However, all dataset may not meet this. To transform the distribution of data to a normal and also scale their range, all data are standardised using z-score method. In order to develop GPR and SVR models, a 5-fold cross validation is implemented to train the models using Train data. Cross-validation aims to validate the performance of models on different subsets of data that was not used in the training process [121, 125]. In this process, the dataset is firstly divided to k equal-sized folds or subsets of data. Then, the algorithm trains the model using k-1 folds of data, namely train folds, and hold out one fold, i.e. validation fold. In other words, the parameters are tuned using k-1 subsets, i.e. training process, and the first fold is used for validation. Then, the second fold becomes the validation fold and the model is trained on the other k-1 folds. This process is summarised in Fig. 7.4. The parameters and hyper-parameters of the model are updated after each iteration. The number of iterations can be a user-defined parameter or a criterion such as a fitness function or condition, e.g. R>0.9. The k-fold validation method enables updating the parameters of the model k times on different k folds of data which results in reducing overfitting.

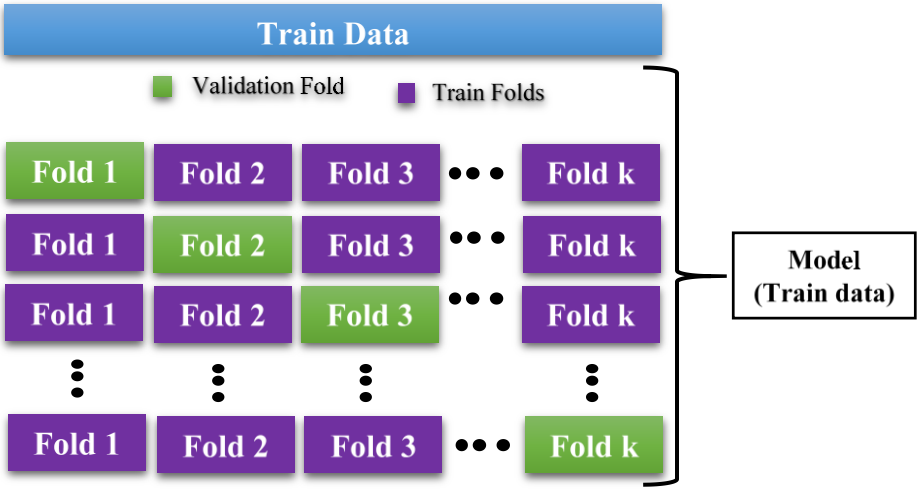


Fig. 7.4 K-fold cross validation used to develop SVR and GPR models using Train data.

After developing several GPR and SVR models with different kernel functions, those models with higher accuracy based on initial error evaluation using R, RMSE and MAE values are chosen for further evaluation. Amongst GPR models with different kernel functions, the model with rational quadratic kernel function, namely rational quadratic GPR (RQGPR) model, outperformed other models. Similarly, the SVR model with quadratic kernel function (QSVR) outperformed other SVR models with different kernel functions. These models are black-box due to the number of parameters and hyper-parameters, but they can be recalled by a computer for accuracy assessment, validation, prediction and further use.

Fig. 7.5 illustrates scatter plots of the predicted versus experimental values of UCS_{28d} obtained by models on Train and Validation data. The values of R, RMSE and MAE are also given in Table 7.6 to have an overall viewpoint on the accuracy and error of prediction. It is recommended by researchers that a regression model which gives $R>0.8$ is acceptable and accordingly $R^2>0.64$ [30, 115]. According to Table 7.6 and Fig. 7.5 (a), all models perform well on Train data. Although QRI perform better than other models on Train data, it gives large error when it comes to Validation data as is represented in Fig. 7.5 (b) and Table 7.6. As was mentioned, the coefficients of QRI are obtained using the analysis of variances. Therefore, they are mainly useful on data they are calibrated and no more.

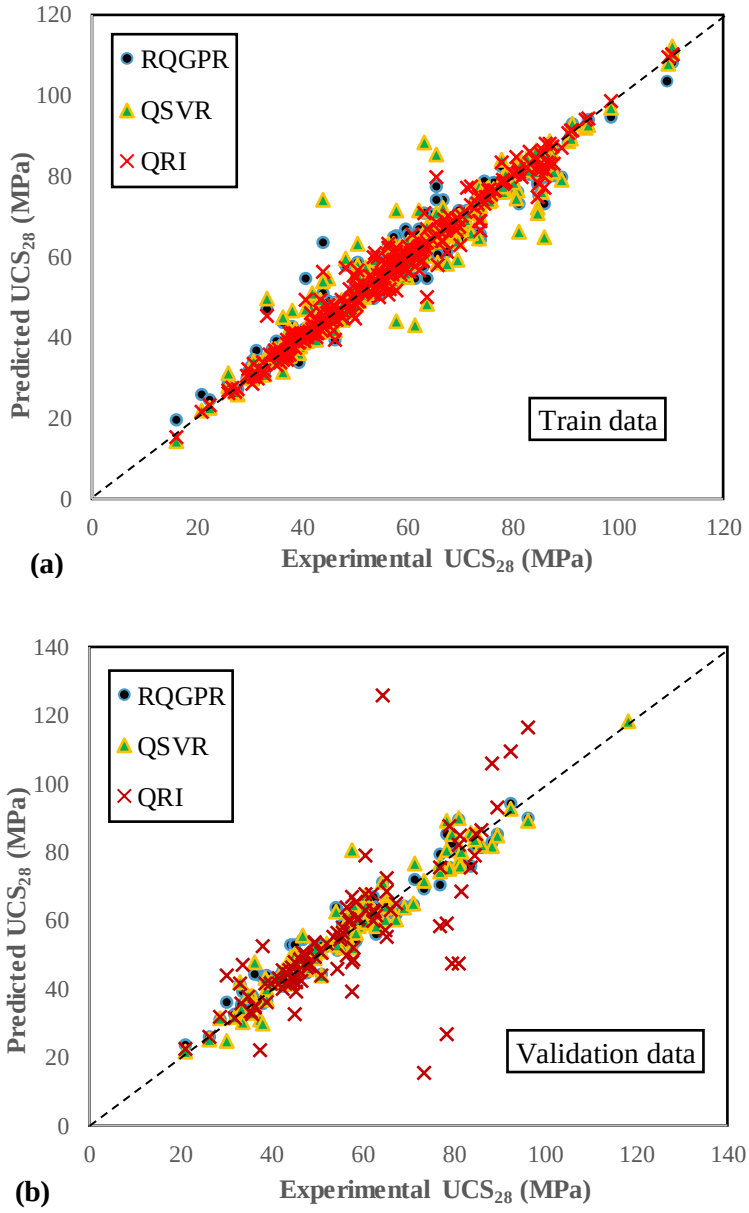


Fig. 7.5 Scatter plots of the predicted versus experimental values of UCS_{28d} obtained by GPR model on (a) Train and (b) Validation data.

Table 7.6 Performance of models on Train and Test data.

Parameter	Train Data			Validation data		
	QRI	QSVM	RQGPR	QRI	QSVM	RQGPR
R	0.99	0.96	0.98	-0.09	0.97	0.97
R^2	0.97	0.92	0.96	0.01	0.94	0.93
RMSE	2.90	4.70	3.66	175.16	4.41	4.55
MAE	1.70	2.86	2.49	166.50	3.02	3.26

RQGPR and QSVR still represent high accuracy on Validation data according to Fig. 7.5 (b) and Table 7.6. Hence, it can be concluded that these models can predict the UCS_{28d} for unseen mixture proportions when the type of constituents is known. This also confirms that these models neither underfit nor overfit

for this scenario. The second scenario questions whether it is possible to use QSVR and RQGPR models for new materials with different properties or proportions. This can be called the predictability on new types of material. As already mentioned, some experimental data provided by [144] are reserved to address this question where type of FA and accordingly the mixtures are new to the model. The experimental and predicted values of UCS_{28d} and the residual error (RE) values for Test data is summarised in Table 7.7. The details of Test data is provided in attached supplementary document for more information.

Table 7.7 The performance of different models for predicting the UCS of SCC with unseen type of material.

No	Mix Label	Experimental UCS_{28} (MPa)	Predicted UCS_{28} (MPa)			Error (MPa)		
			RQGPR	QSVM	QRI	RQGPR	QSVM	QRI
1	**100% OPC**	40.77	42.37	47.47	40.77	1.60	6.70	0
2	**FA05**	42.73	46.03	30.89	-37.93	3.30	-11.84	-80.66
3	**FA10**	44.31	45.12	28.81	-37.93	0.81	-15.51	-82.24
4	**FA15**	48.99	43.96	26.41	-37.93	-5.03	-22.58	-86.92
5	**FA20**	46.43	42.48	23.71	-37.93	-3.94	-22.72	-84.36
6	**FA25**	40.25	40.70	20.69	-37.93	0.46	-19.56	-78.18
7	**FA30**	39.43	38.70	17.36	-37.93	-0.73	-22.07	-77.36
8	**MK05**	48.28	47.64	49.65	46.44	-0.64	1.37	-1.84
9	**MK10**	51.91	51.42	51.48	51.89	-0.49	-0.43	-0.02
10	**MK15**	54.53	54.53	52.75	55.15	0.00	-1.78	0.62
11	**MK20**	57.17	54.72	53.46	56.21	-2.45	-3.71	-0.96
12	**MK25**	53.74	52.59	53.60	55.08	-1.15	-0.14	1.34
13	**MK30**	51.40	50.88	53.19	51.75	-0.52	1.79	0.35
14	**MK05+FA05**	46.88	49.85	37.85	-24.61	2.97	-9.02	-71.49
15	**MK10+FA10**	47.48	51.95	33.16	-24.56	4.46	-14.33	-72.04
16	**MK15+FA15**	53.62	51.97	27.15	-24.51	-1.65	-26.47	-78.13
17	**MK20+FA20**	44.68	49.61	19.83	-24.46	4.93	-24.85	-69.14

According to Table 7, QSVR and QRI models produce large errors for mixtures containing new type of fly ash and only RQGPR model still performs with high accuracy. MAE values as overall error indicator for QRI, QSVR and RQGPR predictions are respectively 2.07, 12.05, 46.21, while RMSE values are 2.65, 15.24 and 60.01. These results impose the deficiency of QRI model for predictive modeling of the UCS. As already noted, QRI is the quadratic form of multiple variable polynomial regression methods, which also reflect the interaction effects of variables. The structure of the QRI model is known in advance. The coefficient of terms in QRI model are calculated using analysis of variances of variables after fitting the model on data using OLS approach. Although QRI was able to successfully model the UCS on Train Data, its prediction performance considerably decreased when it came to Validation and Test datasets. This weakness arises due to the fact that parameters of QRI model

are calibrated for only those data used to develop the model and are not validated for unseen data which is called overfitting.

ML models are commonly developed using a process of validation on different sets of data, which results in better prediction performance. GPR and SVR to apply kernel-based ML algorithms for predictive modeling of the UCS of SCC based on the properties and proportions of materials which have been less applied so far. Both QSVR and RQGPR models are non-parametric, i.e. they are not limited by a functional form, and are kernel-based models. Kernel function convert data from an original input space into a higher dimensional feature space in which a dominated hyper-plane can be found to consider a functional mapping between a set of input variables and the output. Therefore, kernel functions are useful to find the best model. ML methods also use regularisation techniques to add a penalty to model parameters, except for constants or intercepts, to make them robust against problems such as multicollinearity and overfitting. These are the reasons why they perform well on validation data. SVR models commonly map data into a higher dimensional space using kernel transformation functions which was a quadratic function in this paper. The quadratic kernel function was chosen in SVR and GPR models to allow for the quadratic response in the output. The reason why RQGPR outperforms the QSVR model can be realised form Fig. 7.6. The bar chart in Fig. 7.6 represents the experimental and predicted values of UCS_{28d} in Test data. The polynomial lines indicate the trends of experimental values UCS_{28d} when C is partially replaced with different percentages of FA, MK and FA+MK and those predicted by RQGPR and QSVR models. Both experimental and RQGPR increase and start decreasing after a specific replacement of C with FA while QSVR tends to merely decrease which does not conform to experimental results. It can be seen in Fig. 7.6 (a) and (c) that RQGPR model trends follow the experimental, while those of QSVR model are not correct. This incorrect sensitivity or parametric response of predictive models results in prediction errors and vice versa.

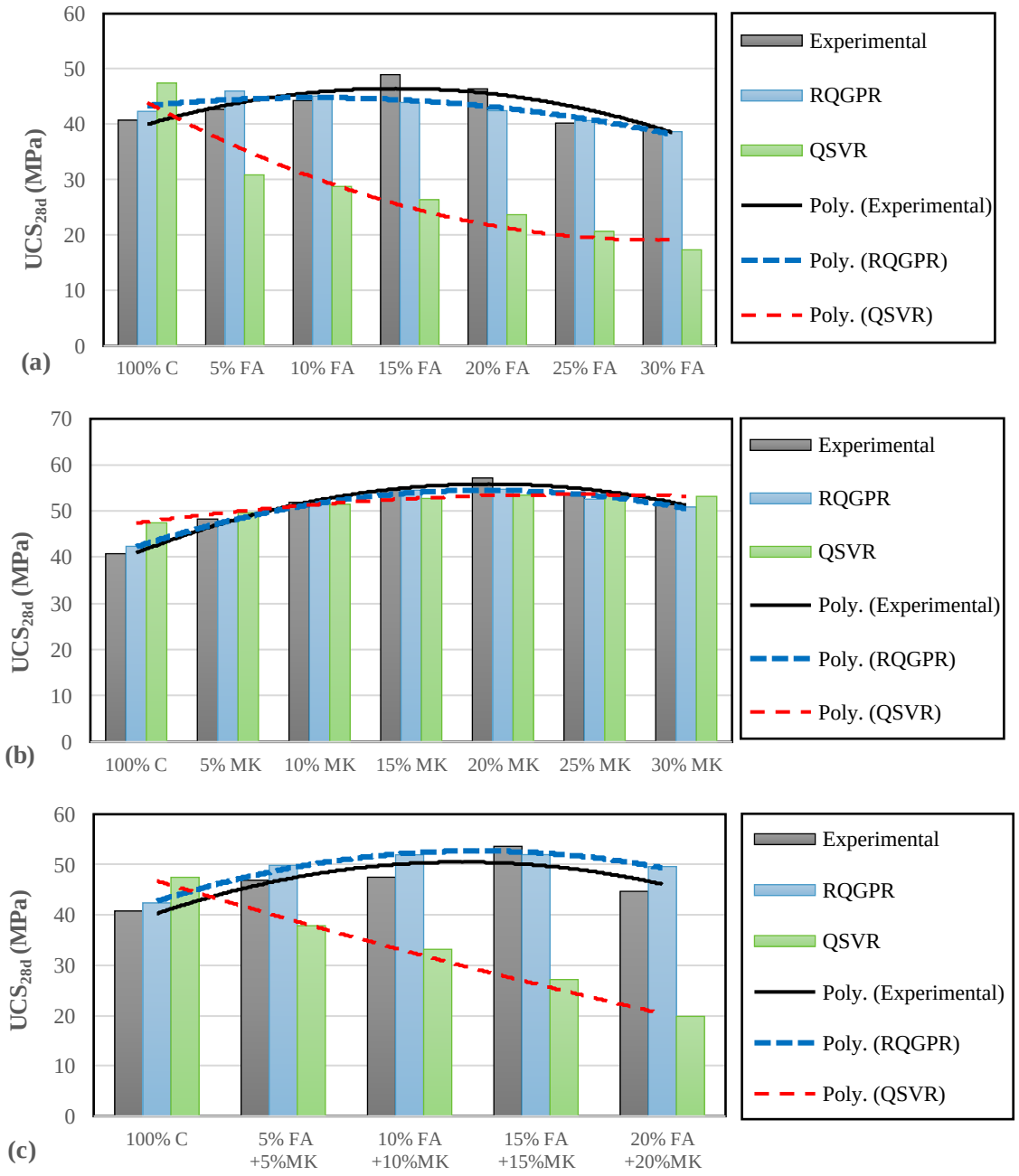


Fig. 7.6 Bar chart values and trend lines obtained by experimental and predicted UCS_{28d} on Test data where C is replaced with different percentages of (a) FA, (b) MK and (c) FA+MK.

The better performance of RQGPR compared to QSVR is due to the fact that RQGPR uses probability distributions over all acceptable relationships fitting the data. This means that the Gaussian probability distribution and more accurate covariance kernels between variables helped better prediction for the considered new types of materials in this paper. Here, according to results, only RQGPR performed well on all subsets of data and met the conditions. Considering the results of model performance analyses in the present paper, RQGPR is proposed as a robust tool for predictive modeling of the UCS_{28d} of SCC based on type and proportions of mixture constituents.

7-6 **Sensitivity analysis**

In order to realise the contribution and importance of considered variables, a sensitivity analysis (SA) can be done on the RQGPR model as the best model found. Several SA approaches have been proposed for different purposes [30, 112, 222] Typically, SA aims to investigate how and which input variables change the output. When it comes to nonlinear models such as RQGPR, a proper SA can be complicated as the output may change differently in local changes of an input variable or due to correlations of variables. According to [222], SA can be used to rank input factors with respect to their relative contribution to the output variability. The method proposed here is based on measuring the amount of output variation or prediction error when an input variable is removed from the model. RMSE between predicted and experimental values of output is chosen as the indictor of error. For this aim, RMSE is calculated for the output using the training data as they are assumed to be available, named Actual RMSE. Then, the values of one variable are changed to zero in the same dataset, while the amounts of all other variables are unchanged. After calculating the output using the RQGPR model, the new RMSE is calculated again. Obviously, the larger variation of RMSE is caused by the most influential factor. This process is done for all variables in RQGPR model and the results are given in Fig. 7.7.

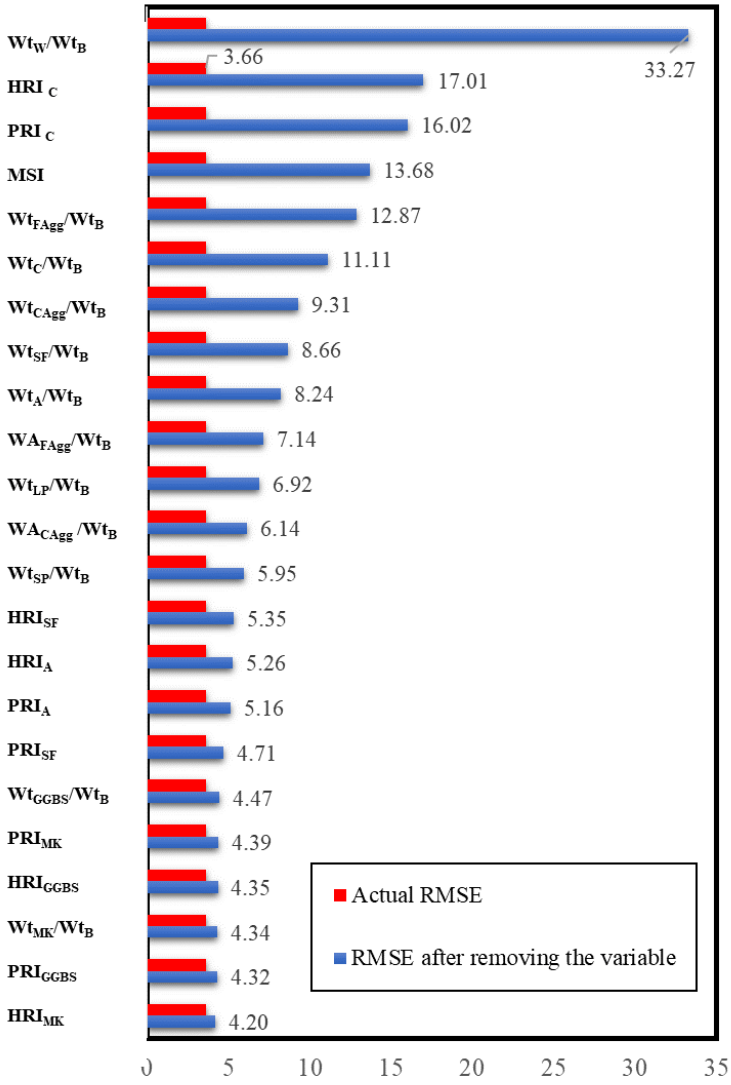


Fig. 7.7 SA for ranking and relative importance of considered factors for prediction of the UCS_{28d} using RQGPR model.

Fig. 7.7 shows that removing Wt_W/Wt_B from the RQGPR model changes the prediction error significantly from RMSE of 3.66 to 33.27. This increase in prediction error demonstrates the importance of Wt_W/Wt_B. Interestingly, the respective largest RMSE increase are caused by HRI_C, PRI_C and MSI which proves the prominence of considering properties and type of materials as influential factors. This indicates the contribution of the type of cement and particle size and grade of aggregates to the UCS_{28d} of Eco-SCC, which have been represented by several researchers using experimental studies such as those researched by [218]. However, such variables have commonly been ignored in previously proposed models. The main objective of this paper was to demonstrate that all influential factors such as type and properties of constituents of composite materials such as Eco-SCC must be considered as variables for predictive modeling of mixture factors such as UCS_{28d}. This issue is validated by the proposed SA and the results shown in Fig. 7.7. In order to have more comprehensive models, better

understanding of effect of variables and decision-making, a more comprehensive dataset containing larger number of experimental data and variables is necessary.

7-7 **Conclusions**

This paper represented that each constituent of concrete mixtures has specific type, chemical and physical properties which influence its properties such as UCS_{28d} where previously proposed models merely relate such properties to the amount or proportions of constituents only. Such models cannot differentiate the type of material and the influence of their properties and output the same results e.g. the same UCS_{28d} for an SCC containing Cement type I and Type V with the same mixture proportions. There are several other uncertainties and influential factors which must be reduced or considered as variables to develop reliable models. For example, this paper only considered UCS_{28d} of cubic specimens while there are experimental data conducted on cylindrical samples which have different UCS. Therefore, such results cannot be merged. This paper also provides new insights into data-pre-processing, feature transformation and selection and dimensionality reduction for development of ML models. HRI and PRI were proposed for classification of cementitious material in terms of hydraulic and pozzolanic reactivity and reducing the dimensionality of variables.

As a project, two separate datasets were considered to test the predictability of the models, validation data had new proportions of materials only and the other set, i.e. test data, contained a new type of material. As already noted, GPR was successful to find a model which can perform well on both considered Test datasets. Showing the capability of ML methods to find an efficient model working well when it comes to new proportion and new type of materials was the main objective of the paper, where only GPR is successful. SVR was used for comparisons and show that a model may perform well on data with new proportions, Validation data here, but it may fail when it comes to new type of material, i.e. Test data. QRI is used to represent the incompetence of classical ANOVA-based approaches for prediction aims. Such models are merely calibrated for the Train data and may not perform well on unseen data. Data pre-processing, regularisation and training process in ML methods enable them to perform accurately in prediction.

This paper aimed to demonstrate that all influential factors such as type and properties of constituents of composite materials such as Eco-SCC must also be considered as variables for an appropriate predictive modeling of mixture factors such as UCS_{28d}. However, such variables have commonly been ignored in previously proposed models. The SA confirmed the importance of considering properties of materials as influential factors as HRI_C, PRI_C and MSI had largest effects on prediction after the amount of water in the mixture compared to other variables. In this paper, RQGPR outperformed other models with regard to the considered variables, dataset and scenarios. Considering the results, RQGPR is proposed as a robust tool for predictive modeling of properties of concrete.

Although several other variables can be considered, this research confirmed that a comprehensive model is achievable that can predict the properties of concrete such as UCS_{28d} of Eco-SCC with high accuracy. Such models can be used worldwide to avoid conducting costly and time-consuming tests, or at least to find the first trial mix design which is difficult due to the large number of materials and factors. An efficient model requires sufficient number of data points, suitable variables and appropriate modeling tool which have been investigated in this paper. Note that the designer must be aware of the properties of materials such as cement and SCMs which are usually reported by their producers and retailers or can be acquired using experiments. Considering the large number of materials and variables, it is expected that a cloud repository is required to collect data and use for development of ML models, similar to image databases used for image processing and computer vision. Models such as RQGPR can also be used for optimisation and multi-objective mixture design purposes, which greatly helps sustainability of the construction and concrete industries.

Chapter 8 Conclusions and recommendations for future work

8-1 Main findings

Optimal mixture design of CMs requires satisfying multiple objectives and constraints such as functional, economic and environmental objectives with respect to emphasis on sustainability issues in recent projects. The optimal mixture design can be the one with the least production cost and CO_2 emission while having the maximum workability and mechanical properties. Designers must also make the best possible decisions at a fast pace. However, CMs are still designed using costly traditional methods such as trial-and-error experiments. Such methods might not lead to the optimal mixture due to the large number influential factors and nonlinear relationships, multiple objectives and constraints while only a limited number of mixture designs and samples can be experimented. This thesis strived to engineer the concept of mixture design of CMs, address the gaps, and propose improved methods and frameworks for multi-objective mixture design optimisation (MOMDO) using AI, ML and optimisation algorithms for sustainable construction and mining projects. In the present thesis, each paper in each chapter investigates and addresses a problem in terms of MOMDO of CMs and ends up with a conclusion section. The main findings and key conclusions of this thesis can be summarised as follows:

1. Each chapter highlights the feasibility of ML algorithms for predictive modeling of complex properties of a pre-specified CM using a collection of experimental data. It was shown that ML algorithms are able to learn from data to discover high-accuracy relationships and models by iteratively tuning their parameters during a training process using a given dataset. Problem statement, determination of influential factors, correct data pre-processing steps, consistency and reliability of the data used to calibrate parameters of ML models are vital to develop accurate ML models. Chapter 1 and 2 highlighted that GP and LGP can generate models that can be converted to equations for manual calculations. Chapter 4-7 compared the performance of different ML algorithms such as traditional multiple polynomial regression methods, ANNs, GPR, SVR, XGBOOST and Ensemble Trees where GPR commonly outperformed other modeling approaches. SEGPR model was able to predict the UCS of UHPC with an accuracy of R, RMSE and MAE 0.999, 0.005 and 0.004 and those of Slump model were 0.999, 0.002 and 0.002. ANNs achieved R, RMSE and MAE of 0.987, 2.3 and 1.4 for prediction of Slump and 0.973, 6.2 and 4.4 respectively for prediction of UCS of UHPC on all data. RQGPR model in Chapter 8 outperformed QSVR and QRI models with R, RMSE and MAE 0.97, 4.55 and 3.26 on Validation data respectively for prediction of the UCS_{28d} of Eco-SCC. More importantly, only RQGPR model could perform accurately when it came to Eco-SCC mixtures with new type of materials. This highlighted that GPR is a robust method for predictive modeling of properties of CMs. Unique features of each ML approach were explained and discussed in Chapters.

2. Optimisation algorithms can be deterministic or stochastic. Deterministic approaches work without any randomness, while there is some randomness in the algorithm in stochastic approaches. If there is some randomness in the algorithm, the algorithm will usually reach a different point every time it is executed, even with the same configuration. However, deterministic algorithms reach the same solution after each run. Metaheuristics such as GAs and PSO are good examples of stochastic algorithms. Such population-based approaches improve multiple candidate solutions after each iteration to enable search the space of solutions. Furthermore, optimisation methods are commonly designed to solve specific problems while metaheuristics are problem-independent techniques that can be applied to a broad range of problems. They have broad applicability and can be combined (hybridisation). The selection of the best ensures that solutions will converge to the optimum, while diversification via randomization allows the search to escape from local optima and, at the same time, increases the diversity of solutions. A good combination of these two major components will usually ensure that global optimality is achievable. Although these algorithms cannot guarantee the optimality of the solution, they have been applied for many benchmark problems and found more successful than other algorithms in terms of finding local and global optimum solution.

The mixture design problem is commonly a multi-objective multi-constraint optimisation problem with a set of linear and or nonlinear objectives and constraints which may conflict. Objectives and constrains might be nonlinear in this subject. Therefore, algorithms such as linear programming and simplex methods are not applicable for all problems except for those where objectives and constraints are linear. In each run, metaheuristics may find different sub-optimal solutions but with a degree of accuracy. Traditional nonlinear programming (NLP) methods usually find one solution which might be a local optima and cannot search neighbourhood of the local optima to find other local optimum solutions or global optimum. When it comes to a large number of variables they fail to find solution. More importantly, when it comes to multi-objective optimisation, a set of trade-off solutions is anticipated rather than one solution as objectives may conflict. One objective may increase with the increase of variable(s) while the other may decrease. This conflict leads to sacrifice of one objective when it comes to optimisation (maximisation/minimisation). Therefore, a set of trade-off solutions is preferable. These trade off solutions are called Pareto-optimum (Pareto-optimal or Pareto front) solutions rather than optimum.

3. This research highlighted that crucial steps of MOMDO are proper project statement, precise understanding and determination of objectives, constraints, influential factors, and the use of robust mathematical methods. It was demonstrated that the accuracy of models or relationships used as objective or constraint is of paramount significance. Two robust metaheuristic optimisation algorithms PSO and NSGA-II have been mainly used in this thesis to convert considered project scenarios to mathematical optimisation problems. More importantly, the results of proposed MOMDO methods have been validated for the considered scenarios and this has been less emphasised in presented literature so far. After trial-and-error experiments, it was found that PSO is faster than GA in terms of

convergence, while NSGA-II is better in avoiding local optima. NSGA-II is more improved than PSO in terms of programming and coding. Different forms of constraints have been considered in NSGA-II that enable the designer to expedite the process of finding more optimal solutions. Chapter 3-6 highlighted the flexibility of proposed MOMDO framework using AI algorithms for the design of CMs such as CPB, UHPC and SCC considering different practical single or multi-objective with or without constraints scenarios. Single–objective optimisation (SOO) algorithm aims to find solutions or mixture designs to satisfy one objective such as maximum UCS only, while multi-objective optimisation (MOO) algorithm discovers Pareto-optimal solutions that are diverse and meet multiple objectives which may contradict such as slump, UCS, production costs and CO_2 emission and considered constraints. In Pareto-optimal space of solutions, each solution is better than others in terms of one objective. This way enables the designer to choose the best mixture design based on priorities. Furthermore, constraints enable to completely reflect the project requirements and limit the space of solutions to ideal. Constraints can be upper and lower limit of design variables, the number of variables, any linear or nonlinear, equality or inequality relationship between variables and even models such as slump or cost as considered in different scenarios in Chapters.

4. In terms of mixture design of CMs, various objectives such as functional, environmental and economic requirements might be important. The optimum mixture can be the one with maximum UCS and Slump, and minimum production cost and CO_2 emission with many constraints such as UCS of more than X and production cost of less than Y which depends on the considered project statement and priorities. Chapter 3 represented the mixture design of CHB for different strength conditions as an SOO problem with constraints that was solved using PSO algorithm. Strength objective models and constraints are developed using GP algorithm. The methodology can find an optimum mixture design with a certain UCS at a specified time as are required by a mine planning engineer. Chapter 4 and 5 considered the mixture design of CPB and UHPC as MOO problem using PSO algorithm. UCS, Slump and production cost were considered as objectives to be maximised or minimised based on the considered scenarios. Chapter 6 proposed NSGA-II for MOMDO of Sustainable SCC containing various SCMs. UCS, Slump and production cost, CO_2 emission were considered as objectives or constraints in different scenarios. In all cases, predictability and performance of models were evaluated and the results found by the optimisation algorithm were confirmed which demonstrates the robustness of proposed approach.

5. In terms of mixture design, the designer commonly mixes different amounts or proportions of materials with specific properties and prepare some specimens to measure the properties such as UCS or slump based on specific methods and conditions. For this case, it can be assumed that all factors such as type of materials are constant except for the amount or proportions of materials which are controllable. Therefore, the only variable of the system is the proportion of material in the mixture such as amounts of cement, water, aggregates and superplasticiser in SCC or amount of tailings, cement and water in CPB as considered in Chapters 2-5. In Chapter 2-5, only proportions of materials

in CMs were influential factors and therefore were considered as variables to develop ML models. Different single objective or multiple objective design scenarios are considered for MOMDO as provided in Chapter 3-5.

6. Available experimental datasets commonly include results of tests such as UCS conducted on different specimens containing proportions of materials with specific chemical and physical properties mixed based on precise methods and conditions. For example in a dataset cement type II might be used while in another Cement type V. Even if all other factors are considered constant expect for type or specific properties chemical composition of cement, the UCS of these two samples can be different due to the difference in chemical composition of cement Type II and V. In fact, chemical and physical properties of these two types of cement can be used to separate them. When dealing with international databases which combines experimental data coming from different resources worldwide, type and properties of materials, specimen size, experimental methods and conditions might be different. However, proposed models in the literature commonly relate the properties to the proportion of constituents only and ignore the effect of their type and properties and other influential factors when it comes to international experimental databases. Such databases can be used to develop comprehensive ML models and MOMDO of CMs, as presented in Chapter 6 and 7. In this case, all influential factors such as chemical and physical properties of materials must be reflected to ML models as variables.

7. In Chapter 6, a comprehensive database was collected through a precise survey of the literature, which comprises of Slump and UCS_{28d} test results conducted Sustainable SCC specimens with different mixture proportions and type of materials. Slump and UCS_{28d} models are developed using ML algorithms. Chemical and physical properties of materials, properties of aggregates, specimens size, experimental methods were considered during the data collection or reflected to models as variables. After comparisons and validations of different ML models, the best Slump and UCS_{28d} in addition to Production costs and CO_2 emission models were used as objective functions or constraints for MOMDO of an optimal Sustainable SCC. In this chapter, NSGA-II was used as the optimisation algorithm to find mixture designs which could meet different MOO or SOO scenarios with or without required constraints. It was demonstrated that experimental datasets coming from different resources worldwide can be merged to compile comprehensive databases. Such databases can be used to develop more comprehensive ML models and MOMDO framework.

8. Chapter 6 and 7 highlighted the importance of reflecting the type and properties of materials and other influential factors to develop comprehensive ML models using international experimental data. Chapter 7 demonstrated that although a model may perform well on mixtures with new proportions, it may fail if the type or properties of a constituent is changed. Chapter 7 provided new insights into variable selection and dimensionality reduction. In this chapter, two new indices, HRI and PRI, were proposed to classify cementitious materials in terms of pozzolanic and hydraulic reactivity. HRI and

PRI combine chemical compositions of materials and transform them into new variables resulting in a meaningful dimensionality reduction.

9. This research proposed MOMDO frameworks using AI algorithms to enable a more sustainable production, optimum consumption of resources, reducing the use of costly materials, reducing overall costs and energy in projects. This method also allows for making the best design decision fast while reducing the costs and efforts of experimental design methods.

8-2 **Suggestions for future work**

This thesis has presented the viability and significance of proper MOMDO of CMs using AI algorithms. Suggestions and recommendations for future work are outlined as follows:

1. First, a comprehensive and reliable online data repository is recommended to be created. It will contain international experimental datasets with all possible influential factors required to develop possible objectives and constraints for design of CMs or any other materials. The biggest challenge in this research was to collect reliable and consistent data with all variables and influential factors. There were several problems such as incorrectly reported parameters or test results, lack of quality assurance, inconsistency and lack of completely reported influential factors. Even if one word such as specimen's size or shape, chemical or physical properties of materials is incorrectly reported, they cannot be used to develop models as the results are different. In fact, such issues are recognisable and the dataset can be cleaned, corrected or updated. With such comprehensive databases, one can come up with new ideas and methods. Importantly, ethical issues, policies and data sharing and privacy issues must not be forgotten which make the data collection process very challenging and slow or even impossible.

2. This research applied a few ML methods for predictive modeling of complicated objectives such as some properties of CMs. Data science and ML methods are being improved progressively and fast, and new ideas are being proposed to reduce computational costs while enhancing the accuracy. When it comes to new datasets, the accuracy of ML methods would be different due to the statistical properties of data and number of variables. Therefore, the applicability and accuracy of new ML methods would have to be investigated. Note that the predictability of ML models increases with the number and broadness of data used for training and calibrating their parameters which depends on the availability of a comprehensive international database.

3. This research used some search-based metaheuristic algorithms such as PSO and NSGA-II due to their usefulness in finding solutions for nonlinear models and multi-objective multi-constraint problems. Such algorithms still need improvements as they may be trapped in local optima that leads to incorrect solutions. However, this can be tackled by repeating the optimisation process and checking the results by an expert designer. It was demonstrated in this thesis that the designer is able to check the accuracy of the results by running the procedure and comparing the results with some available experimental data. Nonetheless, it is vital to develop more robust and efficient optimisation algorithms

with less computational costs and the ability to find optimum solutions. In this research, merely a few objectives were considered for MOMDO of CMs and the applicability of optimisation algorithms to find solutions when it comes to many complicated objectives still needs to be investigated and improved.

4. This thesis mainly considered some functional, environmental and economic objectives and constraints such as strength, slump, production costs and CO_2 emission. It is recommended to consider more objectives such as durability, energy consumption and life cycle analysis (LCA) related objectives or any other important factors such as required objectives or constraint in 3d printed concrete design or pumping CPB to fill out the mined ore can be considered as additional objectives and constraints.

5. The accuracy of ML models is of paramount significance. However, it is a big challenge how to analyse models and interpret their results and performance which are commonly highly nonlinear and black-box. New methods and ideas are required to examine the robustness of an ML model particularly when it comes to big data which is recommended for future studies.

6. With respect to progress of data science and digital technologies, the proposed approach in this research can be used to create an online universal approach for MOMDO of CMs or any material. It can also be expected that such an approach is used to automate the process of design of materials by only determining of project requirements without the need to repeat the tests. Indeed, such an approach enables the design and production of materials more sustainable with less production costs and carbon footprint and perfectly optimal properties.

Chapter 9 References

1. C150M-20 AC. Standard Specification for Portland Cement. ASTM International, West Conshohocken, PA. 2020. doi: 10.1520/C0150_C0150M-20
2. McLellan BC, Williams RP, Lay J, Van Riessen A, Corder GD. Costs and carbon emissions for geopolymer pastes in comparison to ordinary portland cement. Journal of cleaner production. 2011;19(9-10):1080-90.
3. Elchalakani M, Basarir H, Karrech A. Green concrete with high-volume fly ash and slag with recycled aggregate and recycled water to build future sustainable cities. Journal of Materials in Civil Engineering. 2017;29(2):04016219.
4. Liu R, Durham SA, Rens KL, Ramaswami A. Optimisation of Cementitious Material Content for Sustainable Concrete Mixtures. Journal of Materials in Civil Engineering. 2011;24(6).
5. Kim T, Tae S, Roh S. Assessment of the CO2 emission and cost reduction performance of a low-carbon-emission concrete mix design using and optimal mix design system. Renewable and Sustainable Energy Reviews. 2013;25:729-41.
6. Elchalakani M, Basarir H, Karrech A. Green concrete with high-volume fly ash and slag with recycled aggregate and recycled water to build future sustainable cities. Journal of Materials in Civil Engineering. 2016;29(2):04016219.
7. Chen C-T, Gu GX. Machine learning for composite materials. MRS Communications. 2019:1-11.
8. Basarir H, Bin H, Fourie A, Karrech A, Elchalakani M. An adaptive neuro fuzzy inference system to model the uniaxial compressive strength of cemented hydraulic backfill. Mining of Mineral Deposits. 2018.
9. Sivakugan N, Veenstra R, Naguleswaran N. Underground Mine Backfilling in Australia Using Paste Fills and Hydraulic Fills. International Journal of Geosynthetics and Ground Engineering. 2015.
10. Lerche R, Renetzeder H. Lerche, R., and H. Renetzeder. "Development of 'pumped fill'at Grund mine, Preussag AG Metall. Proceedings of the 9th International Conference on the Hydraulic Transport of Solids in Pipes. Rome, Italy1984.
11. Ercikdi B, Cihangir F, Kesimal A, Deveci H. Practical importance of tailings for cemented paste backfill. Paste Tailings Management. 2017:7-32.
12. Niroshan N, Sivakugan N, Veenstra RL. Flow characteristics of cemented paste backfill. Geotechnical and Geological Engineering. 2018;36(4):2261–72.
13. Carneiro A, Fourie AB. An integrated approach to cost comparisons of different tailings management options. 2019 Australian Centre for Geomechanics, Perth,. 2018.
14. Fall M, Benzaazoua M. Advances in predicting performance properties and cost of paste backfill. Proceedings of International Conference of Tailings & Mine Waste2003. p. 12-5.
15. Deilami S, Aslani F, Elchalakani M. Durability assessment of self-compacting concrete with fly ash. Computers and Concrete. 2017;19(5):489-99.
16. Douma OB, Boukhatem B, Ghrici M, Tagnit-Hamou A. Prediction of properties of self-compacting concrete containing fly ash using artificial neural network. Neural Computing and Applications. 2017;28(1):707-18.
17. Shi C, Wu Z, Lv K, Wu L. A review on mixture design methods for self-compacting concrete. Construction and Building Materials. 2015;84:387-98.
18. Venkatesan M, Zaib Q, Shah IH, Park HS. Optimum utilization of waste foundry sand and fly ash for geopolymer concrete synthesis using D-optimal mixture design of experiments. Resources, Conservation and Recycling. 2019;148:114-23.
19. DeRousseau M, Kasprzyk J, Srubar III W. Computational design optimisation of concrete mixtures: A review. Cement and Concrete Research. 2018;109:42-53.
20. Yeh I-C. Optimisation of concrete mix proportioning using a flattened simplex–centroid mixture design and neural networks. Engineering with Computers. 2009;25(2):179.
21. Ghafari E, Bandarabadi M, Costa H, Júlio E. Prediction of fresh and hardened state properties of UHPC: comparative study of statistical mixture design and an artificial neural network model. Journal of Materials in Civil Engineering. 2015;27(11):04015017.
22. Goodspeed CH, Vanikar S, Cook R. High-performance concrete defined for highway structures. Concrete International. 1996;18(2):62-7.
23. Chou J-S, Chiu C-K, Farfoura M, Al-Taharwa I. Optimising the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. Journal of Computing in Civil Engineering. 2010;25(3):242-53.
24. Aïtcin P-C, Flatt RJ. Science and technology of concrete admixtures. Woodhead publishing; 2015.
25. Gandomi A, Alavi A, Arjmandi P, Aghaeifar A, Seyednour R. Modeling of compressive strength of HPC mixes using a combined algorithm of genetic programming and orthogonal least squares. Journal of Mechanics of Materials and Structures. 2010;5(5):735-53.

26. Yeh I-C. Modeling of strength of high-performance concrete using artificial neural networks. Cement and Concrete research. 1998;28(12):1797-808.
27. Yeh I-C. Computer-aided design for optimum concrete mixtures. Cement and Concrete Composites. 2007;29(3):193-202.
28. Yeh I-C. Design of high-performance concrete mixture using neural networks and nonlinear programming. Journal of Computing in Civil Engineering. 1999;13(1):36-42.
29. Wu S, Li B, Yang J, Shukla S. Predictive modeling of high-performance concrete with regression analysis. Industrial Engineering and Engineering Management (IEEM), 2010 IEEE International Conference on: IEEE; 2010. p. 1009-13.
30. Sadrossadat E, Soltani F, Mousavi SM, Marandi SM, Alavi AH. A new design equation for prediction of ultimate bearing capacity of shallow foundation on granular soils. Journal of Civil Engineering and Management. 2013;19(sup1):S78-S90.
31. Ziaee SA, Sadrossadat E, Alavi AH, Shadmehri DM. Explicit formulation of bearing capacity of shallow foundations on rock masses using artificial neural networks: application and supplementary studies. Environmental earth sciences. 2015;73(7):3417-31.
32. Rostami MF, Sadrossadat E, Ghorbani B, Kazemi SM. New empirical formulations for indirect estimation of peak-confined compressive strength and strain of circular RC columns using LGP method. Engineering with Computers. 2018:1-16.
33. Tajeri S, Sadrossadat E, Bazaz JB. Indirect estimation of the ultimate bearing capacity of shallow foundations resting on rock masses. International Journal of Rock Mechanics and Mining Sciences. 2015;80:107-17.
34. Sadrossadat E, Heidaripanah A, Osouli S. Prediction of the resilient modulus of flexible pavement subgrade soils using adaptive neuro-fuzzy inference systems. Construction and Building Materials. 2016;123:235-47.
35. Ghorbani B, Sadrossadat E, Bazaz JB, Oskooei PR. Numerical ANFIS-Based Formulation for Prediction of the Ultimate Axial Load Bearing Capacity of Piles Through CPT Data. Geotechnical and Geological Engineering. 2018:1-20.
36. Gopal P, Bordoloi S, Cai W, Liang H, Boddu R, Sreedeep S, et al. Model Development for Computing Cracking in Soil Reinforced with Fibers from Three Different Bioresources. ADVANCES IN CIVIL ENGINEERING MATERIALS. 2018;7(1):669-93.
37. Mansorkhani AA, Ahmadi S, Rahmani H. Reliable Neural Networks for Proportioning of Concrete Mixes Containing Cement Replacement Materials. Advances in Civil Engineering Materials. 2018;7(1):633-50.
38. Sadrossadat E, Ghorbani B, Zohourian B, Kaboutari M, Rahimzadeh Oskooei P. Predictive modelling of the MR of subgrade cohesive soils incorporating CPT-related parameters through a soft-computing approach. Road Materials and Pavement Design. 2018:1-19.
39. Young BA, Hall A, Pilon L, Gupta P, Sant G. Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods. Cement and Concrete Research. 2019;115:379-88.
40. Garg A, Zhu H-H, Panda B, Paul SC. An Artificial Intelligence Model for Computing Optimum Fly Ash Content for Structural-Grade Concrete. Advances in Civil Engineering Materials. 2019;8(1):56-70.
41. Mousavi SM, Aminian P, Gandomi AH, Alavi AH, Bolandi H. A new predictive model for compressive strength of HPC using gene expression programming. Advances in Engineering Software. 2012;45(1):105-14.
42. Cheng M-Y, Chou J-S, Roy AF, Wu Y-W. High-performance concrete compressive strength prediction using time-weighted evolutionary fuzzy support vector machines inference model. Automation in Construction. 2012;28:106-15.
43. Hoang N-D, Pham A-D, Nguyen Q-L, Pham Q-N. Estimating compressive strength of high performance concrete with Gaussian process regression model. Advances in Civil Engineering. 2016;2016.
44. Behnood A, Behnood V, Gharehveran MM, Alyamac KE. Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. Construction and Building Materials. 2017;142:199-207.
45. Koza JR. Genetic programming II, automatic discovery of reusable subprograms. MIT Press, Cambridge, MA; 1992.
46. Gandomi AH, Alavi AH, Ryan C. Handbook of genetic programming applications. Springer; 2015.
47. Brameier M, Banzhaf W. A comparison of linear genetic programming and neural networks in medical data mining. IEEE Transactions on Evolutionary Computation. 2001;5(1):17-26.
48. Brameier MF, Banzhaf W. Linear genetic programming. Springer Science & Business Media; 2007.
49. Oltean M, Grosan C. A comparison of several linear genetic programming techniques. Complex Systems. 2003;14(4):285-314.
50. Sadrossadat E, Ghorbani B, Oskooei R, Kaboutari M. Use of adaptive neuro-fuzzy inference system and gene expression programming methods for estimation of the bearing capacity of rock foundations. Engineering Computations. 2018;35(5):2078-106.

51. Sadrossadat E, Ghorbani B, Hamooni M, Moradpoor Sheikhkanloo MH. Numerical formulation of confined compressive strength and strain of circular reinforced concrete columns using gene expression programming approach. Structural Concrete. 2018;19(3):783-94.
52. Sadrossadat E, Heidaripanah A, Ghorbani B. Towards application of linear genetic programming for indirect estimation of the resilient modulus of pavements subgrade soils. Road Materials and Pavement Design. 2016;19(1):139-53.
53. Shahin MA, Maier HR, Jaksa MB. Data division for developing neural networks applied to geotechnical engineering. Journal of Computing in Civil Engineering. 2004;18(2):105-14.
54. Francone FD, Deschaine LM. Extending the boundaries of design optimisation by integrating fast optimisation techniques with machine-code-based, linear genetic programming. Information Sciences. 2004;161(3-4):99-120.
55. Trujillo L, Naredo E, Martínez Y. Preliminary study of bloat in genetic programming with behavior-based search. EVOLVE-A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation IV. Springer; 2013. p. 293-305.
56. Smith GN. Probability and statistics in civil engineering. Collins Professional and Technical Books. 1986;244.
57. Alavi AH, Sadrossadat E. New design equations for estimation of ultimate bearing capacity of shallow foundations resting on rock masses. Geoscience Frontiers. 2016;7(1):91-9.
58. EN197 B. 1 (2011). Cement Composition, specifications and conformity criteria for common cements. British Standards Institution (BSI).
59. Aziz NAA, Alias MY, Mohemmed AW, Aziz KA. Particle Swarm Optimisation for constrained and multiobjective problems: a brief review. International Conference on Management and Artificial Intelligence IPEDR2011. p. 146-50.
60. Basarir H, Bin H, Fourie A, Karrech A, Elchalakani M. An adaptive neuro fuzzy inference system to model the uniaxial compressive strength of cemented hydraulic backfill. Mining of mineral deposits. 2018(12, Iss. 2):1-12.
61. Qi C, Tang X, Dong X, Chen Q, Fourie A, Liu E. Towards Intelligent Mining for Backfill: A genetic programming-based method for strength forecasting of cemented paste backfill. Minerals Engineering. 2019;133:69-79.
62. Sivakugan N, Veenstra R, Naguleswaran N. Underground mine backfilling in Australia using paste fills and hydraulic fills. International journal of geosynthetics and ground engineering. 2015;1(2):1-7.
63. Wang D, Tan D, Liu L. Particle swarm optimisation algorithm: an overview. Soft Computing. 2018;22(2):387-408.
64. Sadrossadat E, Ghorbani B, Oskooei R, Kaboutari M. Use of adaptive neuro-fuzzy inference system and gene expression programming methods for estimation of the bearing capacity of rock foundations. Engineering Computations. 2018.
65. Espley G, Beattie H, Pasieka A. Cemented hydraulic backfill within the Falconbridge Group of Companies. CIM Bulletin. 1970;701:1002-10.
66. Lerche R, Renetzeder H. Development of pumped fill at Grund Mine, Preussag AG Metall. Proceedings of the 9th International Conference on the Hydraulic Transport of Solids in Pipes, Rome, Italy1984.
67. Ercikdi B, Cihangir F, Kesimal A, Deveci H. Practical importance of tailings for cemented paste backfill. Paste tailings management. Springer; 2017. p. 7-32.
68. Niroshan N, Sivakugan N, Veenstra RL. Flow characteristics of cemented paste backfill. Geotechnical and Geological Engineering. 2018;36(4):2261-72.
69. Been K, Brown E, Hepworth N. Liquefaction potential of paste fill at Neves Corvo mine, Portugal. Mining Technology. 2002;111(1):47-58.
70. Le Roux K, Bawden W, Grabinsky M. Liquefaction analysis of early age cemented paste backfill. Proceedings of the 8th International Symposium on Mining with Backfill, Beijing, China2004. p. 19-21.
71. Brackebusch F. Basics of paste backfill systems. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts1995. p. 122A.
72. Landriault D. Paste backfill mix design for Canadian underground hard rock mining. 97th Annual General Meeting of CIM Rock Mechanics and Strata Control Session Halifax, Nova Scotia1995. p. 239-8.
73. Grice T. Underground mining with backfill. 2nd Annual Summit œ Mine Tailings Disposal Systems, Brisbane, Nov. 1998:24-5.
74. Testing ASf, Materials. ASTM C143–15 Standard Test. Method for Slump of Hydraulic-Cement Concrete. American Society for Testing and Materials West Conshohocken, PA, USA; 2010.
75. Murata J. Flow and deformation of fresh concrete. Materiaux et Construction. 1984;17(2):117-29.
76. Christensen G. Modelling the flow of fresh concrete: the slump test. Princeton University; 1991.
77. Pashias N, Boger D, Summers J, Glenister D. A fifty cent rheometer for yield stress measurement. Journal of Rheology. 1996;40(6):1179-89.
78. Jewell RJ, Fourie AB. Paste and thickened tailings: a guide. Australian Centre for Geomechanics, The University of Western Australia; 2006.

79. Sellgren A, Addie G, Whitlock L. Technical-economical feasibility of using centrifugal pumps in high-density thickened tailings slurry systems. Paste 2005: Proceedings of the International Seminar on Paste and Thickened Tailings: Australian Centre for Geomechanics; 2005. p. 195-204.
80. Naylor J, Farmery R, Tenbergen R. Paste backfill at the Macassa mine with flash paste production in a paste production and storage mechanism. Proceedings 29th annual meeting of the Canadian mineral processors, Ottawa, Ontario, January1997. p. 21-3.
81. De Souza E, Archibald J, Dirige A. Economics and perspectives of underground backfill practices in Canadian mining. 105th annual general meeting of the Canadian institute of mining, Metallurgy and Petroleum, Montreal, Canada, 15p2003.
82. Fall M, Benzaazoua M. Advances in predicting performance properties and cost of paste backfill. Talings and mine waste'032003. p. 73-85.
83. Belem T, Benzaazoua M. Design and application of underground mine paste backfill technology. Geotechnical and Geological Engineering. 2008;26(2):147-74.
84. Hane I, Belem T, Benzaazoua M, Maqsoud A. Laboratory characterization of cemented tailings paste containing crushed waste rocks for improved compressive strength development. Geotechnical and Geological Engineering. 2017;35(2):645-62.
85. Kesimal A, Yilmaz E, Ercikdi B, Alp I, Deveci H. Effect of properties of tailings and binder on the short-and long-term strength and stability of cemented paste backfill. Materials Letters. 2005;59(28):3703-9.
86. Orejarena L, Fall M. The use of artificial neural networks to predict the effect of sulphate attack on the strength of cemented paste backfill. Bulletin of engineering geology and the environment. 2010;69(4):659-70.
87. Yilmaz E, Belem T, Benzaazoua M. Effects of curing and stress conditions on hydromechanical, geotechnical and geochemical properties of cemented paste backfill. Engineering Geology. 2014;168:23-37.
88. Ouattara D, Yahia A, Mbonimpa M, Belem T. Effects of superplasticizer on rheological properties of cemented paste backfills. International Journal of Mineral Processing. 2017;161:28-40.
89. Erdogan G, Cigla M, Topal E, Yavuz M. Implementation and comparison of four stope boundary optimisation algorithms in an existing underground mine. International Journal of Mining, Reclamation and Environment. 2017;31(6):389-403.
90. Monjezi M, Khoshalan HA, Varjani AY. Optimisation of open pit blast parameters using genetic algorithm. International Journal of Rock Mechanics and Mining Sciences. 2011;5(48):864-9.
91. Musingwini C. Optimisation in underground mine planning-developments and opportunities. Journal of the Southern African Institute of Mining and Metallurgy. 2016;116(9):809-20.
92. Goodfellow RC, Dimitrakopoulos R. Global optimisation of open pit mining complexes with uncertainty. Applied Soft Computing. 2016;40:292-304.
93. Hajihassani M, Jahed Armaghani D, Kalatehjari R. Applications of particle swarm optimisation in geotechnical engineering: a comprehensive review. Geotechnical and Geological Engineering. 2018;36(2):705-22.
94. Yagiz S, Karahan H. Prediction of hard rock TBM penetration rate using particle swarm optimisation. International Journal of Rock Mechanics and Mining Sciences. 2011;48(3):427-33.
95. Boger DV. Rheology and the resource industries. Chemical Engineering Science. 2009;64(22):4525-36.
96. Testing ASf, Cement MCC-o. Standard test method for compressive strength of hydraulic cement mortars (using 2-in. or [50-mm] cube specimens). ASTM International; 2013.
97. Coello CAC. An introduction to multi-objective particle swarm optimisers. Soft computing in industrial applications. Springer; 2011. p. 3-12.
98. Kennedy J, Eberhart R. Particle swarm optimisation. Proceedings of ICNN'95-International Conference on Neural Networks: IEEE; 1995. p. 1942-8.
99. Moore J, Chapman R, Dozier G. Multiobjective particle swarm optimisation. Proceedings of the 38th annual on Southeast regional conference2000. p. 56-7.
100. Sadrossadat E, Basarir H, Karrech A, Durham R, Fourie A, Bin H. The optimisation of cemented hydraulic backfill mixture design parameters for different strength conditions using artificial intelligence algorithms. International Symposium on Mine Planning & Equipment Selection: Springer; 2019. p. 219-27.
101. Tang M-C. High performance concrete—past, present and future. Proceedings of the international symposium on UHPC, Kassel, Germany2004. p. 3-9.
102. Ahmad S, Rasul M, Adekunle SK, Al-Dulaijan SU, Maslehuddin M, Ali SI. Mechanical properties of steel fiber-reinforced UHPC mixtures exposed to elevated temperature: Effects of exposure duration and fiber content. Composites Part B: Engineering. 2019;168:291-301.
103. Abbas S, Nehdi M, Saleem M. Ultra-high performance concrete: Mechanical performance, durability, sustainability and implementation challenges. International Journal of Concrete Structures and Materials. 2016;10(3):271-95.
104. Schmidt M, Fehling E. Ultra-high-performance concrete: research, development and application in Europe. ACI Special publication. 2005;228:51-78.

105. Talebinejad I, Bassam SA, Iranmanesh A, Shekarchizadeh M. Optimising mix proportions of normal weight reactive powder concrete with strengths of 200–350 MPa. Proceedings of the International Symposium on UHPC, Kassel, Germany2004. p. 133-41.
106. Qu D, Cai X, Chang W. Evaluating the Effects of Steel Fibers on Mechanical Properties of Ultra-High Performance Concrete Using Artificial Neural Networks. Applied Sciences. 2018;8(7):1120.
107. Simon M. Concrete mixture optimisation using statistical methods. United States. Federal Highway Administration. Office of Infrastructure …; 2003.
108. Bezerra MA, Santelli RE, Oliveira EP, Villar LS, Escaleira LA. Response surface methodology (RSM) as a tool for optimisation in analytical chemistry. Talanta. 2008;76(5):965-77.
109. Cornell JA. Experiments with mixtures: designs, models, and the analysis of mixture data. John Wiley & Sons; 2011.
110. Lagergren ES, Snyder KA, Simon MJ. Concrete Mixture Optimisation Using Statistical Mixture Design Methods. 1997.
111. Montgomery DC. Design and analysis of experiments. John wiley & sons; 2017.
112. Sadrossadat E, Basarir H. An evolutionary-based prediction model of the 28-day compressive strength of high-performance concrete containing cementitious materials. Advances in Civil Engineering Materials. 2019;8(3):484-97.
113. Duan J, Asteris PG, Nguyen H, Bui X-N, Moayedi H. A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ICA-XGBoost model. Engineering with Computers. 2020:1-18.
114. Asteris PG, Kolovos KG. Self-compacting concrete strength prediction using surrogate models. Neural Computing and Applications. 2019;31(1):409-24.
115. Sadrossadat E, Basarir H, Luo G, Karrech A, Durham R, Fourie A, et al. Multi-objective mixture design of cemented paste backfill using particle swarm optimisation algorithm. Minerals Engineering. 2020;153:106385.
116. Bui DT, Ghareh S, Moayedi H, Nguyen H. Fine-tuning of neural computing using whale optimisation algorithm for predicting compressive strength of concrete. Engineering with Computers. 2019:1-12.
117. Sun L, Koopialipoor M, Armaghani DJ, Tarinejad R, Tahir MM. Applying a meta-heuristic algorithm to predict and optimise compressive strength of concrete samples. Engineering with Computers. 2019:1-13.
118. Dehghani A, Fischer G, Alahi FN. Strengthening masonry infill panels using engineered cementitious composites. Materials and Structures. 2015;48(1-2):185-204.
119. Belhadj T, Abbassi F, Mistou S, Zghal A. Numerical analyses of tube hydroforming problem using artificial neural networks. International Journal of Material Forming. 2010;3(1):295-8.
120. Dutta S, Samui P, Kim D. Comparison of machine learning techniques to predict compressive strength of concrete. Computers and Concrete. 2018;21(4):463-70.
121. Zhang N, Xiong J, Zhong J, Leatham K. Gaussian process regression method for classification for high-dimensional data with limited samples. 2018 Eighth International Conference on Information Science and Technology (ICIST): IEEE; 2018. p. 358-63.
122. Bishop CM. Pattern recognition and machine learning. springer; 2006.
123. Liu Y, Song Y, Keller J, Bond P, Jiang G. Prediction of concrete corrosion in sewers with hybrid Gaussian processes regression model. Rsc Advances. 2017;7(49):30894-903.
124. Rasmussen CE. Gaussian processes in machine learning. Summer School on Machine Learning: Springer; 2003. p. 63-71.
125. Zhang J, Ma G, Huang Y, Aslani F, Nener B. Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. Construction and Building Materials. 2019;210:713-9.
126. Abbassi F, Belhadj T, Mistou S, Zghal A. Parameter identification of a mechanical ductile damage using Artificial Neural Networks in sheet metal forming. Materials & Design. 2013;45:605-15. doi: 10.1016/j.matdes.2012.09.032.
127. Abbassi F, Ahmad F, Gulzar S, Belhadj T, Karrech A, Choi HS. Design of T-shaped tube hydroforming using finite element and artificial neural network modeling. Journal of Mechanical Science and Technology. 2020;34(3):1129-38.
128. Cybenko G. Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems. 1989;2(4):303-14.
129. Konak A, Coit DW, Smith AE. Multi-objective optimisation using genetic algorithms: A tutorial. Reliability Engineering & System Safety. 2006;91(9):992-1007.
130. Coello CC, Lechuga MS. MOPSO: A proposal for multiple objective particle swarm optimisation. Proceedings of the 2002 Congress on Evolutionary Computation CEC'02 (Cat No 02TH8600): IEEE; 2002. p. 1051-6.
131. Sierra MR, Coello CAC. Improving PSO-based multi-objective optimisation using crowding, mutation and∈-dominance. International conference on evolutionary multi-criterion optimisation: Springer; 2005. p. 505-19.

132. Tripathi PK, Bandyopadhyay S, Pal SK. Multi-objective particle swarm optimisation with time variant inertia and acceleration coefficients. Information sciences. 2007;177(22):5033-49.
133. Wang Y, Yang Y. Particle swarm optimisation with preference order ranking for multi-objective optimisation. Information Sciences. 2009;179(12):1944-59.
134. Coello CAC, Pulido GT, Lechuga MS. Handling multiple objectives with particle swarm optimisation. IEEE Transactions on evolutionary computation. 2004;8(3):256-79.
135. Chen S. Constrained particle swarm optimisation. Matlab central file exchange https://www mathworks com/matlabcentral/fileexchange/25986 (Accessed: 28 April 2016). 2016.
136. Okamura H, Ozawa K, Ouchi M. Self-compacting concrete. Structural concrete. 2000;1(1):3-17.
137. Xie T, Ali MM, Elchalakani M, Visintin P. Modelling fresh and hardened properties of self-compacting concrete containing supplementary cementitious materials using reactive moduli. Construction and Building Materials. 2021;272:121954.
138. Okamura H, Ouchi M. Self-compacting concrete. Journal of advanced concrete technology. 2003;1(1):5-15.
139. EFNARC F. Specification and guidelines for self-compacting concrete. European federation of specialist construction chemicals and concrete system. 2002.
140. Gill AS, Siddique R. Strength and micro-structural properties of self-compacting concrete containing metakaolin and rice husk ash. Construction and Building Materials. 2017;157:51-64. doi: 10.1016/j.conbuildmat.2017.09.088.
141. Holland R, Kurtis K, Kahn L. Effect of different concrete materials on the corrosion of the embedded reinforcing steel. Corrosion of Steel in Concrete Structures. 2016:131-47.
142. Güneyisi E, Gesoğlu M. Properties of self-compacting portland pozzolana and limestone blended cement concretes containing different replacement levels of slag. Materials and Structures. 2011;44(8):1399-410. doi: 10.1617/s11527-011-9706-0.
143. Güneyisi E, Gesoğlu M, Özbay E. Strength and drying shrinkage properties of self-compacting concretes incorporating multi-system blended mineral admixtures. Construction and Building Materials. 2010;24(10):1878-87.
144. Kannan V, Ganesan K. Mechanical properties of self-compacting concrete with binary and ternary cementitious blends of metakaolin and fly ash. Journal of the South African Institution of Civil Engineering. 2014;56(2):97-105.
145. Kannan V, Ganesan K. Synergic Effect of Pozzolanic Materials on the Structural Properties of Self-Compacting Concrete. Arabian Journal for Science and Engineering. 2014;39(4):2601-9. doi: 10.1007/s13369-013-0928-z.
146. Almuwbber O, Haldenwang R, Mbasha W, Masalova I. The influence of variation in cement characteristics on workability and strength of SCC with fly ash and slag additions. Construction and Building Materials. 2018;160:258-67. doi: 10.1016/j.conbuildmat.2017.11.039.
147. Zhao H, Sun W, Wu X, Gao B. The properties of the self-compacting concrete with fly ash and ground granulated blast furnace slag mineral admixtures. Journal of Cleaner Production. 2015;95:66-74. doi: 10.1016/j.jclepro.2015.02.050.
148. Nikbin IM, Beygi MHA, Kazemi MT, Vaseghi Amiri J, Rabbanifar S, Rahmani E, et al. A comprehensive investigation into the effect of water to cement ratio and powder content on mechanical properties of self-compacting concrete. Construction and Building Materials. 2014;57:69-80. doi: 10.1016/j.conbuildmat.2014.01.098.
149. Heckert NA, Filliben JJ, Croarkin CM, Hembree B, Guthrie WF, Tobias P, et al. Handbook 151: NIST/SEMATECH e-Handbook of Statistical Methods. 2002.
150. Hinkelmann K. Design and analysis of experiments, volume 3: special designs and applications. John Wiley & Sons; 2012.
151. Ting MZY, Wong KS, Rahman ME, Selowarajoo M. Prediction model for hardened state properties of silica fume and fly ash based seawater concrete incorporating silicomanganese slag. Journal of Building Engineering. 2021;41:102356.
152. Güneyisi E, Gesoğlu M, Algın Z, Mermerdaş K. Optimisation of concrete mixture with hybrid blends of metakaolin and fly ash using response surface method. Composites Part B: Engineering. 2014;60:707-15.
153. Mermerdas K, Guneyisi E, Gesoglu M, Algin Z. Optimisation of concrete mixture with hybrid blends of metakaolin and fly ash using response surface method. 2014.
154. Elemam WE, Abdelraheem AH, Mahdy MG, Tahwia AM. Optimising fresh properties and compressive strength of self-consolidating concrete. Construction and Building Materials. 2020;249. doi: 10.1016/j.conbuildmat.2020.118781.
155. Cibilakshmi G, Jegan J. A DOE approach to optimise the strength properties of concrete incorporated with different ratios of PVA fibre and nano-Fe2O3. Advanced Composites Letters. 2020;29:2633366X20913882.

156. Gandomi AH, Alavi AH, Ting T, Yang X-S. Intelligent modeling and prediction of elastic modulus of concrete strength via gene expression programming. International conference in swarm intelligence: Springer; 2013. p. 564-71.
157. Armaghani DJ, Asteris PG. A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength. Neural Computing and Applications. 2020:1-32.
158. Chen C-T, Gu GX. Machine learning for composite materials. MRS Communications. 2019;9(2):556-66.
159. Siddique R, Aggarwal P, Aggarwal Y, Gupta S. Modeling properties of self-compacting concrete: support vector machines approach. Computers and Concrete. 2008;5(5):123-9.
160. Abbassi F, Ahmad F, Gulzar S, Belhadj T, Karrech A, Choi HS. Design of T-shaped tube hydroforming using finite element and artificial neural network modeling. Journal of Mechanical Science & Technology. 2020;34(3).
161. Alavi AH, Ameri M, Gandomi AH, Mirzahosseini MR. Formulation of flow number of asphalt mixes using a hybrid computational method. Construction and Building Materials. 2011;25(3):1338-55.
162. Sadrossadat E, Basarir H, Karrech A, Elchalakani M. Multi-objective mixture design and optimisation of steel fiber reinforced UHPC using machine learning algorithms and metaheuristics. Engineering with Computers. 2021:1-14.
163. Gandomi AH, Yang X-S, Talatahari S, Alavi AH. Metaheuristic applications in structures and infrastructures. Newnes; 2013.
164. Yang X-S. Engineering optimisation: an introduction with metaheuristic applications. John Wiley & Sons; 2010.
165. Zhang J, Huang Y, Ma G, Nener B. Mixture optimisation for environmental, economical and mechanical objectives in silica fume concrete: A novel frame-work based on machine learning and a new meta-heuristic algorithm. Resources, Conservation and Recycling. 2021;167:105395.
166. Arora JS. Introduction to optimum design. Elsevier; 2004.
167. Uysal M, Sumer M. Performance of self-compacting concrete containing different mineral admixtures. Construction and Building Materials. 2011;25(11):4112-20. doi: 10.1016/j.conbuildmat.2011.04.032.
168. Vivek SS, Dhinakaran G. Fresh and hardened properties of binary blend high strength self compacting concrete. Engineering Science and Technology, an International Journal. 2017;20(3):1173-9. doi: 10.1016/j.jestch.2017.05.003.
169. Li M, Hao H, Shi Y, Hao Y. Specimen shape and size effects on the concrete compressive strength under static and dynamic tests. Construction and Building Materials. 2018;161:84-93.
170. Schlexer Lamoureux P, Winther KT, Garrido Torres JA, Streibel V, Zhao M, Bajdich M, et al. Machine learning for computational heterogeneous catalysis. ChemCatChem. 2019;11(16):3581-601.
171. Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20(3):273-97.
172. Li L-L, Zhao X, Tseng M-L, Tan RR. Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. Journal of Cleaner Production. 2020;242:118447.
173. Shokry A, Audino F, Vicente P, Escudero G, Moya MP, Graells M, et al. Modeling and simulation of complex nonlinear dynamic processes using data based models: application to photo-Fenton process. Computer Aided Chemical Engineering. Elsevier; 2015. p. 191-6.
174. Liu W, Principe JC, Haykin S. Kernel adaptive filtering: a comprehensive introduction. John Wiley & Sons; 2011.
175. Rahla KM, Mateus R, Bragança L. Comparative sustainability assessment of binary blended concretes using Supplementary Cementitious Materials (SCMs) and Ordinary Portland Cement (OPC). Journal of Cleaner Production. 2019;220:445-59.
176. Goldberg DE, Holland JH. Genetic algorithms and machine learning. 1988.
177. Coello CAC, Lamont GB. Applications of multi-objective evolutionary algorithms. World Scientific; 2004.
178. Deb K. Multi-objective optimisation using evolutionary algorithms: an introduction. Multi-objective evolutionary optimisation for product design and manufacturing. Springer; 2011. p. 3-34.
179. Wang L, Ng AH, Deb K. Multi-objective evolutionary optimisation for product design and manufacturing. Springer; 2011.
180. Long G, Gao Y, Xie Y. Designing more sustainable and greener self-compacting concrete. Construction and Building Materials. 2015;84:301-6. doi: 10.1016/j.conbuildmat.2015.02.072.
181. Pacewska B, Wilińska I. Usage of supplementary cementitious materials: advantages and limitations. Journal of Thermal Analysis and Calorimetry. 2020;142(1):371-93.
182. Harrisson AM. Constitution and Specification of Portland Cement. Lea's Chemistry of Cement and Concrete. 2019:87-155.
183. L'Hôpital E, Lothenbach B, Le Saout G, Kulik D, Scrivener K. Incorporation of aluminium in calcium-silicate-hydrates. Cement and Concrete Research. 2015;75:91-103.
184. Pal S, Mukherjee A, Pathak S. Investigation of hydraulic activity of ground granulated blast furnace slag in concrete. Cement and Concrete Research. 2003;33(9):1481-6.

185. Saleh Ahari R, Erdem TK, Ramyar K. Permeability properties of self-consolidating concrete containing various supplementary cementitious materials. Construction and Building Materials. 2015;79:326-36. doi: 10.1016/j.conbuildmat.2015.01.053.
186. Lee J, Lee T. Durability and engineering performance evaluation of CaO content and ratio of binary blended concrete containing ground granulated blast-furnace slag. Applied Sciences. 2020;10(7):2504.
187. Siddique R, Bennacer R. Use of iron and steel industry by-product (GGBS) in cement paste and mortar. Resources, Conservation and recycling. 2012;69:29-34.
188. Dadsetan S, Bai J. Mechanical and microstructural properties of self-compacting concrete blended with metakaolin, ground granulated blast-furnace slag and fly ash. Construction and Building Materials. 2017;146:658-67. doi: 10.1016/j.conbuildmat.2017.04.158.
189. Dinakar P, Sethy KP, Sahoo UC. Design of self-compacting concrete with ground granulated blast furnace slag. Materials & Design. 2013;43:161-9. doi: 10.1016/j.matdes.2012.06.049.
190. Ahmed T, Elchalakani M, Basarir H, Karrech A, Sadrossadat E, Yang B. Development of ECO-UHPC utilizing gold mine tailings as quartz sand alternative. Cleaner Engineering and Technology. 2021:100176.
191. ASTM International WC, PA. ASTM C618-19, Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete,. 2019.
192. Xie T, Visintin P. A unified approach for mix design of concrete containing supplementary cementitious materials based on reactivity moduli. Journal of Cleaner Production. 2018;203:68-82.
193. Nanthagopalan P, Santhanam M. An empirical approach for the optimisation of aggregate combinations for self-compacting concrete. Materials and structures. 2012;45(8):1167-79.
194. Bani Ardalan R, Joshaghani A, Hooton RD. Workability retention and compressive strength of self-compacting concrete incorporating pumice powder and silica fume. Construction and Building Materials. 2017;134:116-22. doi: 10.1016/j.conbuildmat.2016.12.090.
195. Bonavetti V, Donza H, Menendez G, Cabrera O, Irassar E. Limestone filler cement in low w/c concrete: A rational use of energy. Cement and Concrete Research. 2003;33(6):865-71.
196. Moosberg-Bustnes H, Lagerblad B, Forssberg E. The function of fillers in concrete. Materials and Structures. 2004;37(2):74-81.
197. Ilg M, Plank J. Synthesis and properties of a polycarboxylate superplasticizer with a jellyfish-like structure comprising hyperbranched polyglycerols. Industrial & Engineering Chemistry Research. 2019;58(29):12913-26.
198. Houst YF, Bowen P, Perche F, Kauppi A, Borget P, Galmiche L, et al. Design and function of novel superplasticizers for more durable high performance concrete (superplast project). Cement and Concrete Research. 2008;38(10):1197-209.
199. ASTM International WC, PA. Standard Specification for Chemical Admixtures for Concrete. ASTM C494 / C494M-192019.
200. Xie T, Ali MSM, Elchalakani M, Visintin P. Modelling fresh and hardened properties of self-compacting concrete containing supplementary cementitious materials using reactive moduli. Construction and Building Materials. 2021;272:121954.
201. Meko B, Ighalo JO, Ofuyatan OM. Enhancement of self-compactability of fresh self-compacting concrete: A review. Cleaner Materials. 2021;1:100019.
202. Onyelowe KC, Ebid AM, Nwobia LI. Predictive models of volumetric stability (durability) and erodibility of lateritic soil treated with different nanotextured bio-ashes with application of loss of strength on immersion; GP, ANN and EPR performance study. Cleaner Materials. 2021;1:100006.
203. Ilyas I, Zafar A, Javed MF, Farooq F, Aslam F, Musarat MA, et al. Forecasting Strength of CFRP Confined Concrete Using Multi Expression Programming. Materials. 2021;14(23):7134.
204. Javed MF, Amin MN, Shah MI, Khan K, Iftikhar B, Farooq F, et al. Applications of gene expression programming and regression techniques for estimating compressive strength of bagasse ash based concrete. Crystals. 2020;10(9):737.
205. Farooq F, Ahmed W, Akbar A, Aslam F, Alyousef R. Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimisation of models using ensemble learners. Journal of Cleaner Production. 2021;292:126032.
206. Farooq F, Nasir Amin M, Khan K, Rehan Sadiq M, Faisal Javed M, Aslam F, et al. A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC). Applied Sciences. 2020;10(20):7330.
207. Nafees A, Javed MF, Khan S, Nazir K, Farooq F, Aslam F, et al. Predictive Modeling of Mechanical Properties of Silica Fume-Based Green Concrete Using Artificial Intelligence Approaches: MLPNN, ANFIS, and GEP. Materials. 2021;14(24):7531.
208. Aslam F, Farooq F, Amin MN, Khan K, Waheed A, Akbar A, et al. Applications of gene expression programming for estimating compressive strength of high-strength concrete. Advances in Civil Engineering. 2020;2020.

209. Farooq F, Czarnecki S, Niewiadomski P, Aslam F, Alabduljabbar H, Ostrowski KA, et al. A comparative study for the prediction of the compressive strength of self-compacting concrete modified with fly ash. Materials. 2021;14(17):4934.
210. Chou J-S, Chiu C-K, Farfoura M, Al-Taharwa I. Optimising the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. Journal of Computing in Civil Engineering. 2011;25(3):242-53.
211. Vapnik VN. An overview of statistical learning theory. IEEE transactions on neural networks. 1999;10(5):988-99.
212. Heidaripanah A, Nazemi M, Soltani F. Prediction of resilient modulus of lime-treated subgrade soil using different kernels of support vector machine. International Journal of Geomechanics. 2017;17(2):06016020.
213. Bingöl AF, Tohumcu İ. Effects of different curing regimes on the compressive strength properties of self compacting concrete incorporating fly ash and silica fume. Materials & Design. 2013;51:12-8. doi: 10.1016/j.matdes.2013.03.106.
214. Esen Y, Orhan E. Investigation of the effect on the physical and mechanical properties of the dosage of additive in self-consolidating concrete. KSCE Journal of Civil Engineering. 2016;20(7):2849-58. doi: 10.1007/s12205-016-0258-2.
215. Kannan V. Strength and durability performance of self compacting concrete containing self-combusted rice husk ash and metakaolin. Construction and Building Materials. 2018;160:169-79. doi: 10.1016/j.conbuildmat.2017.11.043.
216. Le HT, Ludwig H-M. Effect of rice husk ash and other mineral admixtures on properties of self-compacting high performance concrete. Materials & Design. 2016;89:156-66. doi: 10.1016/j.matdes.2015.09.120.
217. Liu M. Self-compacting concrete with different levels of pulverized fuel ash. Construction and Building Materials. 2010;24(7):1245-52. doi: 10.1016/j.conbuildmat.2009.12.012.
218. Niknezhad D, Kamali-Bernard S, Mesbah H-A. Self-Compacting Concretes with Supplementary Cementitious Materials: Shrinkage and Cracking Tendency. Journal of Materials in Civil Engineering. 2017;29(7). doi: 10.1061/(asce)mt.1943-5533.0001852.
219. Turk K, Karatas M, Gonen T. Effect of Fly Ash and Silica Fume on compressive strength, sorptivity and carbonation of SCC. KSCE Journal of Civil Engineering. 2012;17(1):202-9. doi: 10.1007/s12205-013-1680-3.
220. Uysal M. The influence of coarse aggregate type on mechanical properties of fly ash additive self-compacting concrete. Construction and Building Materials. 2012;37:533-40. doi: 10.1016/j.conbuildmat.2012.07.085.
221. Anjos MAS, Camões A, Campos P, Azeredo GA, Ferreira RLS. Effect of high volume fly ash and metakaolin with and without hydrated lime on the properties of self-compacting concrete. Journal of Building Engineering. 2020;27. doi: 10.1016/j.jobe.2019.100985.
222. Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, et al. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling & Software. 2016;79:214-32.