APPLIED COMPUTATIONAL ANALYSES FOR CONCRETE COMPRESSIVE STRENGTH PERFORMANCE ASSESSMENT

by

Uchenna Emmanuel Anyaoha

B. Eng., Nnamdi Azikiwe University, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES
(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Okanagan)

September 2019

The following individuals certify that they have read, and recommend to the College of Graduate Studies for acceptance, the thesis entitled:

APPLIED COMPUTATIONAL ANALYSES FOR CONCRETE COMPRESSIVE STRENGTH PERFORMANCE ASSESSMENT

submitted by Uchenna Emmanuel Anyaoha in partial fulfillment of the requirements of
the degree of Master of Applied Science .

Dr. Zheng Liu, School of Engineering
Supervisor
Dr. Kasun Hewage, School of Engineering
Supervisory Committee Member
Dr. Seach Chyr (Ernest) Goh, School of Engineering
Supervisory Committee Member
Dr. Joshua Brinkerhoff, School of Engineering
University Examiner

Abstract

The compressive strength of concrete represents a significant and perhaps the most essential mechanical property, measured after a standard curing process for a given concrete mixture produced. Traditional methods such as core cuts and other destructive test methods are used in assessing the strength of concrete. However, concrete strength is influenced by numerous conditions and factors such as its material constituents, mixture designs and the ratio of these material constituents. Additionally, different environmental exposures as well may contribute to the complex problem in producing as well as in estimating the compressive strength of concrete by these traditional methods.

The use of the statistical approach for concrete mixture designs, advanced computational and machine learning approaches for estimating the compressive strength of concrete based on mixture proportions, on account of its industrial importance, has received significant attention. However, previous studies have been limited to a single source and small laboratory-produced data sets used for the analysis of concrete properties. If the adequate nonlinear function is found for several categories and designs of concrete compressive strength condition assessments, then the prediction of concrete strength may tend to become automated. Thereby, reducing the number of destructive testing done in the concrete laboratory as well as reducing the production cost of concrete. In this thesis, some soft computational techniques were employed to address these challenges.

The first study investigates the possibility to explore concrete mixture design to produce favorable and optimal compressive strength results for situations where some experimental tests may be difficult to run due to challenges in obtaining certain constituents which may be expensive or not readily available. With intentions

to then reduce the need for preparing a large number of trial mixes to avoid material wastage, a simple statistical approach for concrete mixture design via a response surface method was proposed to overcome this drawback. In using experimental mixture design data specified for concrete of high performance, the optimization study demonstrated that the proposed method serves as a promising design approach in the concrete domain.

The second study investigates the effects of concrete constituents and the constituents mix proportioning in estimating the compressive strength of concrete. A section of this study further tests the sensitivity analysis on concrete relative to the variables used in evaluating its compressive strength. Firstly, measurements of the linear relationships between a series of variables with the compressive strength of concrete were obtained using their correlation coefficients. After that, an ensemble meta-algorithm was employed to investigate the performance of concrete compressive strength with considered concrete features. Based on the prediction performance, the scalability and performance of the ensemble meta-algorithm were validated.

Lay Summary

The assessment of concrete compressive strength through some traditional methods like concrete coring and the excessive destructive testings before and after construction account for an increase in construction costs as well as a waste of resources in some cases. Thus, it is vital to provide complementary or alternative adequate compressive strength assessments for concrete via computational modeling. In this thesis, the presented research investigates the mixture design of concrete using statistical Box-Behnken design of response surface method, for conditions where concrete trial tests for experiments may be challenging or expensive to run. Secondly, machine learning based approaches for concrete compressive strength performance were carried out to investigate the effect of concrete constituent and their mixture proportioning in estimating concrete compressive strength. Computational results demonstrate the accuracy of the proposed statistical and computational models in both studies.

Preface

In God we trust, all others bring data.

–William Edwards Deming (1900 - 1993)

This thesis is based on the research work completed in the School of Engineering at the University of British Columbia, Okanagan, under the supervision and guidance of Dr. Zheng Liu. All published works are included in this thesis.

Chapter 2, 3 and 4 are based on the versions of the following submitted papers for potential journal publications and a conference paper used with permission of SPIE:

- **Uchenna Anyaoha**, Xiang Peng and Zheng Liu. "Concrete performance prediction using boosting smooth transition regression trees (BooST)." In *Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XIII*, volume: 10971, pages: 109710I). International Society for Optics and Photonics, 2019. (©2019 SPIE)

- **Uchenna Anyaoha** and Zheng Liu. "Soft Computing in Estimating the Compressive Strength of High-performance Concrete via Concrete Composition Appraisal." Submitted to *Construction and Building Materials Journal.* (2019).

I am the principal contributor for these work. Prof. Zheng Liu provided me with some advice on research methodology and experiment design. Mr. Xiang Peng and Mr. Qiwen Jin helped me proofread the manuscripts.

Table of Contents

List of Tables

List of Figures

Glossary

ACI	American Concrete Institute
ANOVA	Analysis of Variance
ANN	Artificial Neural Networks
ASCE	American Society of Civil Engineers
ASTM	American Society for Testing and Materials
Bfs	Blast Furnace Slag
BooST	Boosting Smooth Transition Regression Trees
BBD	Box-Behnken Design
C	Cement
Q_{CmT}	Cementitious content
CA	Coarse Aggregate
CART	Classification and Regression Trees
CCS	Concrete Compressive Strength
DOE	Design of Experiment
FA/T_A	Fine Aggregate-to-Total Aggregate ratio
FA	Fine Aggregate

$\mathbf{F}_{ash}$	Fly Ash
HPC	High Performance Concrete
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
NLR	Non-linear Regression
R^2	Coefficient of Determination
$R^2{}_{adj}$	Adjusted Coefficient of Determination
$R^2{}_{pred}$	Predicted Coefficient of Determination
RMSE	Root Mean Square Error
RSM	Response Surface Methodology
RT	Regression Tree
SF	Silica Fume
SS_E	Sum of Squares (error) or Residual Sum of Squares
SS_R	Sum of Squares (regression)
Su	Superplasticizer
SVM	Support Vector Machine
W	Water
W/C	Water-to-Cement ratio
W/C_{mt}	Water-to-Cementitious Material Content ratio

Acknowledgments

There is no iota of doubt to the academic knowledge I have received from my supervisor Dr. Zheng Liu, and it is with this that I applaud resounding gratitude to him. His educational guidance, generosity, kindness and encouragement throughout my graduate study have been impeccable. Under his leadership, I have attained invaluable research skills as well as obtaining resounding professional skills. Indeed he is a right mentor, and I take it as an honour to have studied and worked under his guidance.

Thanks to my committee members Dr. Kasun Hewage and Dr. Seach Chyr (Ernest) Goh for their valuable time and willingness to serve on the supervisory committee, and to Dr. Joshua Brinkerhoff for his disposition to serve as the university examiner.

Additionally, I offer appreciation to all my colleagues at the Intelligent Sensing, Diagnostic and Prognostic Research Lab (ISDPRL), who have provided encouragement and support throughout my study. Furthermore, this note of appreciation goes to the peer review group consisting of Xiang, Dylan and Amin who helped review my thesis. To this vote of thanks, I also include friends I made both by working in student leadership activities at the university as well as those who showed concern about my welfare - Arnab, Ashley, Kalie, Remy, Ben, Jenny, Yushan, Tanjid, Nibirh and Connor.

Many thanks to the leadership of the College of Graduate Studies for their immense support in improving the lifestyle and academic standard of graduate studies throughout my research study. I would also like to acknowledge the School of Engineering Administrative Assistants, Ms. Shannon Hohl, of the School of Engineering for their ubiquitous impacts and help provided to graduate students during

stressful times via effective communication and management.

Finally, I appreciate my family for encouraging me in all of my pursuits and inspiring me to follow my dreams. I am especially grateful to my parents, who supported me emotionally and financially. Thank you for teaching me that all I have to do is be happy while I make others happy. Thanks to my mother, Mrs. Agatha Anyaoha, for praying for me and transferring to me the entrepreneurial spirit. Thanks to my father, Engr. Joseph Anyaoha, for praying for me and speaking through to me words of wisdom. And above all acknowledgements mentioned above, I praise and thank the good LORD who guides my steps and prepare my ways on Earth.

Chapter 1

Introduction

1.1 Background and Motivation

The compressive strength of concrete is considered a vital mechanical property of concrete, which plays a crucial role in the service life of concrete materials. As the most widely used construction material on earth, concrete requires enormous maintenance throughout its design life cycle [9]. Although, concrete could be in different states; plastic and malleable when newly mixed and strong and durable when hardened, the use of concrete technology helps provide desired outcomes for the properties of concrete [10].

Concrete technology deals with the study of properties of concrete and its practical applications when in different states. Though these states of concrete may be considered as the following [11]:

1. *Plastic*: This state of concrete is considered the plastic state, which refers to concrete when wet. It comprises of the initial stage when concrete is made out of aggregate (fine and coarse), water and cement mixed to form a malleable building material. At this state, concrete can be poured into retaining walls, foundation footings, underground trenches as well as any desired shape for specific projects. Such property of concrete makes it maneuverable compared to other building materials for construction.

2. *Curing*: To help concrete develop strength and durability, the second state

is considered. This curing state takes place after the concrete has been carefully placed and finished. Concrete curing is considered as it comprises of maintaining adequate moisture and temperature controlled over an extended period, typically 28 days [12], but may vary depending on mixture proportions, desired strength and other environmental conditions.

3. *Hardened*: This state of concrete comes right after the curing state, to serve the purpose for which the concrete was produced. When concrete is developed and hardened, its specified strength can be deduced, from which several tests are carried out for analysis. It is the hardened state of concrete that this thesis focuses on. A layout of concrete states and strength formation is shown in Figure 1.1.

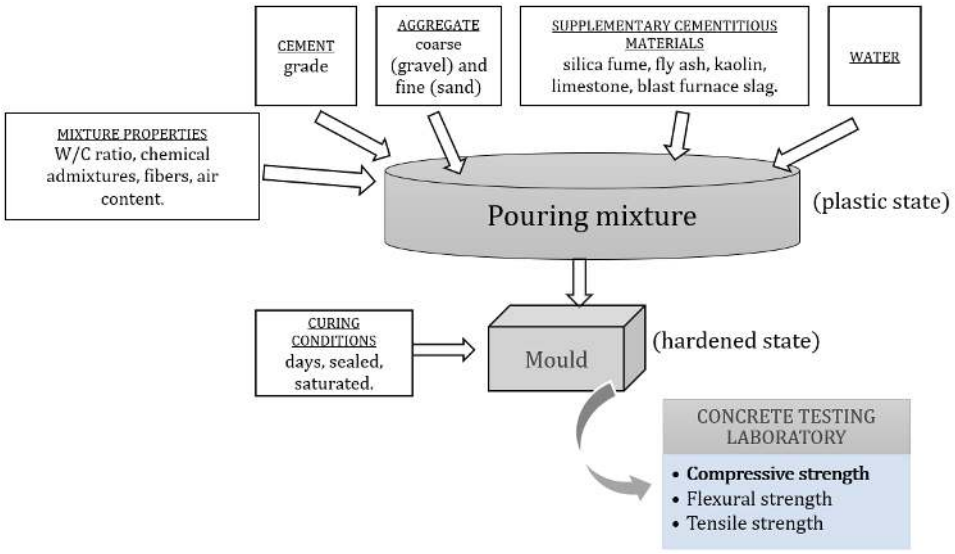


Figure 1.1. Layout of concrete compressive strength formation.

In recent research on concrete, the Concrete Compressive Strength (**CCS**) happens to be of critical importance and also possesses difficulty in assessment due to the complexity of material constituents as well as other conditions, like; mix proportions and working environment/condition of concrete etc. Though several standards and codes such as the American Concrete Institute (**ACI**) 562-16 have

been provided to help with rehabilitation, repair and assessment of concrete structures, studies have shown that concrete is a highly complex material which is difficult to understand [13, 14]. Furthermore, it is well known that Water-to-Cement ratio (**W/C**) chiefly influences CCS. However, even for a given W/C, an enormous variation in concrete properties (primarily compressive strength) may be observed for High Performance Concrete (**HPC**) which consists of several material constituents (cement type, minerals and chemical admixtures, etc.) and mixture proportioning. ACI defines HPC as "a concrete meeting special combinations of performance and uniformity requirements that cannot always be achieved routinely using conventional constituents and normal mixing [15]. Since HPC needs to meet special requirements that cannot be achieved by conventional materials, mixture proportioning and curing practices like the ordinary concrete, therefore, a reasonable choice of material constituents and chemical admixtures can reduce excessive waste of materials which would result in optimal CCS and economical HPC.

With complexity in concrete constituents in recent years, the use of data-driven computational methods as well as machine learning techniques are employed in modeling the properties of concrete [16]. Machine learning is a branch of computer science that can be used for pattern recognition and in understanding complicated relationships among features of data to make accurate decision or predictions [17]. Nonetheless, previous studies in concrete have been limited to small, laboratory dataset with a limited range of material constitutes as input variables, thereby not completely assessing the capability of any technique used in estimating concrete properties. Therefore, as a motivation for improving automation in the concrete domain, it is imperative to assess concrete within a wide range of material constituents using advanced machine learning techniques to estimate the CCS for use in construction. Such assessment would save overall cost as a result of trial and error tests in the concrete laboratories as well as provide high estimation and accuracy. Additionally, a better understanding of the non-linear regression problem related to CCS in the civil engineering domain would be further explored with new computational techniques and experimental settings.

1.2 Research Objectives

The primary objective of this thesis was to assess the accurate estimation of concrete compressive strength using the knowledge of its material constituents while identifying mixture compositions that results in optimal compressive strength. Firstly in the study was the emphasis laid on the design of concrete experiment for optimal strength, which considers a limited number of concrete data to provide mixture designs for optimal CCS. Secondly, a wide range of concrete mixture constituents and mixture proportioning were modeled with state-of-the-art machine learning techniques in predicting the compressive strength of concrete. Furthermore, a new machine learning technique was introduced to compare estimated results with the state-of-the-art techniques.

To further bolster the significance of these studies, the chapters and objectives of this thesis are arranged in a logical order to promote the completion and demonstration of the primary objective. Specific sub-objectives are outlined below:

- Perform an in-depth literature review on design optimization of concrete mixture for optimal strength and on predicting the compressive strength of concrete using mixture constituents and mixture proportioning. The results from these reviews provided insights for the need to proportion concrete mixtures that satisfy concrete strength performance requirements.

- Design of concrete experiment for optimal compressive strength with limited available data.

- Analyze the influence of concrete constituent, its mix proportioning (ratio) and weighing their importance in improving predictability for CCS while investigating adequate models for estimating CCS for high performance using several proportioning of the concrete constituents as explanatory variables. Additionally, tests for sensitivity analysis were investigated, in order to confirm concrete responsiveness with change in explanatory variables.

1.3 Thesis Outline

This thesis is organized into five chapters. Following the Introduction in Chapter 1 is the review in Chapter 2, which is divided into two sections. The first section discusses common practices of concrete mixture design leading to better performance for optimal compressive strength. The second section of Chapter 2 outlines the estimation/prediction of concrete compressive strength for high performance, the existing methods, identifies the knowledge gap and emphasizes the need for enhancing techniques in predicting CCS. Chapter 3 describes procedures of the proposed techniques for each challenge described as well their respective computational analysis. Schematics of these procedures are depicted respectively in Figure 1.2 and Figure 1.3. In Chapter 4, interpretation of computational results as well as discussions for each study in this thesis were presented. The conclusion in Chapter 5 comprises of the contributions and recommendations for future research studies.

In predicting the CCS as shown in Figure 1.3 consists of the proposed machine learning technique compared with other contemporary techniques. The steps consist of data preparation obtained from a created database of 28 days CCS. Furthermore, the data is divided into training and testing set where they are trained and predicted results provided for comparison with other contemporary techniques. Satisfactory performance model created the most efficient technique is then presented for implementation.

A flowchart demonstrating the thesis outline and relationships between thesis chapters and objectives is presented in Figure 1.4.

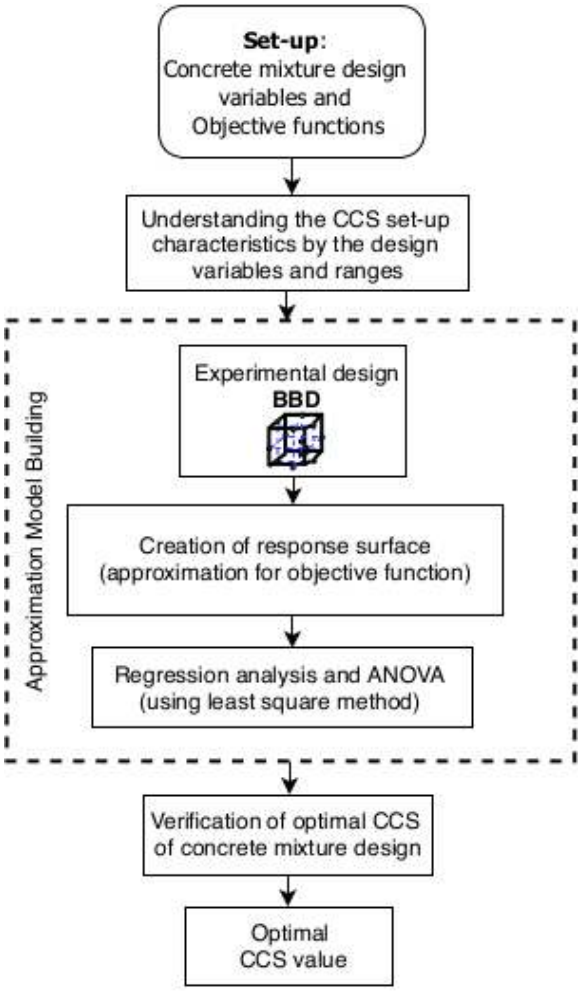


Figure 1.2. Graphical representation of proposed method for the concrete mixture design for an optimal compressive strength.

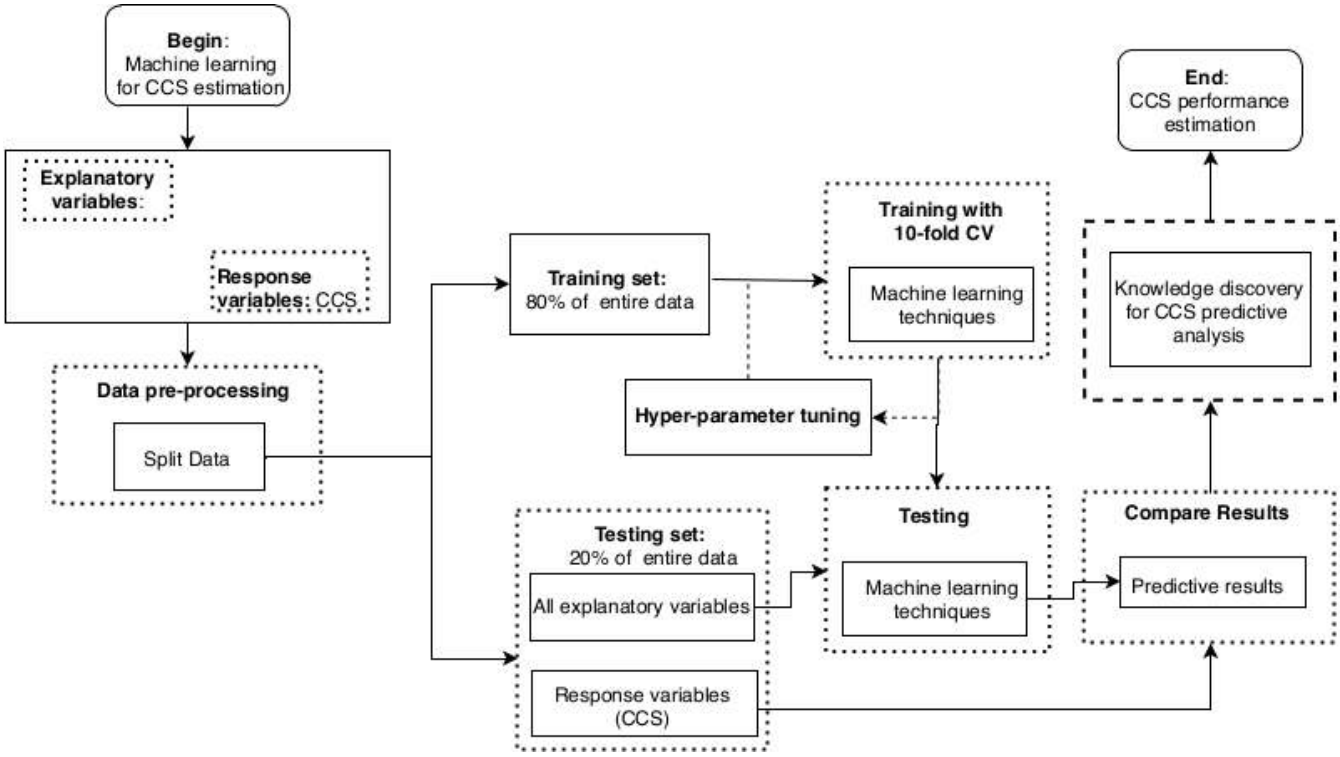


Figure 1.3. Schematic representation of proposed method for the prediction of concrete compressive strength.

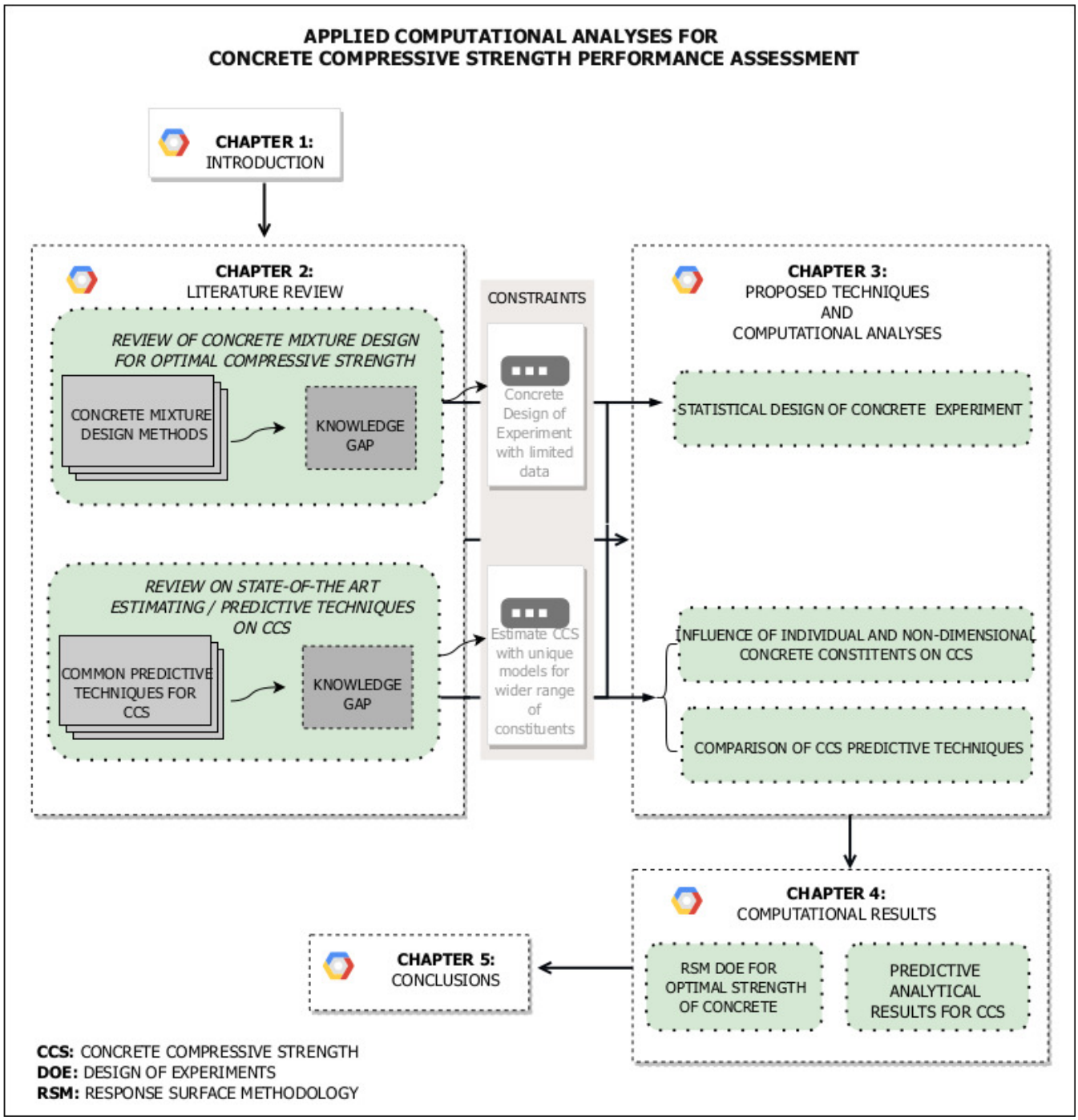


Figure 1.4. Thesis flowchart.

Chapter 2

Common Practices and State-of-the-Art

This chapter is divided into two broad topics of review. The first topic (Section 2.1) discusses review on common practices of concrete mixture design methods for optimal compressive strength. The second topic (Section 2.2) is a review of common contemporary techniques and their working principles used in estimating/predicting concrete compressive strength with knowledge of concrete constituents.

In a subject-specific database, the common practice for review articles is to search for key areas of interest [18–21]. A methodical framework for these reviews were developed with a key area of interest focusing on concrete of high performance and compressive strength. Hence literature reviews were conducted using "Compendex engineering village, Scopus, Science Direct, Google Scholar, UBC library, and American Society of Civil Engineers (**ASCE**) library.

The review of both topics was necessary to identify knowledge gaps in existing concrete mixture design for optimal compressive strength, as well as to determine adequate and efficient techniques in predicting concrete compressive strength. Hence, relevant studies from the databases mentioned were sorted and considered. The reviews also considered documents with reasonable PlumX metrics within the Compendex engineering village. The PlumX metrics can be used to provide perspicacity of people's interaction with an individual piece of a research article in the online environment, thereby determining the reach or impact of such

document [22].

2.1 Concrete Mixture Design Methods for Optimal Compressive Strength

Introduction

The production of concrete with several desired compressive strengths have been a challenge in concrete mixture design. The mixture design of concrete is the selection of material constituents in optimum proportions to produce desired properties during fresh and hardened states for specific applications [23]. With emphasis to concrete on high performance, which is distinct from conventional concrete, the mechanical properties are of paramount importance. Early studies of HPC in 1990, reported that factors including compositions of raw materials, water to cement ratio (W/C), aggregate type, chemical and mineral admixtures have significant effects on the CCS [2].

As high strength is required for construction of high-rise buildings, tunnels, highways and nuclear structures, optimizing material constituents of concrete has received significant attention [24]. Aïtcin [25] reported that Superplasticizer (**Su**) like naphthalene sulfonate, melamine and lignosulfonate aid in the reduction of (W/C) which after that improves the CCS. With the recent paradigm shift towards creating a more sustainable environment, waste materials such as Fine Aggregate (**FA**), oil palm shell, waste glass powder and recycled tire rubber are being used as substitute material constituents for concrete to achieve a high strength [26–29]. However, the incorporation of several mixture design methods further contributes to improved CCS [30, 31].

Conventional Practices for Concrete Design Methods

Over previous decades, a variety of methods have been proposed to design HPC [32]. Swamy [33, 34] presented several designs of HPC in 1996 and 1997, respectively for increasing the durability and strength for HPC using the synergic interactions of material constituents.

Alternatively, a modified mixture proportion design method based on the ACI

method for normal concrete mix design [32], which took into account the efficiency factor of supplementary cementitious materials was proposed by Bharatkumar et al. [35] in 2001. Results from the method showed the possibility to obtain an economical HPC mixture. The study only considered limited durability properties (after 28 days of curing), which contained the coefficient of absorption and sorptivity. To achieve specific HPC for different application purposes, several forms of concrete mixture design are used. Existing mixture design methods focusing on concrete of high performance and strength in the literature are summarized in Table 2.1. The concrete mixture design methods are classified based on the following principles: empirical design method, close particle packing method, compressive strength method and mixture design method based on statistical model.

Empirical mixture design method

The empirical mixture design method is based on practical experiences involving experiments with combinations of concrete material constituents to determine the initial mixture proportions. In this method, the required or specified concrete properties are derived from best estimates of constituent contents and mixture proportions through trials and errors with necessary adjustments.

Early studies on HPC with high CCS reveal that empirical mixture designs were the common practice in concrete technology. In 1990, Mehta and Aïtcin [2] proposed a mixture design method for producing HPC based on experience. The first trial of their design procedure for concrete with 60 to 120 MPa included the following aspects:

- the range of concrete strength is arbitrarily divided into five strength grades;
- from a given strength grade, an estimate of the water content is selected based on experience with high-slump superplasticized concrete mixtures;
- volume of cement paste components are selected with a progressive increase of cementitious contents in 3 options, starting with:
    1. only the portland cement (C) as the only cementitious content,
    2. C + Fly Ash ($\mathbf{F}_{ash}$) or Blast Furnace Slag (**Bfs**) in 75:25 ratio by volume,

Table 2.1. Summary of conventional mixture design methods for concrete focused on high performance and strength in the literature.

Mixture design classification	Reference	Year	Features
Empirical design method	Mehta and Aïtcin [2]	1990	Use a suggested FA:CA ratio of 2:3, then adjust water and superplasticizer dosage as well as the volume fraction of cementitious components.
	De Larrard [36]	1990	Use marsh cone to measure fluidity of paste, adjusting Su dosage with C corresponding to the least water demand.
	Okamura and Ozawa [37]	1995	Use a fix FA and CA, then adjust water-to-binder ratio and superplasticizer dosage.
	Edamatsa et al. [3]	1999, 2003	Use mortar flow and V-funnel test to select FA content, volumetric water-to-powder ratio and Su dosage.
Close particle packing method	Chang [38]	2004	Use Densified Mixture Design Algorithm (DMDA), derived from the maximum density theory and excess paste theory for HPC
	Su and Miao [5]	2003	Use Packing Factor (PF) to control the FA and CA content in mixture proportion
	De Larrard and Sedran [39]	2002	Design concrete based on Compressive Packing Model (CPM) incorporated in a software
	Ng et al. [40]	2016	Mixture design based on packing density and film thickness theories
Compressive strength method	Dinakar and Manu [6]	2014	Conduct with main phases: fix cementitious content, determine W/C using compressive to W/C ratio curve, combination ratio of FA/CA where maximum aggregate size $\leq$ 20mm
	Kheder and Al Jadiri [8]	2010	Mixture design based on the ACI 211.1 methodology for proportioning conventional concrete and the EFNARC method for proportioning concrete with objective function of ratio of cost to CCS
Statistical mixture design method	Simon [41]	2003	Experimental design based on classical mixture approach and mathematically independent variable (MIV) approach
	Kharazi et al. [42]	2013	Obtain statistical relationship based on IV-optimal design with five factors namely; C, W, CA, FA, high-range water reducing admixture (HRWRA)
	Bouziani [43]	2013	Fresh and hardened properties of concrete evaluated based on the effects of sand proportions (river sand, crushed sand and dune sand) in binary and ternary systems.

3. C + F_{ash} or Bfs + Silica Fume (**SF**) in 75:15:10 ratio by volume

- estimate aggregate content with **FA** and Coarse Aggregate (**CA**) for the first grade set at a 2:3 volumetric ratio. Depending on the moisture contents in the aggregate and superplasticizer (if used), adequate moisture correction is made.

Maintaining the grades of strength in the first step requires adjusting the volume of water initially added, water content from the aggregate and superplasticizer. This method is based on experiences, very simple to follow and provides a rational approach of HPC mixture proportioning without extensive laboratory trials. Figure 2.1.

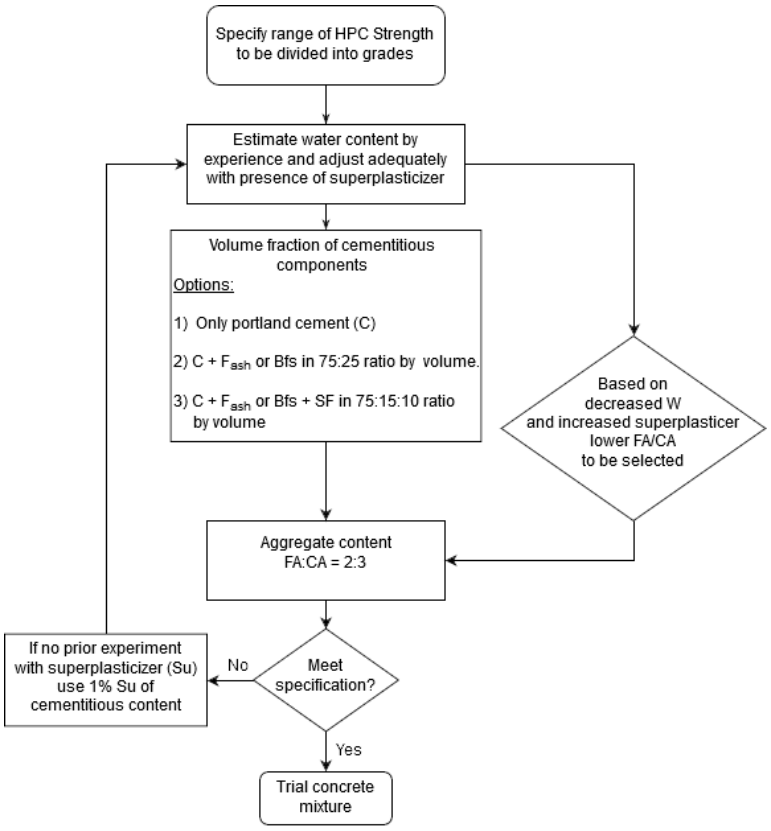


Figure 2.1. Mixture design procedure based on Mehta and Aïtcin's approach [2].

De Larrard [36] in 1990 also proposed a mixture design for HPC achieving an average CCS of 101 MPa. His method provides a convenient way to design HPC mixture using two semi-empirical mix-design tools as well as adopting the granular composition in accordance to Laboratoire Central des Ponts et Chaussées (LCPC) mix-proportioning method for normal strength concrete. Using this approach, De Larrard [36] predicted the strength of concrete using Feret's formula from a limited number of mix-design parameters. The workability in this approach is assumed to be closely related to the viscosity of the mix. The viscosity is computed using the Farris model, a rheological model dealing with polydispersed suspensions. The empirical formula and theoretical model in this approach allow the determination of concrete mixture composition producing a given strength and workability with a minimal number of trial concrete batches.

Later in 1995 and 1997, Okamura et al. [37, 44] proposed a mixture design method for self-compacting HPC. The design procedure included fixing percentage ratios of concrete constituents. Here **CA** content was set at 50% of the solid volume, **FA** content set at 40% of the solid volume, water-to-powder ratio assumed between a range of 0.9 to 1.0 by volume while **Su** dosage and the final water-to-powder ratio were adjusted ensuring self-compatibility. Although this approach promises a simple procedure in designing self-compacting HPC, the requirement for a higher dosage of **Su** to obtain high workability and moderate viscosity results in an increased cost of concrete production. Improving on Okamura's approach, Edamatsa et al. [3, 45] in 1999 and 2003 modified the method by fixing **FA** ratio, volumetric water-to-powder ratio and **Su** dosage. As opposed to Okamura's approach, the method by Edamatsa et al. [3] can be applicable to powder material and various quality of aggregates to enhance HPC. However, further improvement can be done on characterizing the concrete constituent properties, which includes the compatibility between superplasticizers and powder materials. Figure 2.2 shows Edamatsa's mixture design procedure.

Close particle packing method

This method of concrete mixture design works by proportioning various sizes and amount of solid constituents (aggregates) and then applying pastes to fill the "least

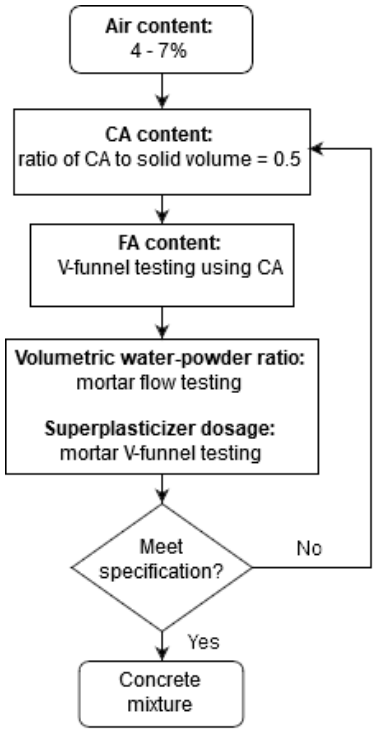


Figure 2.2. Mixture design procedure based on Edamatsa's approach [3].

void" between the aggregates [46, 47]. Research studies have revealed that the packing density of solid constituents in a concrete mixture has significant effects on concrete performance [39, 48, 49]. A higher packing density of aggregate and cementitious materials, imply a smaller volume of voids to be filled with paste and water respectively.

Hwang et al. [50] proposed the use of Densified Mixture Design Algorithm (DMDA) for the design of **HPC**. In another study, Chang [38] used the DMDA approach to optimize the mixture design of **HPC**. This study reveals that the utilization of fly ash and blast furnace slag contribute to the development of long-term strength of **HPC**. The logic of DMDA is a durability design concept used to achieve minimum water and cement content and maximize density by applying fly ash to fill the voids between aggregates and cement paste. With this mechanism, the cement paste acts as the binder for all solid particles while filling the remaining

void between them. However, a study by Hwang and Tsai [4] revealed that the excess amount of paste might lead to abnormal slump loss, bleeding and chemical contraction due to hydration [4]. The design procedure for this method of mixture design is shown in Figure 2.3.

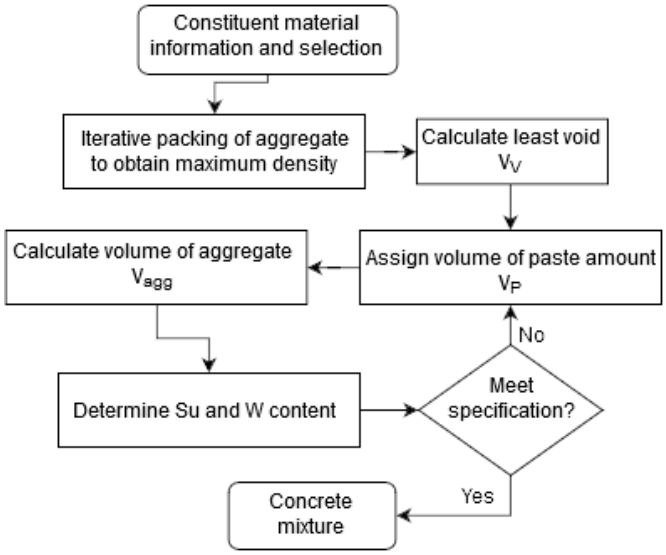


Figure 2.3. Mixture design procedure based on DMDA approach [4].

The amount of **Su** using this approach is determined by its quality and the Water (**W**) content. Whereas under a fixed amount of **W** and **W/C**, the **Su** dosage can be estimated based on past experience [4].

In order to produce economical concrete relative to the cost of production, Su and Miao [5] proposed a mixture design method using a packing factor (PF). The working principle of PF is to have the voids between a loosely piled aggregate framework filled with the optimum amount of binding paste. The authors mentioned that the first step in using this method is to calculate the fine and coarse aggregate content. Thus calculations for the fine and coarse aggregates are done by:

$$FA = \text{PF}\, C_{Lfa} \frac{V_{fa}}{V_{Ta}} \tag{2.1}$$

$$CA = \mathrm{PF}\, C_{Lca} \left(\frac{1 - V_{fa}}{V_{Ta}} \right) \tag{2.2}$$

where *FA* is the fine aggregate content, *CA* is the coarse aggregate content, C_{Lfa} is the unit weight of loosely piled saturated surface dry fine aggregates, C_{Lca} is the unit weight of loosely piled saturated surface dry coarse aggregates, V_{fa}/V_{Ta} is the volumetric ratio of fine aggregate to the total aggregate content, **PF** is the packing factor which represents the ratio of mass of tightly packed aggregates in the mixture to that of loosely packed state. Figure 2.4 depicts the procedure of the mixture design [5].

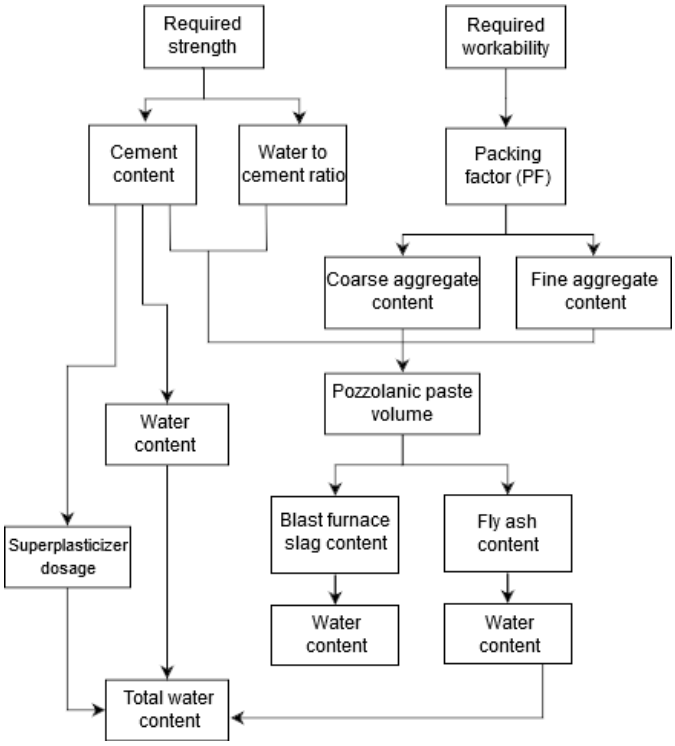


Figure 2.4. Mixture design procedure based on Su and Miao's approach [5].

The packing factor determines the aggregate content and also influences the workability and compressive strength. By adopting this concrete mixture design, less amount of binding materials would be needed. However, the procedure of determining the optimum fine aggregate to coarse aggregate ratio via packing factor

is not well explained and is assumed based on experience.

De Larrard and Sedran [51] proposed a method based on the solid suspension model (SSM), which predict the packing density of particle mixtures. Here the fine aggregate content is used in regulating the maximum paste thickness, which in turn leads to an optimal compressive strength of the mix. The mixture design procedure is based on the fluid consistency, concrete constituents and moderate thermal curing. In using this procedure, to maximize **CCS** the authors concluded that it is first desirable to only use fine aggregate as the aggregate content of the mix, after which a moderate theoretical viscosity (about 10^4 of the water) is chosen while maintaining the matrix final porosity using the following formula:

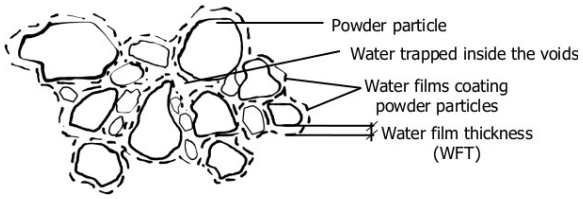
$$\pi_M = \frac{(0.23v_w + v_a)}{1-g} \tag{2.3}$$

where π_M is the matrix final porosity, v_a, v_w and g are respectively the partial volume of air, water and aggregate. Although this method provides promising results in the mixture design of **HPC**, the influence of the maximum paste thickness on compressive strength is not explained. The paste thickness relative to the compressive strength is assumed empirically to carry out the mixture design.

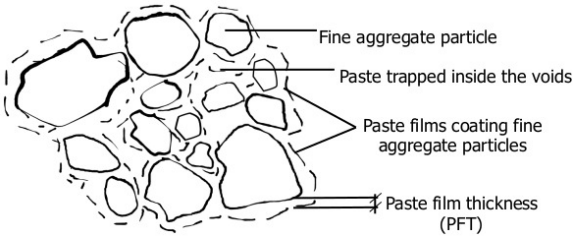
De Larrard and Sedran [39] also proposed another method based on the compressive packing model (CPM), which is the third generation of packing models developed at Laboratoire Central des Ponts et Chaussées (LCPC). This method focuses on the perspective of packing density, which is based on virtual packing density and a compaction index in predicting the granular structure of concrete. Though the models of this method may be incorporated into software for easy simulation, it may become difficult for practitioners to use without purchasing the software.

In a recent study Ng et al. [40] proposed a three-tier mixture design method where the packing and film theories of concrete materials have been studied and developed for mixture design of **HPC** in terms of the water film thickness (WFT), paste film thickness (PFT) and mortar film thickness (MFT) in the concrete. The authors showed that the packing density of solid particles in concrete has significant effect on the WFT, PFT and water film coating the cementitious material, fine aggregate particles and coarse aggregate particles respectively as shown in

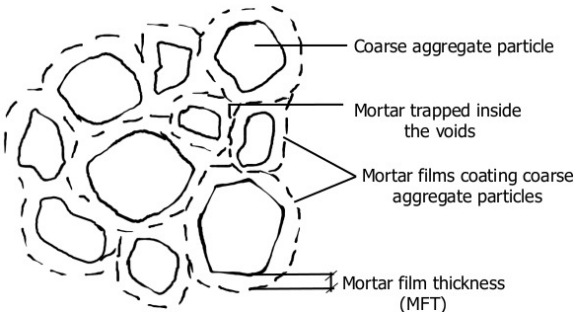
Figure 2.5.



(a) Water film thickness coating



(b) Paste film thickness coating



(c) Mortar film thickness coating

Figure 2.5. Film thickness coatings of the three-tier mixtures.

In this method, the concept is to design the concrete mixture systematically and sequentially in three tiers via a wet packing test, firstly the paste, mortar and concrete by adopting suitable values of WFT, PFT and MFT. This method is focused on the optimization of concrete mixture constituents to achieve better performance. The authors concluded that the MFT, PFT and WFT are governing factors of the rheological performance of concrete. Although they recommended that for a mixture design of **HPC**, the values of WFT should be within 0.14 μm to 0.40 μm, the PFT should be within 20 μm to 60 μm, while the MFT should be selected

according to the flow rate and filling height requirements.

Compressive strength method

This method focuses on determining the constituents of concrete for mixture designs based on required compressive strengths. Dinakar and Manu [6] proposed a methodology for obtaining high strength concrete, which contains metakaolin based on efficiency approach, as shown in Figure 2.6.

This method employed calculations divided into a series of steps to produce a high strength self-compacting concrete. For a required compressive strength at 28 days, the total cementitious content was fixed in the first step, while the percentage of metakaolin was fixed and the efficiency of metakaolin calculated. The efficiency factor (k) at 28 days for the percentage of metakaolin can be evaluated by a formula proposed by Babu and Dinakar [52] in the second step. The water content is calculated in the third step while the aggregate contents are determined using aggregate grading recommended by DIN 1045 [53] standards.

The concrete designed with this methodology and the established efficiency values for metakaolin method achieved strengths of 80, 100 and 120 MPa with metakaolin percentages of 7.5%, 15% and 22.5%. The compressive strength of concrete produced using this proposed method exceeded very high strength, making it a sufficient method of mixture design for high strength concrete. Additionally, the proposed method considered gradation of all aggregate contents. However, to achieve an optimal mixture proportion, a continuous adjustment of all concrete constituents are required.

Habibi and Ghomashi [7] recently developed an economical mixture design relative to compressive strength results and concrete slump. The method, as shown in Figure 2.7 was used to optimize mixture designs for different levels of compressive strength with optimum minimum cost (cost-strength ratio) as the objective function of the design. Though the guidelines which the authors used in this design is based on ENFARC [54].

However, to produce an economical mixture design relative to cost using the compressive strength method, every step of the procedure will require adjustments of all material constituents to achieve an optimal mixture proportion of lost cost.

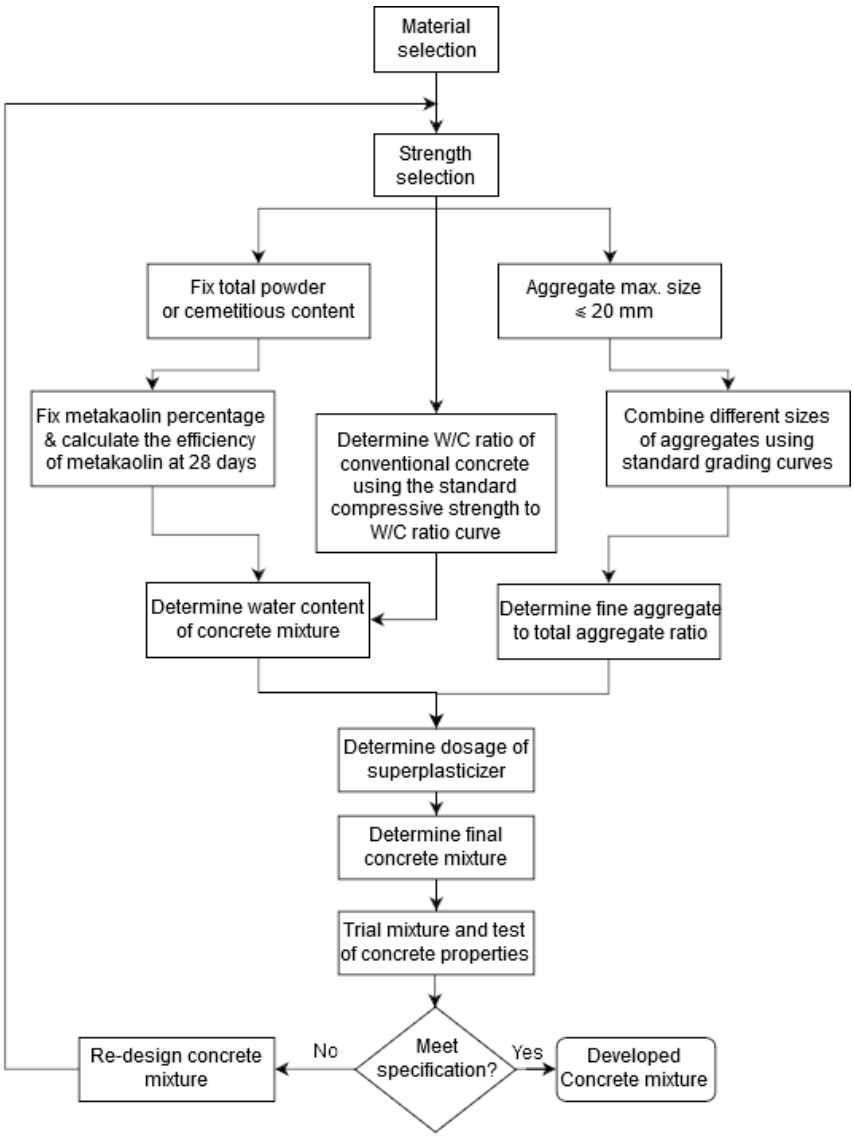


Figure 2.6. Mixture design procedure based on Dinakar and Manu's approach [6].

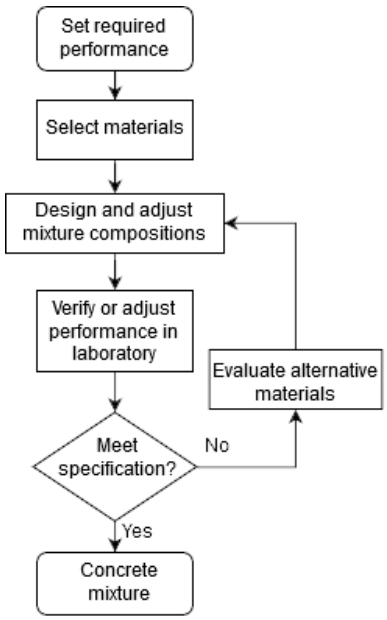


Figure 2.7. Mixture design procedure based on Habibi and Ghomashi's approach [7].

With the need to produce mixture designs of high strength, Kheder and Al Jadiri [8] formed a new method by combining the standard ACI 211.1 [55] method for proportioning conventional concrete and ENFARC [54] method for proportioning self-compacting concrete. Their approach emphasizes the dependency of material constituents to **CCS**. For instance, **W** content depended on the maximum aggregate size (MAS) and the concrete strength, the **CA** content depended on MAS and fineness modulus of the **FA** content while the **W/C** and volumetric ratio of water to powder were determined by the **CCS**. The procedure of this method is shown in Figure 2.8.

Statistical mixture design method

This method focuses on using statistical factorial models in identifying objective functions by using the knowledge of the effects of critical parameters such as the content of cementitious materials, volumetric **W/C** ratio, contents of chemical admixtures etc. during concrete mixtures.

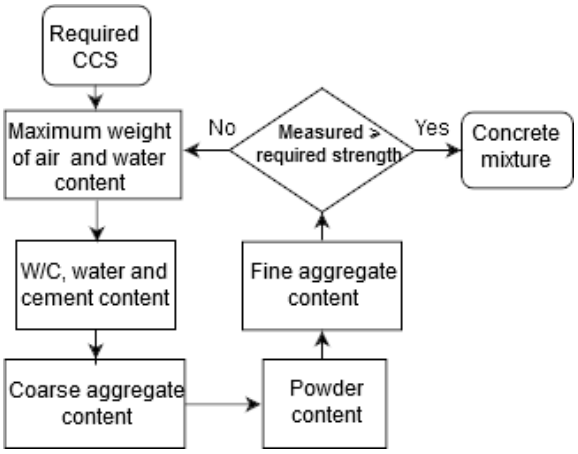


Figure 2.8. Mixture design procedure based on Kheder and Al Jadiri's approach [8].

Simon [41] conducted a technical report on using statistical factorial models for concrete mixture design comprising of six key parameters, namely; water (**W**), Cement (**C**), silica fume (**SF**), coarse aggregate (**CA**), fine aggregate (**FA**) and high-range water reducing admixture (**HRWRA**). Statistical factorial design models were used to find optimum proportions for a concrete mixture to meet the conditions of slump of 50 mm to 100 mm, 1-day compressive strength of 22.06 MPa, 28-day compressive strength of 51.02 MPa, 42-day charge passed in American Society for Testing and Materials (**ASTM**) C1202 "rapid chloride" test (RCT) less than 700 coulombs and minimum cost. This portrayed the efficacy of statistical design method in meeting several desired conditions in concrete mixture designs. Although using a statistical approach to mixture optimization requires a significant investment in trial batches and testing (runs) required to fit the model for each response and provide control and replicate runs for estimating repeated trial runs. Simon [41] reported that if the response is repeated adequately by the models, the number of trial batches can possibly be reduced by about 50%.

Kharazi et al. [42] used a statistical mixture approach based on IV-optimal design [56] for concrete mixes with five mixture components. The IV-optimal design comprises of a systematic procedure in a constrained design space that selects design points from a set of candidate points which are highly dependent on the

order of the fitted model [56]. The design space represents the region in which trials of the experiment are feasible. This design approach provided a cost-effective means for concrete performance optimization after investigating test results from 20 experimental runs to develop adequate models. In the study, the key performance criteria were slump and CCS at 3, 7, 28, 56 and 91 days. Additionally, an optimization analysis was performed to obtain the optimum material composition to meet all desired performance criteria simultaneously. An advantage of using the models established by statistical mixture approach is that they can predict the desired concrete properties and also assess the effect of mixture design parameters/components effectively unlike the traditional concrete mixture design method.

While Ellis et al. [57] used a simplex mixture design in modeling the setting time and compressive strength in sodium carbonate activated blast furnace slag mortars, Bouziani [43] had used a statistical method of classical mixture approach called the simplex-lattice mixture design. The study consists of using the simplex-lattice mixture design comprising of three factors and five levels to evaluate the effects of three types of sand (river sand (RS), crushed sand (CS) and dune sand (DS)) in binary and ternary combination, on properties of fresh and hardened concrete. The simplex-lattice design is a space-filling design that creates a triangular grid of combinations as portrayed in Figure 2.9. In the figure, the factors RS, CS and DS are represented as x_1, x_2, x_3 respectively, while the number of combinations (C) is calculated as follows:

$$C = \frac{(q+m-1)!}{m!\,(q-1)} \tag{2.4}$$

where q represents the number of factors and m represents the number of levels. With the equation given, three factors and five levels equates to 21 number of combinations to be treated.

Additionally, by using this approach, a mathematical model that describes the individual and combined effects of the three types of sand (RS, CS and DS) can be established. A second-degree order model with three non-independent variables

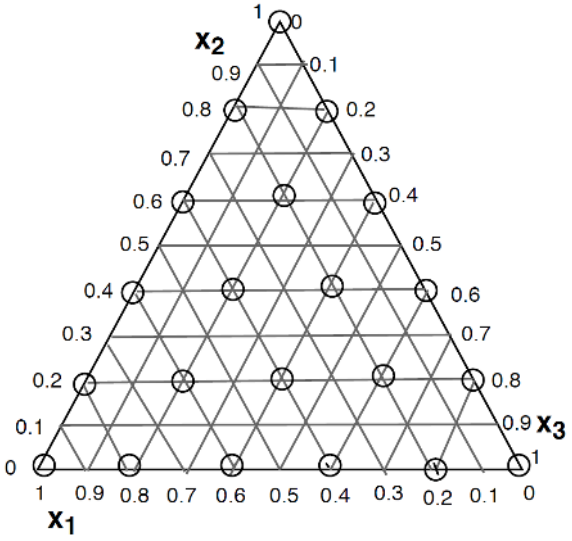


Figure 2.9. Illustration of simplex-lattice design with three factors and five levels.

(proportions of RS, CS and DS) and five levels was used and expressed as:

$$Y = b_1 * \mathrm{RS} + b_2 * \mathrm{CS} + b_3 * \mathrm{DS} + b_4 * (\mathrm{RS} \cdot \mathrm{CS}) + b_5 * (\mathrm{RS} \cdot \mathrm{DS}) + b_6 * (\mathrm{CS} \cdot \mathrm{DS}) \tag{2.5}$$

where the model's coefficients (b_i) represent the contribution of the associate variables on the response Y. These model coefficients are determined by a standard least-square fitting usually executed using statistical software.

Although the classical mixture design approach employs the use of statistical models, all the factors are constituent proportions of the mixture, which must sum up to be 1. The makes the last component proportion dictated by the sum of all other constituent proportions. In other words, in this mixture approach, the factors are considered not independent. Though an advantage of this approach is that it allows the experimental region of interest to be defined more naturally. However, it becomes complicated in the analysis of the results when using this classical mixture design approach of the statistical method. The statistical mixture design method provides accuracy in design and also avoid extensive repeated experiments, it requires some statistical expertise which may be difficult for individuals with no

statistics background knowledge.

Summary

From the literature review, it is apparent that the statistical mixture design provides more accuracy as well as a means for mixture design modeling and optimization. Although, one of the benefits of using the statistical mixture experiment approach is placing restrictions in the form of mathematical equations on certain regions in the design space which are undesirable. It becomes complicated to analyze the results with the classical mixture design approach using a certain number of components.

The classical approach of observing one constituent factor at a time and studying the effect of the factor on concrete property is complicated, especially in a multivariate system. With such constraints, adequate procedures and appropriate models to navigate such constraints in mixture designs for an optimal compressive strength of concrete are required. Moreover, with the advent of newer and expensive concrete constituents, it would be economical to reduce the number of the trial batches in concrete production. In an attempt to address these challenges, and fill in the knowledge gap, the Box-Behnken Design (**BBD**) of experiment on Response Surface Methodology (**RSM**) is proposed.

2.2 Prediction of Concrete Compressive Strength for High Performance

Introduction

Apart from the optimization of concrete mix proportion to achieve desired high compressive strength, a lot of questions arise in the construction industry about the efficacy of predicting **CCS** from the knowledge of mixture proportions using statistical analysis and machine learning techniques [58]. Not only does this question focus on normal concrete, but it also possesses major thoughts in high strength concrete, which is made by mixing some other cementitious materials [59]. Although, due to the need in cost reduction, concrete production has taken a lot of direction with constituents consisting of recycled materials [60]. However, the subject of adequately predicting concrete strength with the respective constituents is

paramount. Hence, it is critical to perform researches to understand the complexity in constituent interactions [61], as well as finding adequate and more accurate nonlinear function to analyze the compressive strength of concrete [62].

Existing Techniques for Concrete Strength Prediction

Proceedings from scientific studies to understand concrete strength interaction with its constituents have led to the use of statistical and computational models, including machine learning techniques for the prediction of **CCS** [63–65]. Although these techniques proved to be sufficient to predict CCS, the unique comprising constituents of concrete and their rational distribution during concrete mixture predominantly affects the predictions of CCS.

A widely used technique is the predictive techniques by linear or Non-linear Regression (**NLR**) methods [66]. In 2002, Bhanja and Sengupta [66] investigated CCS using 5% to 30% partial replacement of silica fume as a cementitious material to estimate the strength of concrete at 28 days. The validity of their nonlinear model portrays the importance of non-dimensional variables of concrete in CCS prediction. However, their study emphasized the effect of only one cementitious material in CCS prediction. Later in 2009, Zain and Abd [67] emphasized the importance of linear regression in predicting the 28 days CCS with more than one cementitious material as well as having water to cement ratio being a non-dimensional explanatory variable. In their study, they employed the use of multilinear regression, which provided a quick and accurate estimation of CCS on site, which is valuable to the construction industry.

Although the use of **NLR** has wider implementation for regression problems, Classification and Regression Trees (**CART**) have been fairly used in predicting relationships between variables due to its simplicity of graphical representation [68]. In **CART** the same predictor may be used in different levels and several times in both classification and regression problem. However, in regression problems such as concrete property predictions, due to its superiority in logic rules, Regression Trees (RT) are widely considered than other modeling techniques. Chou and Pham [69] in their study on ensemble approach in predicting concrete strength achieved high predictive accuracy by combining regression tree model with models from

other predictive technique.

Recently, the need for higher predictive accuracy for **CCS** has been on the rise and given the non-linear characteristics of concrete, various machine learning techniques especially Artificial Neural Network Artificial Neural Networks (**ANN**) have been investigated [59, 70–72]. Although ANN consists of several architectures which provide reasonable predictive results, it is quite known to result in a black-box due to the difficulty it possesses in understanding procedures to which the predictive models are obtained [73]. Nevertheless, the multi-layer perceptron Multilayer Perceptron (**MLP**) is a simple and most popular representation of the ANN architecture used in predicting properties of concrete [26]. In a study of classifying concrete based on its strength while obtaining its predictive values, Khashman and Akpinar [65] employed artificial neural network for **CCS**. Their study infers that non-destructive system portrays high efficiency in classifying and predicting **CCS** for use in real life application. Although using the same neural network technique in a study for predicting **CCS**, Deshpande et al. [74] still affirm the capability of the ANN. However, they observed that tree-based model techniques provided excellent performance in the analysis of varying complexity and accuracy. This tree-based machine learning techniques provide useful information during analysis for regression problem in predicting **CCS**.

With the widespread of machine learning techniques for concrete performance prediction, Gupta [75] was among the first to predict **CCS** using Support Vector Machine (**SVM**), in a study with a value of $R^2 = 0.99$. The study illustrated the capability of **SVM** as an efficient technique in modeling **CCS** using small dataset. Yan and Shi [76] also performed a study which demonstrated the efficacy of **SVM** in predicting concrete properties. Their study concluded that **SVM** demonstrated good performance than other compared model. However, like most research studies, their experimental data used in the study were few with small compressive strength boundaries.

The working principles of these commonly used predictive techniques in the literature and industry practices for estimating **CCS** are as follows:

Non-linear regression

The Multiple non-linear regression (**NLR**) model determines the relationship between two or more explanatory variables and response variables by fitting a linear equation to the sample data. In prediction modeling, the use of a linear regression model is usually the first attempt due to its wide acceptance and simplicity in application [77, 78]. In addressing computational modeling problem, NLR fits a hyperplane to an n-dimensional space where n represents the number of explanatory variables. A system with n explanatory variables x's and one response variable, y, the general least square problem is to determine the unknown parameters β_i of the linear model. The general representation is shown in Equation 2.6.

$$y = \beta_0 + \sum_{i=1}^{N} \beta_i x_i + \varepsilon \tag{2.6}$$

where y represents the concrete compressive strength **CCS**, x represents the concrete explanatory variables,$(i = 1,2,...N)$, β_i, $(i = 1,2,...N)$ represents regression coefficients and ε is the error term. The use of **NLR** model is vastly applied in civil engineering problems and has been used in this study for comparison.

Multi-layer perceptron

The multi-layer perceptron (**MLP**) is a simple and most popular representation of the artificial neural network (ANN) architecture consisting of a set of layers (input layer, hidden layer and output layer), where a given number of neurons in the hidden layers can decipher complexities and non-linearity in concrete predictability [26]. An artificial neural network (ANN) is considered any computational technique that simulates after the structural function of the biological neural network [73]. The ANN has been found to be extensively adequate in modeling the relationship of concrete relative to its mechanical properties [16, 79].

Though there are several learning algorithms used in training a **MLP** neural network, the most widely and effective is the back-propagation algorithm. The algorithm makes adjustments to the weights and bias value between neurons during the training process to simulate input-output relation. Here the input values to a neuron are derived by the product of the output of the connected neuron and

the synaptic strength of the connection between them. Activation of neurons in a hidden layer can be expressed as follows:

$$net_j = \sum_{i=1}^{n} w_{ij} x_i, \tag{2.7}$$

where net_j is the weighted sum of the jth neuron for the input received from the preceding layer with n neuron, w_{ij} is the weight of connection between the jth neuron and the ith neuron in the preceding layer, x_i is the output of the ith neuron in the preceding layer. The output of the jth neuron out_j is calculated by Equation 2.8 with a sigmoid transfer function as follows:

$$out_j = f(net_j) = \frac{1}{1 + e^{-knet_j}} \tag{2.8}$$

where k controls the function gradient. The sigmoid nonlinearity is activated in all layers of the nerual network except in the input layer [73]

To train and update weights w_{ij} in each cycle h, Equation 2.9 is employed:

$$w_{ij}(h) = w_{ij}(h-1) + \Delta w_{ij}(h) \tag{2.9}$$

where $\Delta w_{ij}(h)$ is expressed in Equation 2.10 as:

$$\Delta w_{ij}(h) = \eta \delta_{pi} x_{pi} + \alpha \Delta w_{ij}(h-1), \tag{2.10}$$

where η is the learning rate parameter, δ_{pi} is the propagated error, x_{pi} is the output of the ith neuron for record p, α is the momentum parameter, and $\Delta w_{ij}(h-1)$ is the change in w_{ij} in the previous cycle. A typical structure of MLP with three layers used in the study is shown in Figure 2.10.

Classification and regression tree

Regression tree Regression Tree (**RT**) is a part of Classification and Regression Trees **CART**, which is a simple machine learning algorithm that provides a foundation for constructing predictive tree-based models for categorical or numerical data analysis [68]. Breiman et al. [68] introduced this binary tree which can be op-

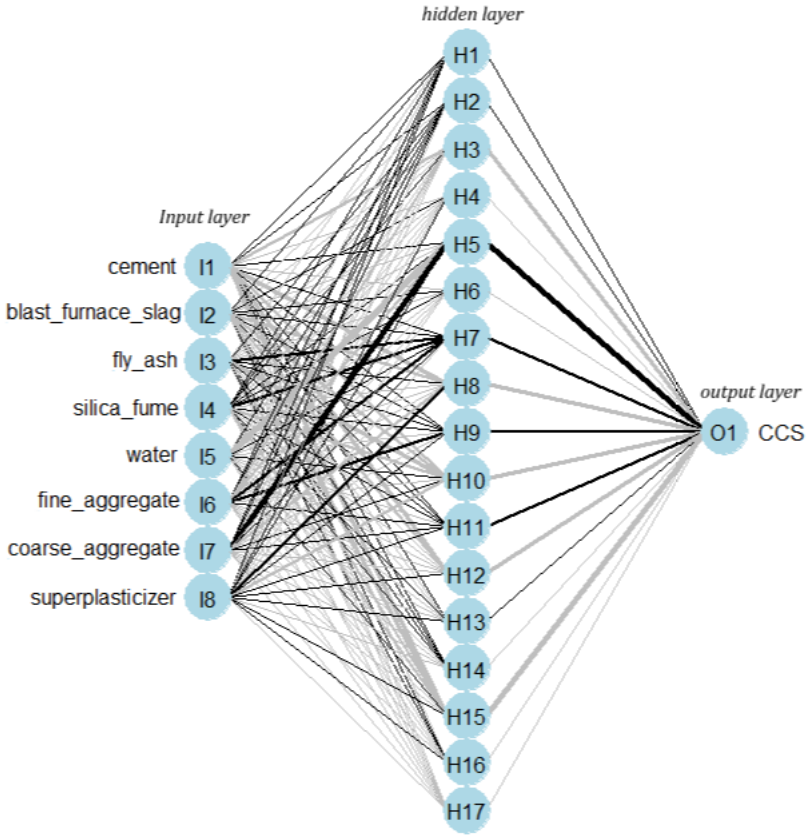


Figure 2.10. Graphical representation of ANN.

timized by a learning process which prunes saturated trees and selects among the obtained sequence of nested trees. The procedure of dichotomizing binary trees in this model employs some measures: Gini, which is applied to symbolic target fields whereas a least-squared deviation method is used in selecting continuous targets automatically. Given a node d in a CART, the Gini index $g(d)$ is defined as follows:

$$g(d) = g(d) = \sum_{i \neq j} p(i \mid d) p(j \mid d), \tag{2.11}$$

where i and j are target response categories

$$p(i \mid d) = \frac{p(i,d)}{p(d)}; \;\; p(id) = \frac{\pi(i) n_i(d)}{n_i}; \;\; p(d) = \sum_i p(i,d) \tag{2.12}$$

where $\pi(i)$ is the prior probability value for the ith category, $n_i(d)$ represents the number of records in the ith category of node d, while n_i represents the number of records of the ith category in the root node. To determine the improvement after a split in the tree using the Gini index, only records at the root node and node d are used to compute n_i and $n_i(d)$ respectively.It is due to the superiority in these logic rules, CART are widely considered than other modeling techniques.

Support vector machine

Vapnik [80] introduced support vector machine (**SVM**) which have been used in several civil engineering application [81, 82] with the basic idea to solve linear regression problems by adopting a nonlinear transformation (mapping) of the input data into a higher m dimensional feature space. This transformation is usually done using a kernel function which may consist any of the following; linear, polynomial, sigmoid and Gaussian Radial Basis Function (RBF) [83]. The linear model in the feature space, $f(x, \omega)$, is expressed in mathematical notation as follows:

$$f(\mathbf{x}, \omega) = \sum_{m}^{j=1} \omega_j g_j(\mathbf{x}) + b \tag{2.13}$$

where g_j(x), j = 1,...,m is a set of nonlinear transformations from the input space, b is a bias term and ω represents the weighted vector estimated by minimizing the regularized risk function of the empirical risk. Estimating quality is measured by the loss function L_ε proposed by Vapnik [83]:

$$L_\varepsilon[y, f(x,\omega)] = \begin{cases} 0 & \text{if } \mid y - f(x,\omega) \mid \leq \varepsilon \\ \mid y - f(x,\omega) \mid & \text{otherwise} \end{cases} \tag{2.14}$$

SVM regression computes linear regression in the high dimensional feature space by using its ε-insensitive loss function and reduces model complexity by minimizing $\|\omega\|^2$. An illustration of the procedure is depicted by including nonnegative

slack variables ξ_i and ξ_i^*, where i = 1, ..., N to identify training samples that deviate from the ε-insensitive zone. The SVM regression can thus be formulated as a minimization of the function as follows:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \tag{2.15}$$

$$\text{subject to} \begin{cases} y_i - f(x_i, \omega) \leq \varepsilon + \xi_i \\ f(x_i, \omega) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}, i = 1, ..., N$$

Transforming the optimization problem into the dual problem, can be solved by

$$f(x) = \sum_{i=1}^{n_{SV}} (\alpha_i - \alpha_i^*) K(x, x_i) \quad \text{subject to} \tag{2.16}$$

$$0 \leq \alpha_i^* \leq C, \ 0 \leq \alpha_i \leq C$$

where n_{SV} = number of support vectors (SVs) and the kernel function i

$$K(x, x_i) = \sum_{i=1}^{m} g_i(x) g_i(x_i) \tag{2.17}$$

The selection of kernel function (i.e., linear, radial basis, polynomial, or sigmoid function) parameters usually depend on the kernel type and the implementing software used. Though selecting the kernel types and function parameters should reflect the distribution of the training data set.

Summary

Ideally, the literature on predicting concrete properties contains studies with applied machine learning algorithms attempting to identify which algorithm has the best predictive performance. However, the claims in these studies mostly apply to few or single-sourced experimental data. Although making definitive conclusions about which algorithm or technique is the ”best” remains arduous - following a principle of No Free Lunch (NFL) Theorem [84], which summarizes that a model with good performance may conduct poorly on some types of problems. Thus,

the best technique or algorithm is often but not always dependent on the data or problem from which it learns.

In an attempt to explore the estimation of **CCS** performance of an extensive experimental boundary, a new technique; Boosting Smooth Transition Regression Trees (**BooST**) is proposed.

Chapter 3

Computational Methods for Concrete Compressive Strength Analyses

In this chapter, the proposed techniques and computational analysis for the mixture design of concrete for high performance and strength as well as the prediction of concrete compressive strength performance are proposed. The proposed response surface design (Box-Behnken design) of experiment is presented in Section 3.1, where a brief overview of the pertinent background information is given. The section also contains the diagnostic evaluation procedure for the proposed technique. In Section 3.2, the proposed technique for predicting concrete strength performance using Boosting Smooth Transition Regression Trees (**BooST**) is presented. The overall structure, as well as the working principle of the machine learning technique, is also presented. Additionally, a comparison of this technique with other contemporary machine learning techniques is evaluated and presented, as a means of demonstrating the effectiveness of the **BooST** when applied in concrete performance prediction.

3.1 Response Surface Design of Experiment Methodology for Optimal Concrete Compressive Strength

In recent studies, the use of several response surface Design of Experiment (**DOE**) do not only provide means for initial assessments in experimental designs but also serves with it, a range of procedures for statistical analysis and optimization for these designs [56, 85, 86]. To investigate the effect of fly ash replacement on early and late compressive strength, a DOE simplex-centroid design was adopted and investigated by Yeh [86] in 2006. The potential of using this design method was used to determine the effect of fly ash replacement from 0% to 50 % in concrete. Yeh concluded that in using a simplex-centroid mixture DOE, a much smaller number of experiments need to be performed to obtain reasonable data. Such a study to obtain meaningful data with a small number of experiments further motivated some research studies. In 2014, Ahmad and Alghamdi [87] proposed a statistical approach to obtain optimum proportioning of concrete mixtures using varying key level factors affecting **CCS**. Their study further validated the efficacy of producing an optimal concrete strength using a DOE method. In a recent study, to save cost and promote the use of environmentally friendly materials in concrete production, Lejano and Gagan [88] used the central composite design to establish the design of experiment for demonstrating the optimization of **CCS** with pig-hair fibres and green mussel shells as a partial cement substitute. Using this Response Surface Methodology **RSM**, their study resulted in reducing the number of experimental runs and saving experimental cost.

Although a major focus on concrete experimental design is to provide optimal strength with constituents, there is additional need of exploring other experimental design methods that would provide favorable results for any experimental boundary where some experimental results are missing or difficult to obtain. Hence, an RSM using Box-Behnken Design **BBD** is proposed in this section to achieve better optimal strength having corresponding mixture compositions with a minimal number of experimental runs (experiments).

Response Surface Methodology (**RSM**) is a collection of mathematical and statistical techniques useful for the analyses of several experimental input variables

in a process to obtain an optimal response of interest [56]. The sequential procedure of RSM to obtain an optimal response in a region of operability for any process can be likened to involve the ascent or descent of the current operating conditions through paths of improvement into a region of optimum. That is to say; it is a technique that escalates the sequential movement of a process through a direction to attain maximum response. In any experiment having several input variables, the main effects of such input variables and their interactions to yield response can be effectively investigated via RSM approach. Nevertheless, with requirements of several runs in experiments, the main benefit and advantage of RSM over other statistical techniques in evaluation is the relatively reduced number of experimental runs necessary to evaluate a response from a given set of variable interactions in the process [89]. Three general steps comprise **RSM**: experimental design, modeling and optimization. The following processing features sets RSM to be a convenient and sophisticated technique in the concrete domain:

3.1.1 Material Composition and Distribution of Dataset

For the illustration of RSM on HPC, using this dataset, a reasonable distribution of data points were provided throughout a region of interest. The levels and their respective factors (water to cementitious content ratio, fine aggregate to total aggregate ratio and cementitious content) were considered in this study according to their influence on HPC as reported by previous experiments and literature [87, 90–97]. Neville and Aïtcin [93] from their study concluded that the expected maximum ratio of water to the total cementitious material in the HPC mix would be 0.45. However, for this study, the Water-to-Cementitious Material Content ratio (W/C_{mt}) considered is between 0.30 to 0.45. W/C_{mt} was the leading factor used to obtain other factors' levels due to their availability in the dataset [98]. Levels of other factors; Fine Aggregate-to-Total Aggregate ratio (FA/T_A), (0.4 to 0.5) and Cementitious content (Q_{CmT}), (460 to 500)kg/m^3 were for this reason explored via RSM to obtain an optimal concrete compressive strength at 28 days. Since **BBD** of RSM does not require a large number of runs, the dataset was prepared and categorized to focus on performance-based results into these influencing factors with three levels.

3.1.2 Multivariate Design of Experiment for Concrete

It is essential to explore the effects of input variables to yield a true approximation of the response in any process or experimental designs. Some experimental techniques hold some parameters constant while investigating the effects of the independent variables. Such a procedure can be time-consuming when some input variables have multiple levels which contribute to how a response is influenced. However, with Box-Behnken Design [99] and Central Composite Design approach in **RSM**, all parameters and effects of input variables can be considered concurrently [56].

An essential step in RSM is to find an apt approximation for the true functional relationship between a response and its sets of independent variables. This comprises of exploring response optimal operating conditions through experimental methods to obtain sets of functional conditions to suit a high-order model. One cannot say in its entirety that RSM is not analogous to the usual regression problem. However, with cognizance of its process, the RSM technique is much broader and explicit in the analysis of an experiment. RSM technique includes importantly, the use of coded input variables, fitted model assessment, post-ad hoc analyses carried out depending on the fitted model and outcome and visualizing of the response surface [100]. All these procedures aid in defining the relationship between the input variables and response while finding an optimal response to an experiment.

As stated earlier, for the optimization of **CCS** in this study, the input factors considered are the W/C_{mt}, FA/T_A and Q_{CmT} with experimental levels of 0.3 to 0.45 , 0.4 to 0.5 and 460 to 500 kg/m^3 respectively. Box-Behnken Design **BBD** was employed for the experimental design, which essentially allows independent estimates of error to be obtained as well as develop a response surface model.

When a **BBD** is visualized as a cube, it does not contain experimental points at the vertexes of the cubic region created by the experimental variable limits [56, 101]. Mostly this is as a result of experimental points at those regions being either expensive to run or practically impossible to assess. However, for this case, it is due to the nature of the dataset that all the points at the vertexes are unavailable. Nevertheless, when a design contains points at the regions of both vertexes and between the vertexes of a cuboidal region, it is considered to be a hybrid design [56].

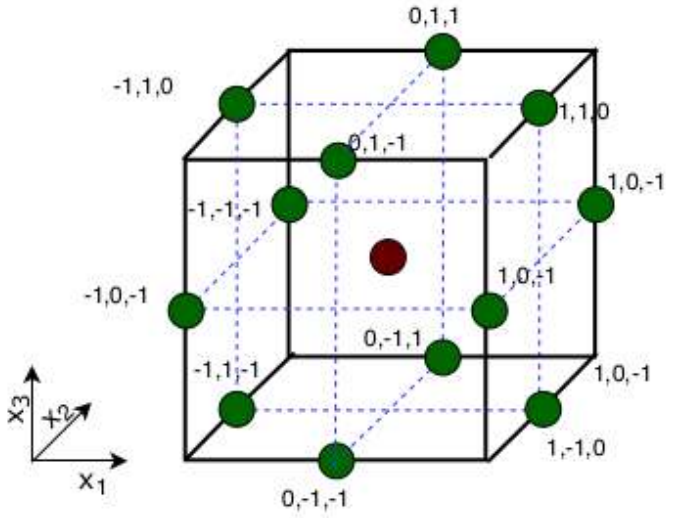


Figure 3.1. Schematic diagram of the BBD as a function of three independent variables x_1, x_2 and x_3 in a 2^3 factorial design.

As shown in Figure 3.1 the schematic diagram consists of levels of the factors at the midpoints of all edges represented by the green dots while the center point is in the center of the cube represented by the red dot. It shows the cuboidal representation of the BBD having the coded variables positioned as levels of the selected factors(variables). However, it is to be noted that unlike central composite design (CCD), BBD may not be used for an experimental design having only two factors. Borkowski [102] discussed the peculiarity of both CCD and BBD.

Coded variables of the original variables were used to analyze the results. Using the coded variables makes an easy clarification of analysis because the magnitude of the model coefficients becomes dimensionless, establishing a direct comparison. Another reason for using the coded variables is that it aids in the easy interpretation of results. Table 3.1 shows the coded variables and corresponding responses.

The coded variables levels can be obtained as

$$\frac{(actual\ level - centre\ value)}{half\ range\ value} \tag{3.1}$$

Table 3.1. Coded variable design and response for concrete of high performance at 28 days.

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	**CCS**
W/C_{mt}	Q_{CmT}	FA/T_A	**(MPa)**
-1	-1	0	44.52
-1	1	0	68.50
1	-1	0	33.72
1	1	0	35.31
0	-1	-1	41.05
0	1	-1	44.13
0	-1	1	42.66
-1	0	-1	58.99
1	0	-1	37.72
0	0	0	43.94
0	0	0	43.90
0	0	0	43.92
-1	-1	0	44.52
1	-1	1	38.63
-1	0	-1	53.58
0	0	-1	43.58
-1	0	-1	58.52
0	-1	-1	41.05

So, in order to obtain a coded level for $Q_{CmT} = \frac{(Q_{CmT}-480)}{20}$. x_1, x_2, x_3 represent the coded variables W/C_{mt}, Q_{CmT} and FA/T_A respectively. The terms -1, 0 and 1 represents *minimum*, *centre* and *maximum* level respectively for the actual values of the variables experimental boundary.

The design in Table 3.1 contains three center points which ensure an allowance for an independent estimate of the error to be obtained. Lawson [101] stated that by including center points, a design would have uniform precision and the variance of a predicted value will be the same at the origin in coded factor levels. Again, the center points investigate the presence of curvature in a system [56]. In other words, when the points are added to the center of a design, the test for curvature tests the hypotheses of the sum of quadratic effects in the system.

3.1.3 Computational Modeling of Concrete Design

An illustration for RSM to provide an interaction between input variables and yield response is vital. Therefore, it becomes essential to develop a relationship via a mathematical model fitted for the factorial design. RSM methodology comprises of a body of methods for exploring optimal operating conditions through different experiments using the given results to focus on sets of conditions to suit a higher-order model [100]. For instance, in using the central composite design (CCD) in RSM, if the assumption of linearity of factor effects in an experimental design best fit a model, then its first-order linear model could be adopted. However, due to the impact of the interaction between factor effects in a model, the first-order linear model may be insufficient in estimating the curvature of the response surface, thereby resulting to adopting a quadratic or second-order response surface model [56]. And this could be true for concrete since the concrete compressive strength is a highly nonlinear function of age and its other compositions [16]. For this study, the full quadratic (second-order) model is given as:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_i \mathrm{x}_i + \sum_{j=1}^{k} \beta_{ii} \mathrm{x}_i^2 + \sum_{i<j}^{k} \sum^{k} \beta_{ij} \mathrm{x}_i \mathrm{x}_j + \varepsilon \tag{3.2}$$

where y in this case given in Equation 3.2 is the predicted value for the concrete compressive strength with k number of factors or independent variables. β_0, β_i, β_{ii} and β_{ij} are the regression coefficients for intercept, linear, quadratic and interaction coefficients respectively while x_i and x_j are coded independent variables. ε represents the random experimental error effect. The matrix notation of this same model ($y = \mathrm{x}\beta + \varepsilon$) is expressed in Equation 3.3:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & . & . & x_{1k} \\ 1 & x_{21} & x_{22} & . & . & x_{2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_{n1} & x_{n2} & . & . & x_{nk} \end{bmatrix}}_{x} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}}_{\varepsilon} \tag{3.3}$$

The significance of the stated model represented in Equation 3.2 and 3.3 are evaluated by analyzing the model statistics given in section 3.1.4.

3.1.4 Model Diagnostic Evaluation

As earlier discussed, it is usually essential in RSM to find an appropriate approximation for the true functional relationship between sets of independent variables that yield a response via a model. However, the adequacy of such model obtained needs to be investigated to confirm that it fits to be a functional approximation of the true system. The investigation of this model also aids to verify if there is any violation of least squares regression assumptions; as an adequate fit of the model ensures that the results from the regression model are not misleading [56]. There are several exploratory statistics employed to investigate if a model provides an adequate fit for the approximation of a true system. These statistics are:

$$R^2 = \frac{SS_R}{(SS_R + SS_E)} = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{3.4}$$

Coefficient of Determination (R^2) which is calculated to identify the amount of reduction in the variability of expected response by using a given set of variables in a model. SS_R and SS_E are the sums of squares (regression) and residual sum of squares respectively. SS_T is the total sum of squares derived from the summation of Sum of Squares (regression) (SS_R) and Sum of Squares (error) or Residual Sum of Squares (SS_E).

Predicted Coefficient of Determination (${R^2}_{pred}$) which represents the predicted R^2, indicates the percentage of variation explained when the regression model predicts response for new observation data.

$${R^2}_{pred} = 1 - \frac{PRESS}{SS_T} \tag{3.5}$$

PRESS represents the prediction error sum of squares. To compute PRESS, the value obtained from the difference between actual and ith predicted values (prediction errors) are squared and summed up this summation results in the prediction error sum of squares.

$$R^2{}_{adj} = 1 - \frac{(n-1)}{(n-p)}(1-R^2) \tag{3.6}$$

The Adjusted Coefficient of Determination ($R^2{}_{adj}$) represents the percentage of variation explained by only the independent variables which affect the dependent variable. In Equation 3.6, n represents the number of observations while p represents the number of model parameters.

A way to assess the strength of fit is by finding the standard error of residual, which considers the magnitude of residual that shows how close the estimates may be from the actual values. Such statistic is represented as the mean square of Equation 3.20.

$$\text{MSE} = \frac{\sum_i e_i^2}{df} = \frac{SS_E}{df} \tag{3.7}$$

However, in analysis of variance, mean squares are used to determine when factors are significant. The mean square of error term (MSE) measures the amount of variance in the response variable that can be explained by a model. It is obtained by dividing the residual sum of squares by the degrees of freedom. In Equation 3.7, df represents the degree of freedom while e represents the residual.

3.1.5 Summary

In this section of the chapter, Box-Behnken Design (**BBD**) of **RSM** is proposed to provide favorable results for concrete of high performance and compressive strength having boundaries/constraints where some experimental results are missing or difficult to obtain. Hence, exploring the to achieve better optimal strength with corresponding mixture compositions of a minimal number of experimental runs (experiments). Computational results and discussions are presented in section 4.1 of Chapter 4.

3.2 Prediction of Concrete Strength Performance Using Boosting Smooth Transition Regression Trees (BooST)

In light of the challenges discussed in the literature which constitutes estimating concrete of high performance and compressive strength, a new machine learning technique is proposed. In this section of the thesis, the **BooST** model is developed for predicting 28 days **CCS** of high performance. Additional to exploring prevailing predictive techniques in comparison with the **BooST** for **CCS**, this study adds to the concrete domain body of knowledge by the following:

- To investigate and develop adequate models for predicting CCS of HPC using concrete constitutes and the several proportioning of these constituents as explanatory variables.
- To investigate the prediction of CCS using a limited number of explanatory variables and the influences these variables cause relating to concrete strength.
- To investigate the capability of the predictive techniques in evaluating complex structures of HPC compressive strength data using a wide range of experimental boundaries of concrete explanatory variables.

To illustrate the efficacy of this machine learning technique, its methodology is presented in this section. Additionally, the details of data description and preprocessing as well as the theoretical background of **BooST** are provided.

3.2.1 Boosting Smooth Transition Regression Tree (BooST)

Boosting smooth transition regression tree (BooST) is an ensemble of trees algorithm. The model from BooST is a non-parametric model that uses smooth transition regression trees (STR-Tree) [103] as base learners, making it possible to approximate the derivative of any underlying non-linear relationship between explanatory and response variables [1]. The STR-Tree is a type of regression tree with soft splits instead of a discrete (sharp) ones. This feature of the STR-Tree makes it simply differentiable with respect to the regressors, thereby ensuring analytical

computation of partial effects. Dai et al. [104] demonstrated how non-parametric model could recover partial effects without fitting regression function. The ideology behind the use of BooST is to substitute the traditional classification and regression trees (CART), which are not differentiable, by smooth logistic trees [1].

Actually, the BooST framework has a similar mechanism as the boosting framework [105], where the decision trees are grown sequentially using information from the previous trees. The principal algorithm behind the BooST model is the boosting algorithm [105], which is considered to be greedy with iteratively combining weak learners (models) to create much stronger ones as defined by Friedman [106].

Friedman's definition above conforms to a predictive estimation which system constitutes of a response variable y and a set of multiple explanatory variables $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ [106]. Employing a training set $\{y_i, \mathbf{x}_i\}_1^N$ of known (y, $\mathbf{x}$)-values, the goal is to find a function $F_0(\mathbf{x})$ that maps $\mathbf{x}$ to y such that over the distribution of y given $\mathbf{x}$ values, the expected value of some specified loss function $L(y,F(\mathbf{x}))$ that is convex on $F_0(\mathbf{x})$ is minimized as:

$$F_0(\mathbf{x}) = \arg\min_{F(x)} \sum_{i=1}^{N} L(y_i, F(\mathbf{x}_i)) \tag{3.8}$$

With the assumption that the algorithm is computed until $m-1$ iteration, the gradient of the cost function becomes:

$$z_m(\mathbf{x}_i) = \left[\frac{\partial L(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})} \tag{3.9}$$

where z_m is the gradient at the m-th iteration, assuming the algorithm is computed until $m-1$ iterations, L is the loss function and m is the number of sub-model ($m = 1, 2, ..., M$). The boosting algorithm used here is based on the $steepest-descent$ algorithm where at each iteration of the algorithm a step in the opposite direction of the loss function L is taken [106]. Therefore, using a two-step procedure a_m parameter is computed:

$$a_m = \arg\min_{a} \sum_{i=1}^{N} [-z_m - h(\mathbf{x}_i; a)]^2 \tag{3.10}$$

where a is the split point of $\mathbf{x}_i$ while $h(\mathbf{x}_i; a_m)$ is the base learner of the model (in this case of the study, it is a STR-Tree [103]). Then, the step size or expansion coefficient for the base learner $h(\mathbf{x}_i; a_m)$ in the direction of -z_m is computed by solving for:

$$\rho_m = \arg\min_{\rho} \sum_{i=1}^{N} L\left(y_i, F_{m-1}\left(\mathbf{x}_i\right) + \rho h\left(\mathbf{x}_i; a_m\right)\right), \tag{3.11}$$

where ρ is the weight of each new tree in the direction of -z_i. After the above two-step least square procedure, the model F_m is updated at the m-th step:

$$F_m\left(\mathbf{x}\right) = F_{m-1}\left(\mathbf{x}\right) + \rho_m h\left(\mathbf{x}_i; a_m\right), \tag{3.12}$$

and the predictions of the final model given by

$$\hat{y_i} = \hat{F}_M\left(\mathbf{x}_i\right) = \hat{F}_0 + \sum_{m=1}^{M} \rho_m h\left(\mathbf{x}_i; \hat{a}_m\right), \tag{3.13}$$

where M is the total number of base learners and F_0 is the initial estimation.

Because the shrinkage parameter λ controls the learning rate of the process and reduces the risk of overfitting. Hence, the updated equation for equation 3.12 and 3.13 becomes:

$$F_m\left(\mathbf{x}_i\right) = F_{m-1}\left(\mathbf{x}_i\right) + \lambda \rho_m h\left(\mathbf{x}_i; a_m\right), \tag{3.14}$$

and

$$\hat{y_i} = \hat{F}_M\left(\mathbf{x}_i\right) = \hat{F}_0 + \sum_{m=1}^{M} \lambda \rho_m h\left(\mathbf{x}_i; \hat{a}_m\right) \tag{3.15}$$

where $0 < \lambda \leq 1$. Algorithm 3.2 presents the simplified **BooST** model for quadratic loss. Although the shrinkage parameter λ is not required in estimating equation 3.10 and equation 3.11, theoretically and empirically results has shown its necessity for convergence and consistency of Boosting [107]. More theoretical details are available in literature [1].

3.2.2 Concrete Data Collection and Summarization

In this study, a database containing the 28 days compressive strength of concrete was created. The database contained experimental results from investigations presented in published articles on CCS [16, 79, 108–122] as summarized in Table 3.3.

Table 3.2. Simplified BooST training steps (Adapted from [1]).

Algorithm: BooST

initialization $\mathrm{F}_0 = \bar{y} := \frac{1}{N}\sum_{i=1}^{N} y_i$;
for m = 1, ..., M **do**
    make $\mathrm{z}_m = y - F_{m-1}$;
    grow a STR-Tree in z_m having sets of terminal nodes $\mathbb{T}_m$, $\hat{z}_{i,m} = \sum_{k\in\mathbb{T}_m} \beta_{k,m} B_m(\mathbf{x}_i; a_{k,m})$
    make $\rho_m = \arg\min_{\rho} \sum_{i=1}^{N} [z_{i,m} - \rho \hat{z}_{i,m}]^2$;
    update $\mathrm{F}_m = F_{m-1} + \lambda \rho_m \hat{z}_m$
end

The database contained 2456 datasets of mixture composition with comparable physical and chemical properties from which 848 datasets consisting of 28 days CCS data were used for this study. These mixture compositions are considered to be the original (raw) data, whereas the mixture proportioning of these compositions as obtained in several literatures are considered to be the non-dimensional data. Both the raw data, their proportioning (non-dimensional data) are considered explanatory variables, while the 28 days CCS is considered to be the response variable.

A graphical distribution and correlation plots of all variables are shown in Figure 3.2. From the figure, the unit values of both the explanatory and response variables in both axes (x-axis and y-axis) are the same as given in Table 3.4. The values for the correlation in the figure suggests the relationship values among the variables. Although the values of interest are the corresponding values relative to **CCS**, which serve as the response variable. A positive value indicates a positive relationship and vice versa. Diagonally displayed in the figure depicts the distribution of all variables, where the response variable (**CCS**). The series of scatter plots below the diagonally displayed distribution of Figure 3.2 shows the representation of individual sample observations of all the variables. These visualization scatter-plots support the correlation values of all explanatory variables. Along the bottom row of the figure maps the correlation plots of explanatory variables to the response variable (**CCS**).

The classification of all variables for the case of this study is identified as fol-

Table 3.3. Compressive strength database.

References	Laboratory	No. of concrete mixtures	Sample sizes	Curing age (days)	Type and condition of curing	Type of cement	Type of filler	Type of aggregate	
								FA	CA
Videla and Gaedicke [109]	Chile	40	200	1, 3, 7, 28 and 56	Cured at a temperature of 20 ± 3 °C and relative humidity [RH] >90%	Rapid hardening portland blast-furnace slag cement (ASTM C 595) [123]	Silica fume	Crushed siliceous fine aggregate with fineness modulus of 2.93	Crushed siliceous coarse aggregate with maximum sizes of 12 mm and 20 mm
Yeh [16] Yeh [108]	Taiwan	Combined mixtures	1030 103	3, 7, 14, 28, 56, 90, 100, 180, 270 and 365	[Moist] Normal condition	Ordinary portland cement (ASTM type I)	Blast-furnace and fly ash	Natural sand	Natural gravel <20 mm
Sobhani et al. [79]	Iran	32	96	28	Moist	Ordinary portland cement (ASTM C 150 type II) [124]	Silica fume and siliceous filler (with more than 99.0% SiO_2)	Fine aggregate with fineness modulus of 3.2 and specific gravity of 2.53 g/cm^3	Coarse aggregate with specific gravity of 2.56 g/cm^3
Kim et al. [110]	Korea	8	32	3, 7, 28, and 90	[Moist] Cured at a temperature range of 23.0 ± 1.0 °C	Ordinary portland cement (type I)	Fly ash (class F)	Sea sand with fineness modulus of 2.73 and specific gravity of 2.58 g/cm^3	Crushed stone with maximum size of 19 mm, fineness modulus of 6.83 and specific gravity of 2.61 g/cm^3
Lam et al. [111]	Hong Kong	24	192	3, 7, 28, 56, 90 and 180	Cured in water at 27 °C according to Hong Kong practice	Ordinary portland cement (ASTM type I)	Fly ash (class F) and condensed silica fume	Natural river sand	Crushed granite with sizes of 10 mm and 20 mm
Leung et al. [112]	Kazakhstan	10	10	28	Cured in water (dried at 105 °C for 24 hours and stored in a desiccator for further 24 hours) before testing.	Ordinary portland cement	Fly ash (ASTM C 618-99 class F) [125], silica fume with average diameter of 0.1 μm (ASTM C 1240-99) [126]	Natural river sand	Crushed granite with maximum size of 10 mm
Liu [113]	United Kingdom	5	20	7, 28, 90 and 180	Moist	CEM I 42.5N (BS EN 197-1 [127])	Fly ash	Sand 0/4 mm with fineness modulus of 2.78	4/10 mm and 10/20 mm gravel
Nepomuceno et al. [115]	Portugal	60	120	7 and 28	Cured in water at a temperature of 20 ± 2 °C	CEM I 42.5R and CEM II/B-L32.5N [128]	Limestone powder, Fly ash and granite filler	Natural sand with fineness modulus of 1.49 and 2.71	Crushed granite 3/6 and 6/15 with fineness modulus of 5.08 and 6.47 respectively
Bouzoubaa and Lachemi [117]	Canada	9	27	1, 7 and 28	[Moist] Cured in lime-saturated water at a temperature of 23 ± 2 °C at 100% relative humidity	Ordinary portland cement (ASTM type I)	Fly ash (ASTM class F)	Natural sand	Crushed limestone with maximum size of 19 mm
Ghezal and Khayat [118]	Canada	21	21	1 and 28	[Moist] Cured in lime-saturated water	CSA Canadian cement (type 10)	Ground limestone filler	Natural siliceous sand with fineness modulus of 2.36	Crushed limestone with a maximum size of 20 mm
Bui et al. [119]	United States of America	14	14	28	Normal condition	Ordinary portland cement	Fly ash	Natural fine aggregate	Natural coarse aggregate
Oner and Akyuz [116]	Turkey	32	224	7, 14, 28, 63, 119, 180 and 365	[Moist] Cured in lime-saturated water at a temperature of 20 ± 2 °C (ASTM C 192-88) [129]	CEM I 42.5 [128]	Ground granulated blast-furnace slag GGBS (ASTM C 989 grade 100 class) [130]	Crushed limestone with fineness modulus of 2.25	Crushed limestone of 19 mm maximum size with fineness modulus of 5.61
Sonebi [120]	United Kingdom	21	90	7, 28, and 90	Normal condition	CEM I 42.5N	Pulverished fuel ash	Quartzite sand with fineness modulus of 2.74	Crushed basalt aggregate with nominal size of 20 mm
Wongkeo et al. [122]	Thailand	36	144	3, 7, 28 and 90	Cured in lime-saturated water at 23 ± 1 °C (ASTM C109)	Ordinary portland cement (type I)	Fly ash and undensified silica fume (grade 920-U)	River sand	Crushed limestone with 82.75% size of 19 mm
Patel et al. [121]	Malaysia	21	42	1 and 28	Normal condition	Ordinary portland cement	Fly ash	Natural fine aggregate	Natural coarse aggregate
Johari et al. [114]	Malaysia	13	91	1, 3, 7, 14, 28, 90, 180 and 365	Moist	Ordinary portland cement	Silica fume, fly ash, metakaolin and ground granulated blast-furnace slag (GGBS)	Natural river sand	Quartzite natural gravel with nominal maximum size of 10 mm

lows:

- *Original explanatory variables known as raw data*: Cement (C), Blast furnace slag (Bfs), Fly ash (F_{ash}), Silica fume (SF), Water (W), Fine aggregate (FA), Coarse aggregate (CA), Superlasticizer (Su), constitute eight original explanatory variables.

- *Non-dimensional explanatory variables*: Water to Cement ratio (W/C), Fine aggregate to total aggregate ratio (FA/T_A), Water to total material content ratio (W/T_M), Total aggregate to cementitious material content ratio (T_A/C_{mt}) and Water to cementitious material content ratio (W/C_{mt}). Due to the wide variety of cementitious materials used for high strength concrete, in this study, Cement (C), Blast furnace slag (Bfs), Fly ash (F_{ash}) and Silica fume (SF) were used as cementitious materials.

- *Response variable*: 28 day compressive strength of concrete.

The summary of all attributes from the database consisting of 28 days CCS are shown in Table 3.4. Data sets from literature often contain some inaccuracies. For instance, some details about the fly ash used may be not be indicated, whereas different chemical composition may constitute admixtures such as the superplasticizers obtained from various sources or manufacturers [131]. However, the CCS is a function of these discrepancies of constituent materials. The interaction of these materials both in measured values as well as in their mix proportioning further complicates the computation and estimation of CCS. Therefore, the following machine learning computational techniques are employed to the concrete complex structure to model the CCS of HPC.

3.2.3 Concrete Data Pre-processing and Feature Engineering

To provide useful information of explanatory variables as well as provide accurate predictive models for CCS, the contemporary techniques reviewed in subsection 2.2 as well as the **BooST** technique from sub-section 3.2.1 are investigated. Since the estimate of the effects of explanatory variables (as shown in Figure 3.2) on CCS provides more interpretation in concrete properties, models from these techniques are developed which can advance the knowledge in the concrete domain

Table 3.4. Summary of the concrete compressive strength variables.

Attributes	Range		
	Minimum	Maximum	Average
Explanatory variables (raw data)			
Cement (C) (kg/m^3)	102.00	670.00	289.50
Blast furnace slag (Bfs) (kg/m^3)	0.00	440.00	58.95
Fly ash (F_{ash}) (kg/m^3)	0.00	420.00	85.11
Silica fume (SF) (kg/m^3)	0.00	93.00	3.34
Water (W) (kg/m^3)	95.00	390.40	185.50
Fine aggregate (FA) (kg/m^3)	354.20	1300.00	766.30
Coarse aggregate (CA) (kg/m^3)	477.00	1440.60	935.80
Superplasticizer (Su) (kg/m^3)	0.00	32.20	5.83
Explanatory variables (non-dimensional data)			
Water to Cement ratio (W/C)	0.25	1.88	0.76
Fine aggregate to Total aggregate ratio (FA/T_A)	0.20	0.68	0.45
Water to Total material content ratio (W/T_M)	0.04	0.18	0.09
Total aggregate to Cementitious material content (T_A/C_{mt})	1.67	11.05	4.13
Water to Cementitious material content ratio (W/C_{mt})	0.24	1.19	0.44
Response variables (output value)			
28 days- Concrete Compressive Strength (CCS) (MPa)	8.54	117.50	44.18

by acknowledging relevant non-dimensional explanatory variables, thus reducing the complexities and number of trial testings in the concrete laboratory. The experimental settings, as well as the pre-processing of data provided for the analysis of models in this study, are as follows:

Model development

Ten different models were developed in this study using the techniques above viz.; ANN, RT, SVM, NLR and BooST to predict the 28 days CCS of HPC. To determine the explanatory variables (excluding original data), a linear relationship between CCS and the non-dimensional variables usually analyzed as mix proportions of concrete constituents were obtained using correlation coefficients shown in Figure 3.3 as obtained by the simplified formula in Equation 3.16.

$$\mathrm{Cor}(x,y) = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}} \tag{3.16}$$

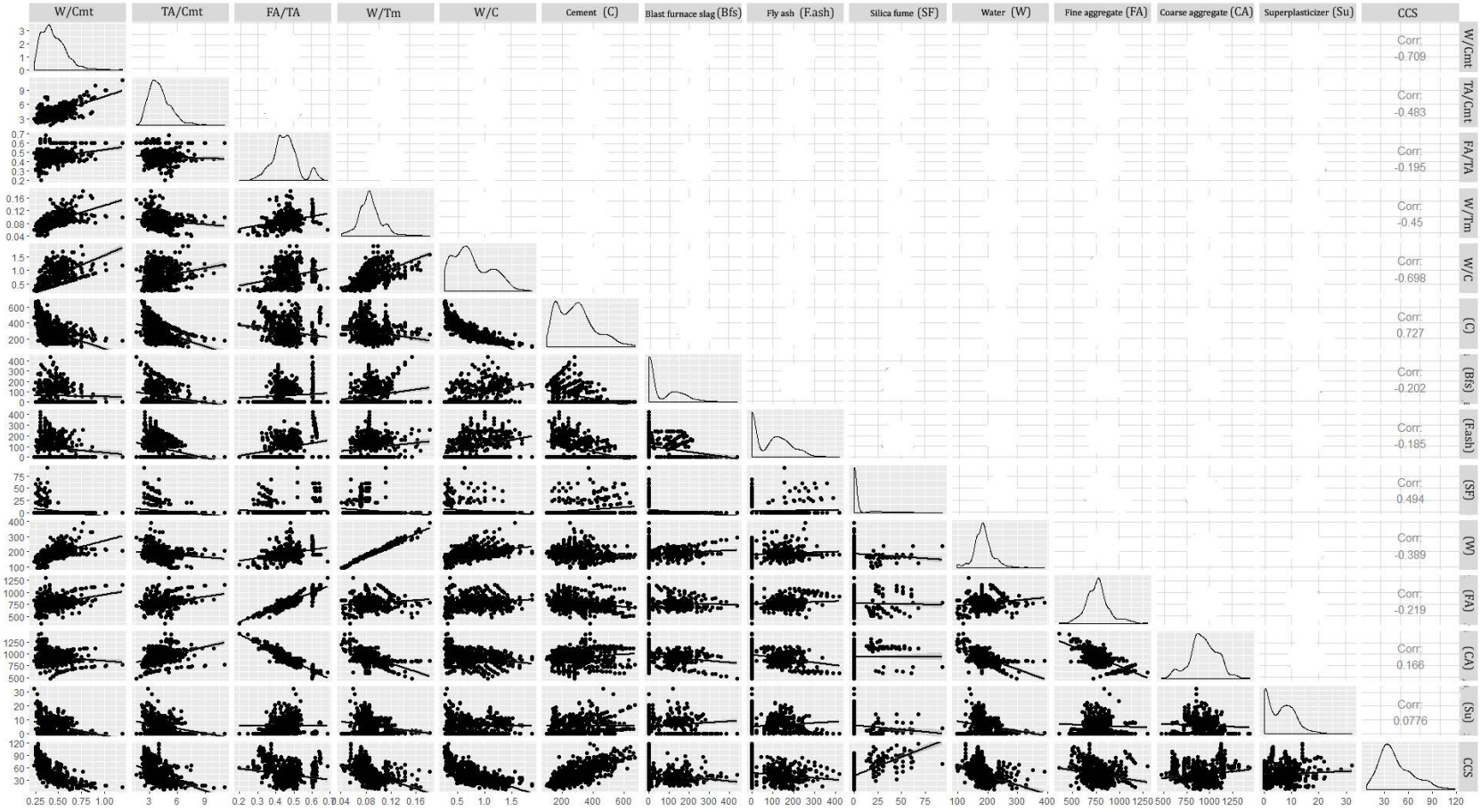


Figure 3.2. Graphical distribution plots for all variables. The units of all values of the variables at x-axis and y-axis are as same in Table 3.4. The values insider the plot correspond to the correlation values associated with all individual variables for concrete production, while the diagonal distributions depicts the distribution of the variables.

where the correlation coefficient of an explanatory variable x relative to the response variable y is represented by $\mathrm{Cor}(x,y)$, $\bar{x}$ and $\bar{y}$ represent the individual average values of the explanatory and response variable respectively while N represents the number of data samples.

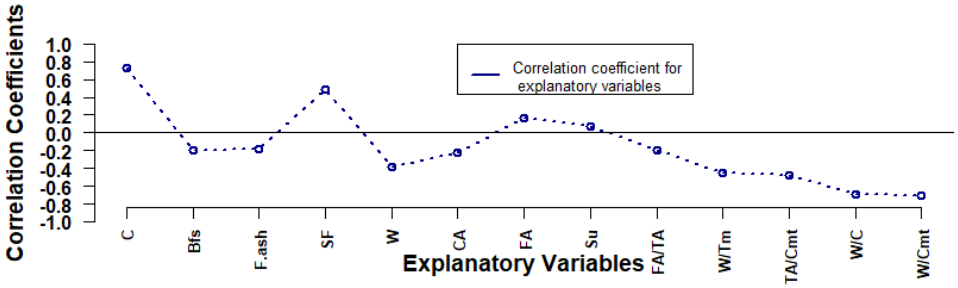


Figure 3.3. Correlation coefficients for all explanatory variables with respect to CCS.

The procedure to validate the effect of mix proportioning of concrete constituent to CCS was adopted from experiments in the laboratory and the literature [79, 132–134]. In simplifying the procedure to understand the explanatory-response variable relationship, the sets of models were divided into three sets. The first set model for each technique (MLP 1, RT 1, SVM 1, NLR 1 and BooST 1) consists of using the original explanatory variables to estimate CCS. This aids in understanding the predictive capability of each technique in predicting CCS using concrete constitutes with no mixture ratio of the constituents. Details of set 1 are shown in Table 3.5.

Table 3.5. Summary of the model development procedure for the first set.

Sr. No.	Explanatory variables	MLP model	RT model	SVM model	NLR model	BooST model
1	C, Bfs, F_{ash}, SF, W, FA, CA, Su	MLP 1	RT 1	SVM 1	NLR 1	BooST 1

In the second set, with decreasing order of correlation coefficient value of the non-dimensional explanatory variable (FA/T_A, T_A/C_{mt}, W/C_{mt}, W/C) to CCS, the individual non-dimensional explanatory variables were added to other original explanatory variable without duplicating their presence as part of the original explanatory variable in the set. Such a procedure allowed the study of the influence of

these individual non-dimensional explanatory variable effects without duplicating their presence on other material constituents for CCS. It also allowed for sensitivity analysis study in predicting CCS using each technique. Summary of each model is shown in Table 3.6.

Table 3.6. Summary of the model development procedure for the second set.

Sr. No.	Explanatory variables	MLP model	RT model	SVM model	NLR model	BooST model
1	C, Bfs, F_{ash}, SF, W, Su, FA/T_A	MLP 2	RT 2	SVM 2	NLR 2	BooST 2
2	W, Su, T_A/C_{mt}	MLP 3	RT 3	SVM 3	NLR 3	BooST 3
3	Bfs, F_{ash}, SF, FA, CA, Su, W/C	MLP 4	RT 4	SVM 4	NLR 4	BooST 4
4	FA, CA, Su, W/C_{mt}	MLP 5	RT 5	SVM 5	NLR 5	BooST 5

Lastly, in the third set, each non-dimensional explanatory variables were added to the first set consecutively in their decreasing order of correlation coefficient to the compressive strength of concrete (CCS). Thus, the FA/T_A was added to all original variables, after which T_A/C_{mt} was added to the previous sum of explanatory variables and so on. Such a procedure allowed to study the combined effect of each added parameter on the model performance. A table constituting the model developed via this methodology is shown in Table 3.7.

Table 3.7. Summary of the model development procedure for the third set.

Sr. No.	Explanatory variables	MLP model	RT model	SVM model	NLR model	BooST model
1	C, Bfs, F_{ash}, SF, W, FA, CA, Su, FA/T_A	MLP 6	RT 6	SVM 6	NLR 6	BooST 6
2	C, Bfs, F_{ash}, SF, W, FA, CA, Su, FA/T_A, W/T_M	MLP 7	RT 7	SVM 7	NLR 7	BooST 7
3	C, Bfs, F_{ash}, SF, W, FA, CA, Su, FA/T_A, W/T_M, T_A/C_{mt}	MLP 8	RT 8	SVM 8	NLR 8	BooST 8
4	C, Bfs, F_{ash}, SF, W, FA, CA, Su, FA/T_A, W/T_M , T_A/C_{mt}, W/C	MLP 9	RT 9	SVM 9	NLR 9	BooST 9
5	C, Bfs, F_{ash}, SF, W, FA, CA, Su, FA/T_A, W/T_M , T_A/C_{mt}, W/C_{mt}	MLP 10	RT 10	SVM 10	NLR 10	BooST 10

The models used by the six different machine learning techniques used for this study were implemented by an open-source statistical software called **R**. Table 3.8 lists the individual hyper-parameters of each techniques that were tuned to obtained the best results for the models.

Experimental settings

Figure 3.4 provides a detailed description of the machine learning procedures performed for this research study. As illustrated by the given flowchart, after the data

Table 3.8. Model hyper-parameter range of settings.

Technique	Hyper-parmeters	Settings
MLP	Hidden layer [135]	1
	Learning function rate	$0.2 \sim 0.3$
	Learning function	backpropagation
	Momentum	0.2
RT	Pruning	True
	Initial count	0
	Complexity parameter (cp)	$0.01 \sim 0.02$
	cv.Fold	10.0
SVM	Kernel	RBF
	Regularization term (C)	1.0
	Regression precision (ε)	0.1
	Type	eps-regression
	gamma	$0.08 \sim 0.3$
NLR	Intercept	True
BooST	d_max	4
	gamma	[0.5,5] by 0.01
	Shrinkage parameter (v)	$1.0 \sim 0.2$

pre-processing, the collected data was randomly divided into two sets where 80% of the data set was used of training the unique models of each technique while the remaining 20% was held-out for testing after the training sets have been validated by the range of model parameters settings.

To minimize the bias associated with the random sampling of the training and hold-out data samples when comparing the predictive accuracy of machine learning models, the use of *k*-fold cross-validation algorithm is adopted [136, 137]. In an early study of cross-validation and bootstrap for accuracy estimation and model selection, Kohavi [138] reported that ten-fold cross-validation yields optimal computational time and reliable variance. Therefore, in this study, a stratified ten-fold cross-validation is adopted. This procedure divides the training data into ten sets or folds. In each fold, 90% of the datasets are used in training, while the remaining 10% is used for validation of the trained data after being tuned for efficiency by adequate hyper-parameters set out in Table 3.8. With the hyper-parameter tuning for the models of each technique, approximately two to four hours was required for training each model, with exemption to the NLR and RT models which took ap-

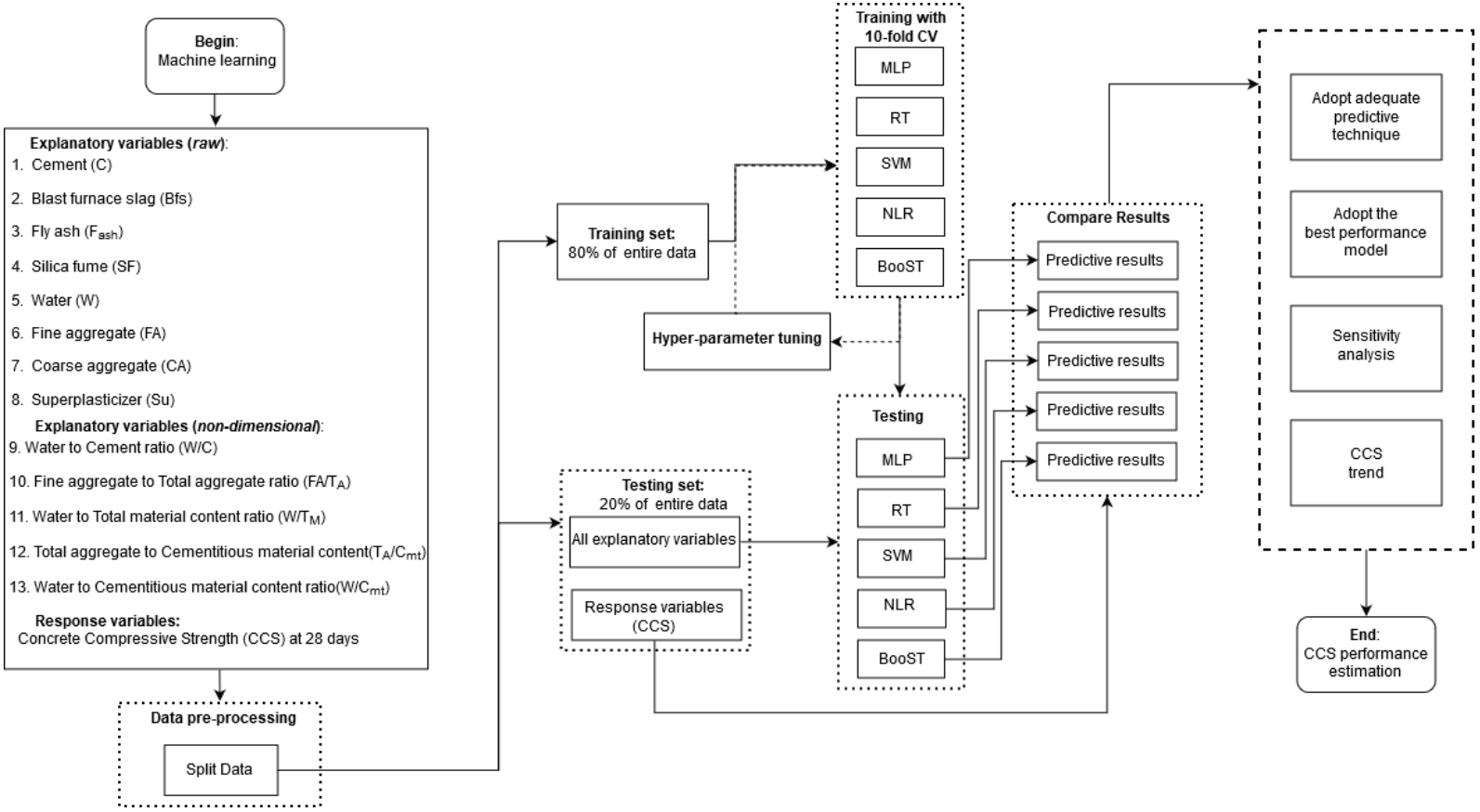


Figure 3.4. Machine learning computational flowchart.

proximately one to five seconds. At ten rounds of training for model building, the 90% subsets of each fold are selected for training while one-tenth (10%) are used for validation, as shown in Figure 3.5. The trained models are afterwards evaluated by comparing predicted CCS values with the held-out testing sets.

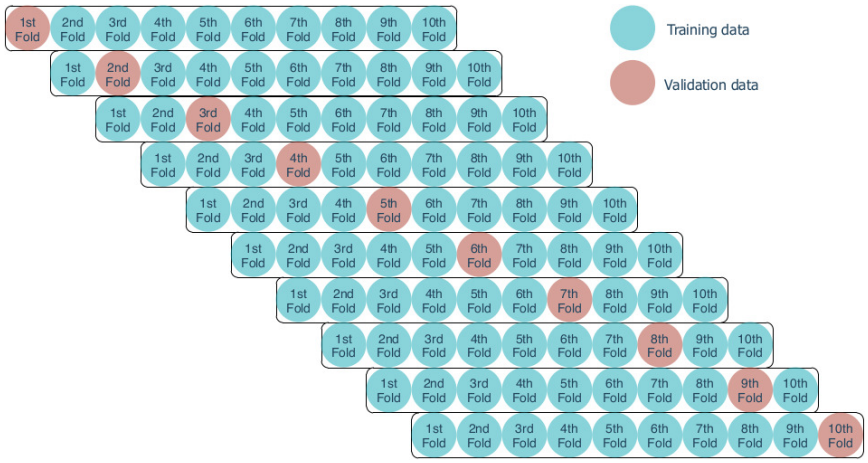


Figure 3.5. Graphical representation of 10-fold cross-validation procedure.

The cross-validation calculation for accuracy estimation is expressed as an average of the k individual accuracy. If the cross-validation accuracy is denoted as CV and the accuracy measure of each k fold is A, therefore the CV of the entire cross-validation would be expressed as follows:

$$CV = \sum_{i}^{k} \frac{A_i}{k} \tag{3.17}$$

3.2.4 Predictive Performance Evaluation Metrics

The following performance statistics are widely used in concrete predictive evaluation and were employed to evaluate the proposed machine learning techniques in this study:

Coefficient of determination

The coefficient of determination, (R^2) is a measure of how well the explanatory variables account for the measured response variable. It provides the goodness-of-fit for a model. The higher the R^2 value, the better the predictive power of such a model. The mathematical expression for R^2 is given as follows:

$$R^2 = \left(\frac{N\Sigma y_i \hat{y}_i - (\Sigma y_i)(\Sigma \hat{y}_i)}{\sqrt{N(\Sigma y_i^2) - (\Sigma y_i)^2}\sqrt{N(\Sigma \hat{y}_i{}^2) - (\Sigma \hat{y}_i)^2}} \right)^2 \tag{3.18}$$

Mean absolute error

The Mean Absolute Error (**MAE**) is a statistic that measures the average magnitude of errors of a models predictive values. The **MAE** is commonly used in evaluating other engineering predictive models. The **MAE** equation is given as follows:

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

Root mean squared error

The Root Mean Squared Error (Root Mean Square Error (**RMSE**)) is a statistical measure of predictive accuracy. It is the square root of the mean square error. The **RMSE** measures the amount of variance in the response variable that can be explained by a model. The **RMSE** is given by the equation:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2} \tag{3.20}$$

In Equation (3.18), (3.19) and (3.20) above, y_i and $\hat{y}_i$ represent the actual value and the predicted value of CSC respectively while N represents the number of data samples. A lower **MAE** and **RMSE** value would signify lower values in overall errors in a model, which would mean better predictive power of such model.

3.2.5 Summary

This section of the chapter presented the working principle of Boosting Smooth Transition Regression Trees (BooST) as a machine learning technique proposed for the prediction of concrete compressive strength (**CCS**). The BooST technique can provide sufficient estimation of variable effects in nonlinear models, thereby providing adequate interpretation about the mapping between the covariates and the dependent variables in analyzing and prediction for non-linear regression problems. Additionally, the section also presented dataset and the experimental settings for several models created for analysis of **CCS**. Computational results and discussions are presented in section 4.1 of Chapter 4.

Chapter 4

Experimental Results

Two significant study demonstrations that aid performance assessments of optimal concrete compressive strength have been presented and discussed: Design of experiment methodology and Prediction of concrete compressive strength. Box-Behnken Design (**BBD**) has been proposed as a response surface method (**RSM**) for the design of concrete experiment where conditions such as experimental boundary constraints or expensive experimental runs are met. While Boosting Smooth Transition Regression Trees (**BooST**) have been proposed as a machine learning technique to estimate the **CCS** with knowledge of concrete constituents.

The computational results and discussions of the studies as earlier discussed are presented in this chapter.

4.1 Analytical Results of Response Surface Methodology for Optimal Concrete Compressive Strength

The statistical assessment via Analysis of Variance (**ANOVA**) and findings from the optimization study for CCS due to the RSM approach is presented in this section of the study.

4.1.1 Statistical Analysis of Design Experiment

The **ANOVA** was carried out to investigate the significance of the input variables W/C_{mt}, Q_{CmT}, FA/T_A and their interaction effects on the response (CCS). This

analysis is also employed as a check on the adequacy of the eventually fitted surface model if it would be an adequate approximation of the true response function. Additionally, with the challenges in examining the influence of several explanatory variables concerning corresponding changes in the dependent variable, ANOVA serves beneficial in describing the interactions between the variables [139]. Analysis of the experimental data was evaluated at a 5% significant level, that is to say, that the statistical significance of the experimental input parameters (W/C_{mt}, Q_{CmT}, FA/T_A) on the response (**CCS**) was examined at a 95% confidence level. The significance of a variable is indicated by the p-value ($p \leq 0.05$). It is to be noted that a check for a pure quadratic curvature effect was also carried out and its significance indicates a second-order model being adequate to represent the relationship between the response and its sets of input variables.

Table 4.1. Analysis of variance (ANOVA) for the RSM model fit for the concrete design (performance statistics R^2, $R^2{}_{adj}$, $R^2{}_{pred}$ and RMSE, are also provided for the analysis) with the response = (CCS).

Source	Df	Sum of Sq.	Mean Sq.	F value	P-value
FO(x_1, x_2, x_3)	3	1069.30	356.43	153.11	2.10e-07
TWI(x_1, x_2, x_3)	3	240.43	80.14	34.43	6.38e-05
PQ(x_1, x_2, x_3)	3	30.54	10.18	4.37	0.04
Residuals	8	18.62	2.33	–	–
Lack of fit	2	0.61	0.31	0.10	0.90
Pure error	6	18.01	3.00	–	–
$*$ R^2: **0.9863**,	$R^2{}_{adj}$: **0.9709**,	$R^2{}_{pred}$: **0.9409**,	RMSE: **1.5258**		

From the ANOVA table above, FO(x_1, x_2, x_3), TWI(x_1, x_2, x_3), and PQ(x_1, x_2, x_3) represent the linear, quadratic and interaction effects respectively. The sums of squares for the linear effect FO(x_1, x_2, x_3) is considered simultaneously for the three linear terms (x_1, x_2, x_3) in the model. Same applies for both the interaction (cross product) effects TWI(x_1, x_2, x_3) and the quadratic effect PQ(x_1, x_2, x_3). The terms in these effects are all considered collectively.

The ANOVA table shows that all linear, interaction and quadratic terms are significant. However, since the check for the pure quadratic curvature effect PQ(x_1, x_2, x_3) is significant, it indicates that the quadratic model provides a good approximation to the response function over the experimental region. This, therefore, implies that

the linear model would not adequately represent the relationship between the factor variables and the response (CCS) [101].

The lack of fit from Table 4.1 is insignificant ($p > 0.05$), indicating that predictions made from the quadratic model in this experiment can be considered as accurate as running additional experiments, as long as no lurking variable is altered before running the addition experiment within the experimental boundary. In other words, the insignificance of the lack of fit test indicates that the general quadratic model is adequately suited for prediction and providing a good approximation to the true response (CCS) function over the experimental region. This outcome of the quadratic model from the ANOVA table also supports and is consistent with the findings of [16] about the non-linearity of concrete.

Since the linear, quadratic and interaction effects are all significant, the individual terms in the general quadratic model are to be retained. As a result of the empirical modeling of the experimental design using RSM, the second-order model with the ultimate goal to determine optimum response (CCS) is given. This model is presented using the coded variables for easy interpretation of results as earlier discussed in the previous section (see sub-section 3.1.2).

$$\begin{aligned} CCS = \; & 47.440 - 10.906x_1 + 6.389x_2 + 4.352x_3 - 5.623x_1x_2 - 1.155x_1x_3 \\ & + 3.022x_2x_3 - 0.130x_1^2 - 1.698x_2^2 + 4.324x_3^2 \end{aligned} \tag{4.1}$$

As a result of mathematical modeling of experimental design using RSM technique, the built RSM model to predict CCS in this study is given in Equation 4.1. The adequacy of this built RSM second-order model in Equation 4.1 is checked with the magnitude of the coefficient of determination statistics R^2, $R^2{}_{adj}$, $R^2{}_{pred}$ and as well as **RMSE** given in Table 4.1. These statistics, measure how well the input variables considered in the model accounts for the response as well as the resulting residuals. Also, it can be observed from Equation 4.1 that among all variable coefficients, x_1 has the highest absolute value amongst other variable coefficients. This can indicate that the x_1, which is the coded variable of W/C_{mt} has a stronger influence in the model. The percentage of influence of each parameter in the built RSM model is illustrated in Figure 4.1. This effect of W/C_{mt} as

a significant influence in concrete is consistent with the findings from literature studies[16, 87, 92–97, 140, 141].

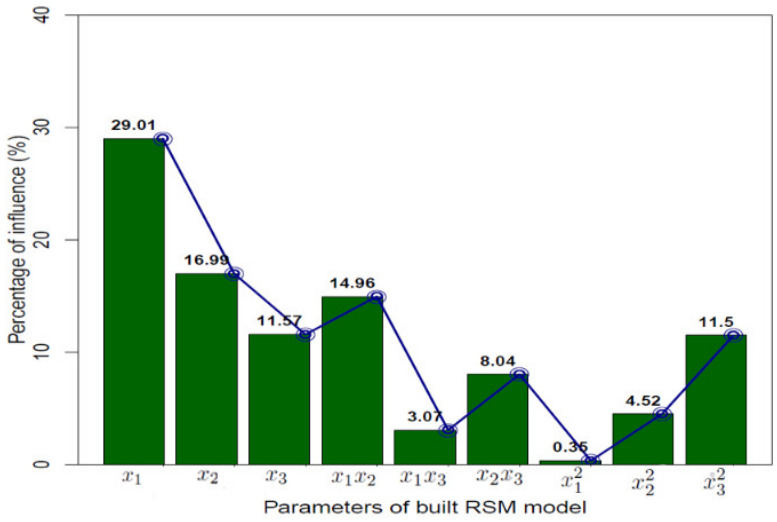


Figure 4.1. Percentage of influence of parameters in the built model for CCS.

However, according to Figure 4.1 showing the percentage of influence of the model parameters, the most interactive parameter effect is found to be x_1 and x_2, likely due to the binding effect of the cementitious content which is significant in concrete.

Though the residual analysis is vital for the RSM technique and is usually depicted by some diagnostic plots according to the residual analysis assumptions, these assumptions are i) normality and ii) equal (constant) variance. The normality assumption test plots the studentized residual against the theoretical quantiles to verify if the data plausibly came from a normal theoretical distribution. If the residuals plot approximately along a straight line, the normality assumption is satisfied [56]. The normal quantile-quantile plot for the RSM model in this study is depicted in Figure 4.2. The figure shown portrays to satisfy the normality assumption as the studentized residuals plot approximately along a straight line. The assumption check for equal variance is verified by plotting the studentized residual against the predicted response variable (CCS) built for the RSM model, as shown

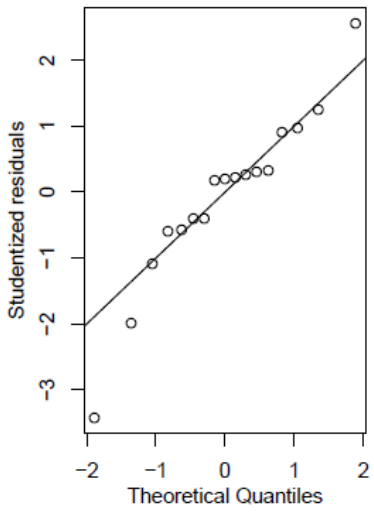


Figure 4.2. Normal quantile-quantile plots of studentized residuals.

in Figure 4.3. The idea behind the plot is to verify that the studentized residual scatters randomly without any definite pattern, suggesting a constant variance of the original measurement for all measured (actual) response values [56]. With verification of the stated assumptions, the built RSM model appears to be a good approximation of the true response surface.

The coefficient of determination, R^2 is a measure of the variability of the response value in the fitted model. As shown in Table 4.1, the R^2 is found to be 0.9863, which approximately equals 1. This indication signifies that variations in the response variable are well explained by the input factors being a good fit of the model. However, an increased value of R^2 does not necessarily imply the model is a good one because having additional factors to a model enhances the value of R^2 regardless whether such added factor is statistically significant [56, 142]. However, the $R^2{}_{adj}$ is a statistic that is adjusted for the given number of factors which are significant in a model. Unlike R^2, the $R^2{}_{adj}$ will often decrease when non-significant factors are added to a built model [142]. Notice, that in Table 4.1 the value of $R^2{}_{adj}$ is 0.9709 which is slightly lower than R^2, although, it still approximates to 1 which validates that the input factors well explain the variations in the response

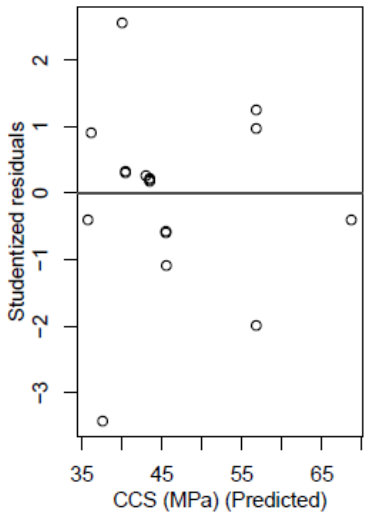


Figure 4.3. Studentized residuals against predicted response (CCS) for the built RSM model.

variable by a good fit of the model. The other statistic, R^2_{pred} is a measure of the variation in new data within the experimental region that is explained by the RSM model. As shown from Table 4.1 the value of R^2_{pred} for the RSM model is 0.9409 indicating that the model can explain approximately 94.09% of the variability in the estimation of new response value obtained, compared to the 98.63% variability in the dataset within the experimental region. Myers et al. [56] suggested that if the difference between the magnitudes of the two statistics R^2_{adj} and R^2_{pred} is less than 0.2, then the statistics are adequate. For the difference between the magnitudes of these two statistics for the built RSM model which is less than 0.2, it can be said that the statistics agree to signify that the input factors well explain variations in the response variable by a good fit of the model. Lastly, the root mean square error (**RMSE**) statistic assess the strength of fit for the model by considering how far off the model estimated values are from the observed or true values of CCS. The value of **RMSE** from Table 4.1 is approximately 1.53, which appears not to be very large. From results of the statistics R^2, R^2_{adj} and R^2_{pred} in this study, it can be concluded that Equation 4.1 is considered a good RSM model for optimization in

this study.

4.1.2 Optimization Study

Since the response surface model is fit for the experimental data of CCS at 28 days, therefore it is essential to establish the optimal CCS for the data. This generates also the optimal levels of the independent variables (W/C_{mt}, Q_{CmT}, FA/T_A) for the corresponding optimal CCS. In this study, the optimization was performed using ridge analysis, which for simplicity can be described as a method of steepest ascent. The method of steepest ascent is a procedure that moves sequentially through a path of ascent to attain a maximum value of response [142]. It operates as a subject of having the optimum to be at a distance of radius from the origin in the coded unit within the experimental region.

The optimal response (**CCS**) and the corresponding independent variables (W/C_{mt}, Q_{CmT}, FA/T_A) in this study were obtained using a statistical software "**R**" (open source software). This software contains several packages used for graphics and statistical computation. Lenth [100] demonstrated the call functions for the BBD and CCD for RSM. However, Lawson [101] and Borkowski [102] stated some discrepancies between CCD and BBD.

The optimization results are presented in Table 4.2. based on the experimental factors (W/C_{mt}, Q_{CmT}, FA/T_A) with the ranges of their respectively levels (0.3 to 0.45), (460 to 500)kg/m^3 and (0.4 to 0.5) as considered for the optimization process. With the considered factors and their respective levels given above, optimization to obtain the maximum strength was carried out with the Box-Behnken Design (BBD). After the optimization process given in this study, 15 different optimal results were generated and presented in Table 4.2. The results include the maximum levels of the factors satisfied to generate corresponding responses. However, from the optimization results, the maximum CCS resulted at 73.31MPa with optimal variables of W/C_{mt} at 0.308, Q_{CmT} at 493.44 and FA/T_A at 0.492. The 3-dimensional (3D) and contour plots of the CCS with mixtures values are illustrated in Figure 4.4.

From the 3D response surface plot in Figure 4.4, the magnitude of the compressive strength of HPC is plotted on the y-axis of the plot. The plots give graphical

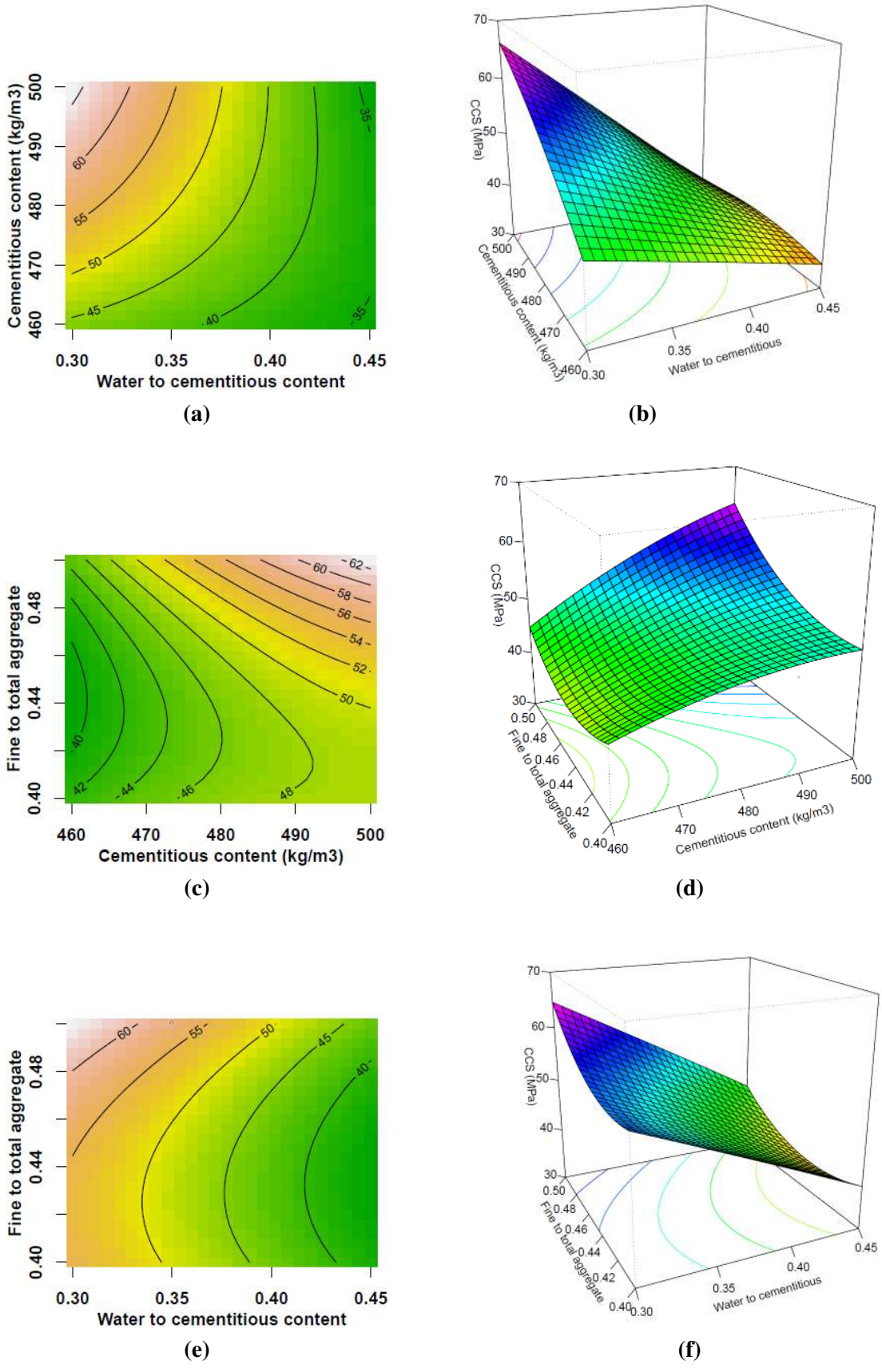


Figure 4.4. Response surface plots: (a, c, e) Contour plot, (b, d, f) 3D plot illustrating surface with a different variation of CCS having variable interactions W/C_{mt}, Q_{CmT}, FA/T_A.

Table 4.2. The Optimization Results for CCS.

W/C_{mt}	Q_{CmT} (kg/m^3)	FA/T_A	CCS (MPa)
0.375	480.000	0.450	47.44
0.369	480.980	0.452	48.81
0.363	481.960	0.454	50.23
0.358	482.960	0.456	51.74
0.352	483.960	0.458	53.29
0.347	484.960	0.461	54.93
0.342	485.960	0.464	56.65
0.337	486.940	0.467	58.43
0.332	487.920	0.470	60.31
0.328	488.860	0.474	62.25
0.324	489.820	0.477	64.29
0.320	490.740	0.481	66.41
0.316	491.640	0.485	68.62
0.312	492.540	0.489	70.90
0.308	**493.440**	**0.492**	**73.31**

view of the response (CCS) with variations of the variables within the experimental region.

4.2 Predictive Analytical Results for Concrete Strength Performance

In each concrete experiment from literature, most of the experimental boundaries of explanatory variables are all from individual references which can result to the tendency of the machine learning techniques used in these analyses to learn from and only behave with respect to a limited range of experimental boundaries. Hence, concerning specified CCS, this calls for investigation of concrete constituents with respect to high-performance concrete strength.

This section of the chapter presents the analytical results from the machine learning techniques discussed in predicting the compressive strength of concrete. The predictive techniques investigated in this study were used for the ten models developed, as explained in sub-section 3.2.3 and using tuned experimental settings

and cross-validation procedure in sub-section 3.2.3. To evaluate the **CCS** predication performance, three criteria are used in this study as earlier discussed: **RMSE**, R^2, and **MAE**. In statistics, RMSE and MAE are two metrics frequently used to evaluate the prediction modeling accuracy. RMSE is used to measure the differences between predicted values and the observed values while MAE measures the average magnitude of the errors without considering their direction. Since RMSE gives a relatively high weight to large errors, it is more useful when large errors are particularly undesirable. The equations of MAE and RMSE have been given in Equations (3.19) and (3.20), respectively. Equation (3.18) is the formula of R^2, which describes the fitting degree of models in a $[0, 1]$ range, where the larger the R^2 value is the better the model fits the data.

Concrete Compressive Strength Predictions

The respective developed models were evaluated as discussed in sub-section 3.2.4 using the mean absolute error (**MAE**), root mean squared error (**RMSE**) and the coefficient of determination, (R^2). The performance statistics of each model is shown in; Table 4.3 for MLP, Table 4.4 for RT, Table 4.5 for SVM, Table 4.6 for NLR and Table 4.7 for BooST.

Table 4.3. Performance statistics for models developed using MLP.

MLP model	MAE	RMSE	R^2
MLP 1	5.37	7.03	0.88
MLP 2	5.68	7.38	0.86
MLP 3	8.09	10.66	0.72
MLP 4	5.46	7.38	0.86
MLP 5	7.65	10.13	0.74
MLP 6	5.70	7.27	0.87
MLP 7	5.33	7.07	0.88
MLP 8	5.45	6.97	0.88
MLP 9	4.92	6.48	0.90
MLP 10	5.21	6.79	0.88

In the overall performance from each machine learning techniques in predicting CCS, the models developed from the NLR and RT performed lower as compared to models from MLP, SVM and BooST. As seen in Table 4.6 and Table 4.4, the error values of *MAE* and *RMSE* are larger, indicating inadequacy to predict CCS

Table 4.4. Performance statistics for models developed using RT.

RT model	MAE	RMSE	R^2
RT 1	8.59	11.11	0.69
RT 2	8.41	10.85	0.71
RT 3	8.02	10.47	0.73
RT 4	7.27	9.40	0.78
RT 5	7.80	9.92	0.75
RT 6	8.41	10.85	0.71
RT 7	8.95	11.27	0.68
RT 8	7.69	10.30	0.74
RT 9	7.81	10.14	0.74
RT 10	7.65	9.38	0.78

Table 4.5. Performance statistics for models developed using SVM.

SVM model	MAE	RMSE	R^2
SVM 1	4.76	6.55	0.89
SVM 2	4.96	6.83	0.88
SVM 3	7.11	10.00	0.76
SVM 4	5.13	7.21	0.87
SVM 5	6.86	8.76	0.81
SVM 6	4.85	6.64	0.89
SVM 7	4.77	6.48	0.89
SVM 8	4.67	6.43	0.89
SVM 9	4.53	6.21	0.90
SVM 10	4.61	6.36	0.90

Table 4.6. Performance statistics for models developed using NLR.

NLR model	MAE	RMSE	R^2
NLR 1	6.88	8.91	0.80
NLR 2	7.25	9.42	0.78
NLR 3	10.62	13.48	0.55
NLR 4	9.03	11.50	0.68
NLR 5	10.61	13.15	0.58
NLR 6	6.79	8.76	0.81
NLR 7	6.68	8.71	0.81
NLR 8	6.57	8.65	0.81
NLR 9	6.56	8.66	0.81
NLR 10	6.59	8.65	0.81

with wider range values of explanatory variables. Additionally, from the presented tables establish that the overall R^2-values of the predictive models were higher for MLP, SVM and BooST than for NLR and RT. In other words, the models from

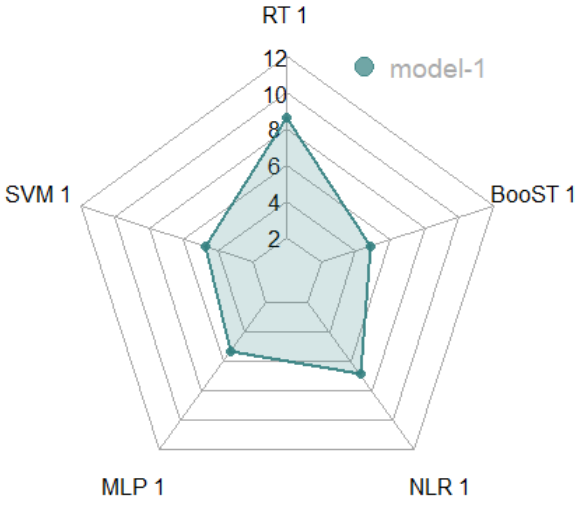
Table 4.7. Performance statistics for models developed using BooST.

BooST model	MAE	RMSE	R^2
BooST 1	4.84	6.21	0.90
BooST 2	5.16	6.98	0.88
BooST 3	6.86	10.46	0.73
BooST 4	5.02	6.88	0.88
BooST 5	6.88	10.58	0.72
BooST 6	4.90	6.38	0.90
BooST 7	4.89	6.39	0.90
BooST 8	4.68	6.21	0.90
BooST 9	4.67	6.23	0.90
BooST 10	4.55	6.31	0.90

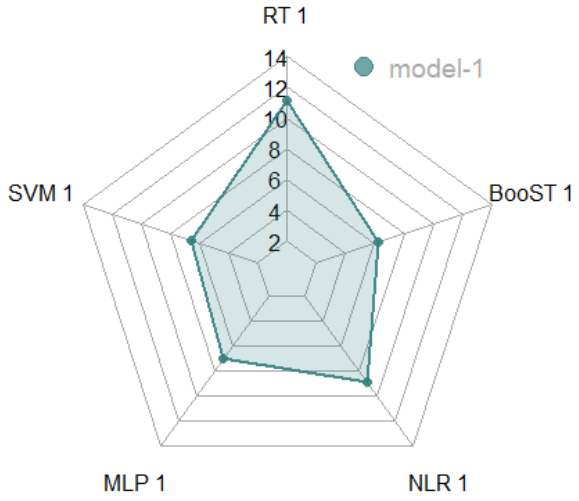
these advanced techniques (MLP, SVM and BooST) were able to explain more of the variance in CCS for concrete data with a wide range of explanatory (input) values. However, additional computational training time were required for these advanced techniques. With the hyper-parameter tuning for the models of these advanced techniques, approximately two to four hours were required for training each model, while the NLR and RT models took approximately one to five seconds for training time.

For the first set of models described in Table 3.5, while the performance of MLP 1, SVM 1 and BooST 1 were some what similar, the BooST model displayed the highest R^2-value and lowest RMSE. Figure 4.5 depicts the performance statistics for the first set of models with original explanatory variables as described in Table 3.5.

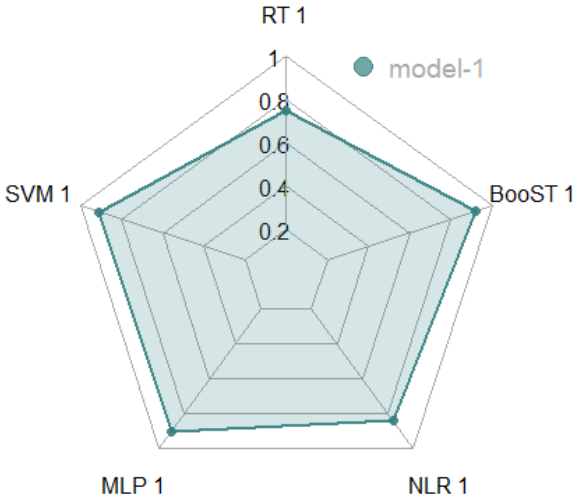
The values shown in the radar chart (spider plot) of Figure 4.5a, 4.5b and 4.5c are representative of same units with MAE, RMSE and R^2 respectively. Since lower values of MAE and RMSE indicates a better predictive model, a good model in the radar chart (spider plot) shown in Figure 4.5a and 4.5b will tend to move closer to the inner pentagon with lower values. While a high value of R^2, which indicates a better model will have the individual plots diverging to the larger pentagon. Though the BooST 1 model indicated dominance over the performance of SVM 1 and MLP 1, each technique is well-suited for predicting CCS with original explanatory variables, and therefore selection between the techniques can be made based on their ease of implementation.



(a) Mean absolute error (MAE).



(b) Root mean square error (RMSE).



(c) Coefficient of determination (R^2).

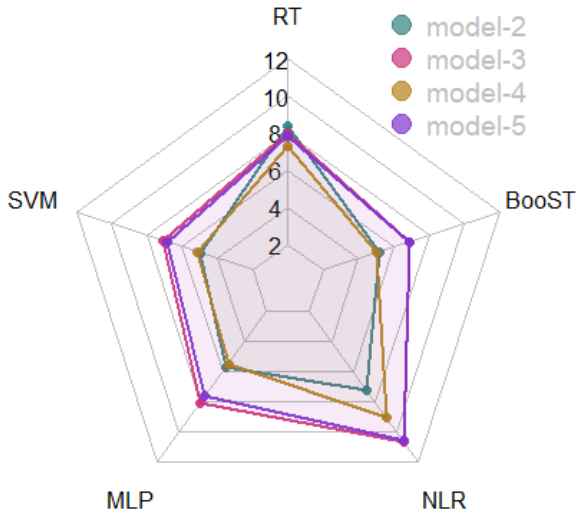
Figure 4.5. Radar chart of performance statistics for first set of models.

With no increase in model performance of the second set of models described in Table 3.6 over the first set of models, model assessment of the second set emphasizes the performance of model-2 and model-4 in the set. It shows the importance of using non-dimension explanatory variables W/C and FA/TA as alternatives in predicting CCS as presented in Table 4.3 - 4.7. The radar chart(spider plot) of the second set is shown in Figure 4.6.

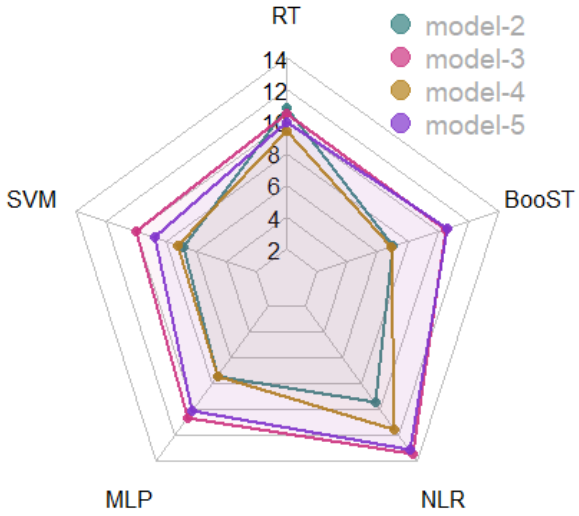
The plots from Figure 4.6a and 4.6b depicts the lower error observed by model–2 (SVM 2, MLP 2 and BooST 2) and model–4 (SVM 4, MLP 4 and BooST 4). Figure 4.6c also supports the previous figures. However, the overall estimate of the techniques in selecting models of the second set suggests that model–2 and model–4 are quite significant in predicting CCS as there was no duplication of explanatory variables. With approximations in values, BooST, MLP and SVM still display sufficiency as techniques for estimating CCS.

The third set of models were analyzed to assess the sensitivity of concrete in predicting compressive strength as well as verify the combined effect of the non-dimensional explanatory variables in predicting CCS as described in Table 3.7 . However, there appear duplication of explanatory variables in original form and non-dimensional form, which is evident from the results in Table 4.3 - 4.7 showing no much variation in values between model–6 to model–10. Figure 4.7 shows the radar chart for the third set of models.

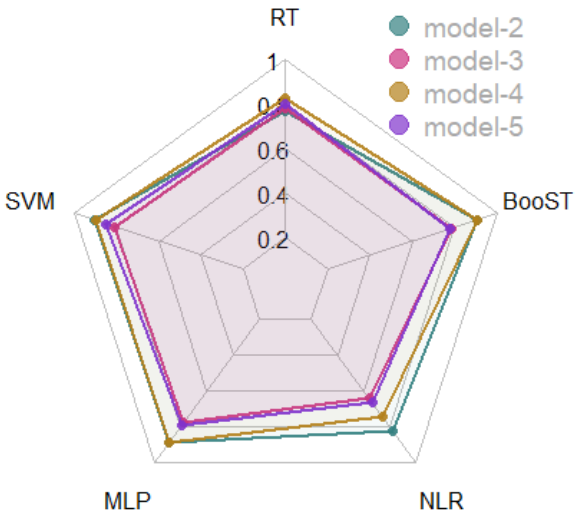
With no duplication of explanatory variables in the first and second sets of models, the model from the first set can be selected as the performing model for CCS prediction. Additional display of prediction of CCS against the individual test data instances for each technique is shown in Figure 4.8. The plots shown in Figure 4.8 depicts the best data fit from the predictive techniques, which compare the actual laboratory experimental results with the results from the predictive models. The vertical axis represents the concrete compressive strength in MPa of the measured samples, while the horizontal axis represents the index of instances in the test dataset used in predictive analysis. The figure shows that SVM, MLP and BooST models were emphasized good fit to the source data as compared to NLR and RT models which performed poorly. The SVM, MLP and BooST models, obtained reasonable results for most instances of the test data, except for few estimated instances with perhaps larger deviations (i.e., the worst estimated instances can be



(a) Mean absolute error (MAE).

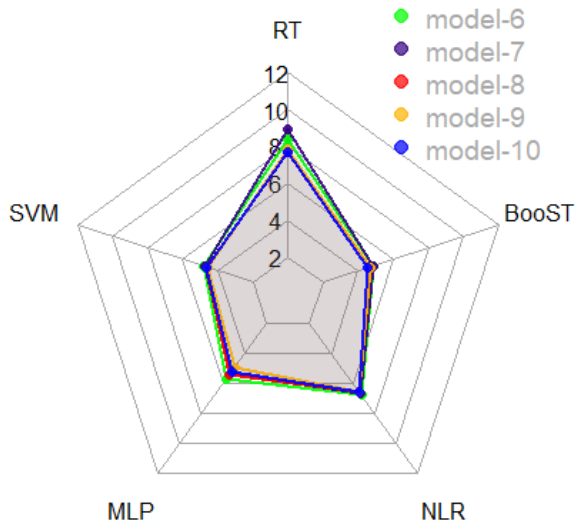


(b) Root mean square error (RMSE).

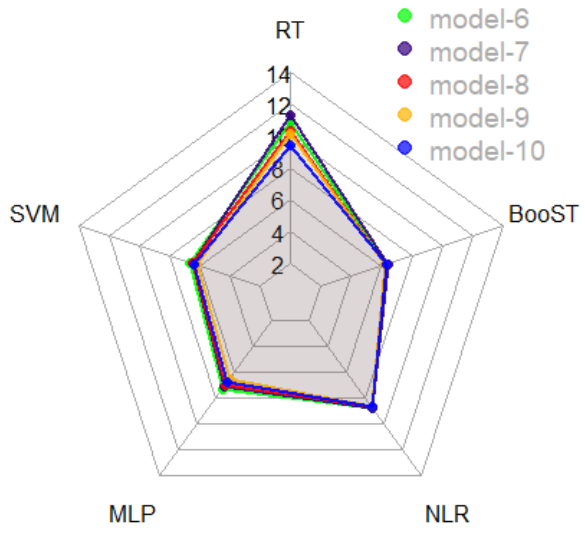


(c) Coefficient of determination (R^2).

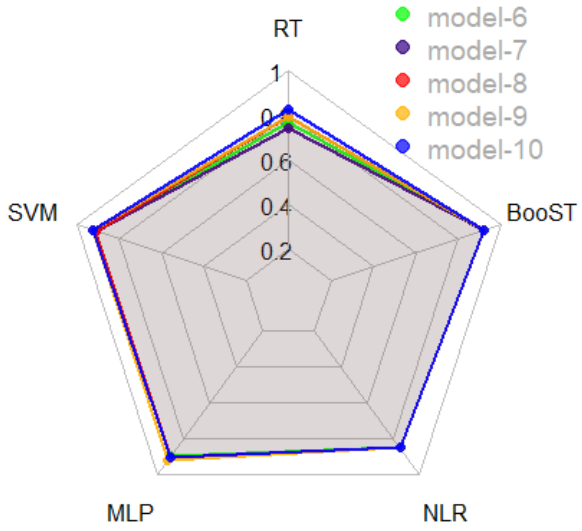
Figure 4.6. Radar chart of performance statistics for second set of models.



(a) Mean absolute error (MAE).



(b) Root mean square error (RMSE).



(c) Coefficient of determination (R^2).

Figure 4.7. Radar chart of performance statistics for third set of models.

seen above 150th instances as shown by SVM, MLP and BooST models). Additionally comparative plots of the actual CCS against predicted CCS for all models (model–1 to model–10) are provided in the Appendix. As shown in Figures A.1 to A.10, the more the observed data points congregates in a straight line about the diagonal, the better the performance is judged.

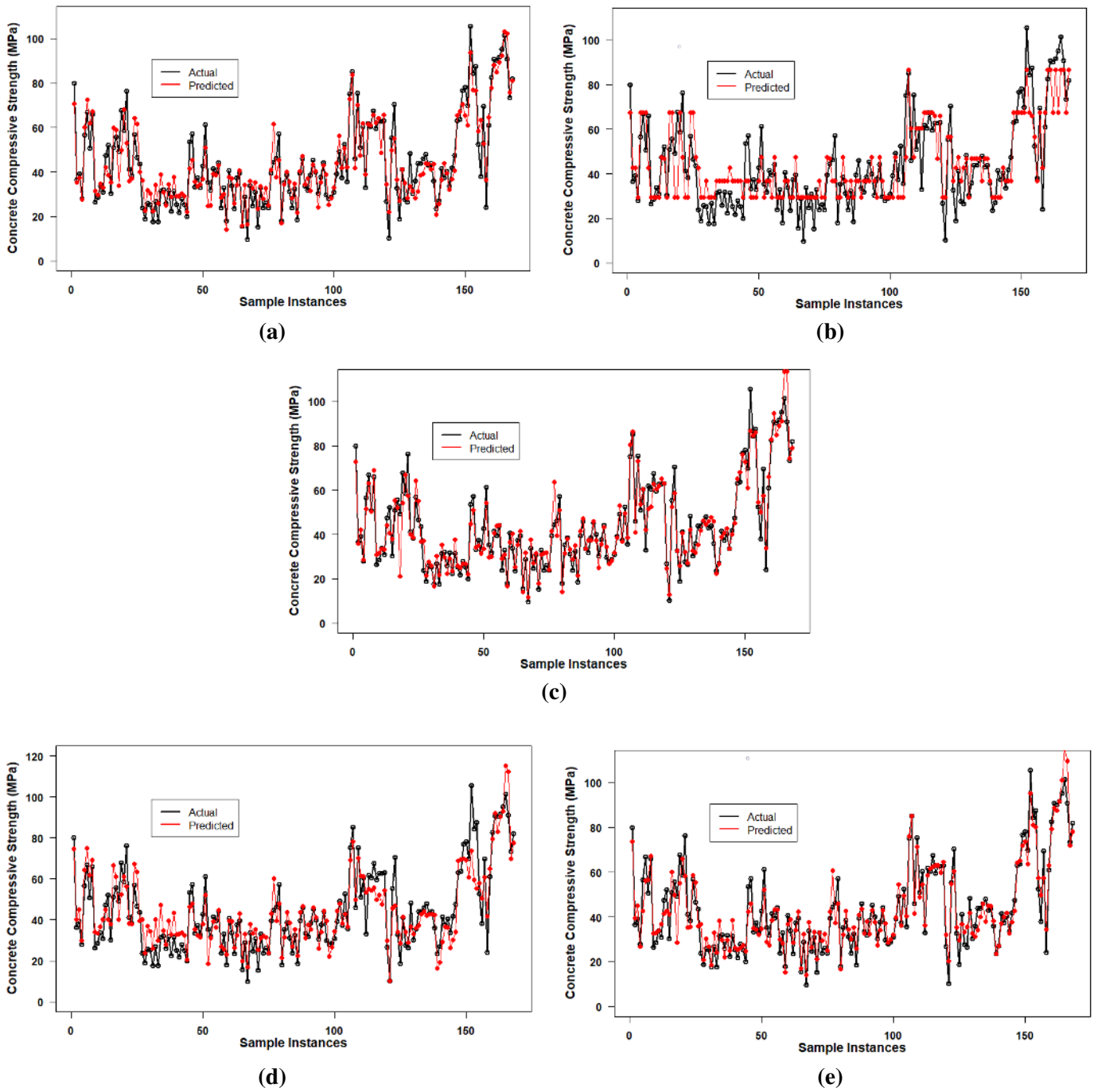


Figure 4.8. Actual and predicted plots for the compressive strength of concrete using each predictive technique **(a)** MLP 1 **(b)** RT 1 **(c)** SVM 1 **(d)** NLR 1 **(e)** BooST 1.

Chapter 5

Conclusions

5.1 Summary

In this thesis, statistical and machine learning techniques were studied, and respective approaches proposed to address the challenges existing in the concrete domain for the mixture design of concrete experiments and the prediction for concrete compressive strength of high performance.

The existing techniques for the stated subject matters were outlined and discussed in Chapter 2 of this thesis. Firstly, the concrete mixture design methods discussed in Section 2.1 of Chapter 2 can still provide reasonable optimal compressive strength. However, with expensive concrete constituents used for producing concrete of high performance, excessive trial experiments would lead to expensive concrete production for a specified strength. Therefore, in Section 3.1 of Chapter 3 Response Surface Method (**RSM**) design (**BBD**) of experiment have been proposed to address such constraint.

Secondly, challenges about adequate estimation of concrete compressive strength (**CCS**) with consideration to constituents and mixture proportioning have been investigated with machine learning techniques. Section 2.2 of Chapter 2 presented the commonly used techniques in predicting **CCS** whereas a new technique proposed in Section 3.2 of Chapter 3 was compared with the existing techniques in attempt to investigate the estimation of **CCS** performance of large experimental boundaries.

Considering the increase in technological advancements in construction and civil engineering, increasing knowledge for improved performance of concrete properties can reduce the overall time spent in the concrete laboratory for testings and alternatively avoid increased concrete production costs.

5.2 Contributions

The studies carried out in this thesis, promise significant contributions to the body of knowledge in the concrete domain. The following phenomena were observed:

Concrete Mixture Design for Optimal Compressive Strength

In this study, a design (**BBD**) of Response Surface Methodology (RSM) has been used to optimize the 28 days compressive strength of high-performance concrete with mix design comprising water to cementitious content (W/C_{mt}) , fine to total aggregate (FA/T_A) and total cementitious content (Q_{CmT}). However, though the HPC is a complex material with properties that pose nonlinear relationship with its constituents, as well as difficulty in modeling its behaviour, the importance in using Box-Behnken Design (BBD) of RSM appears successful in optimizing the HPC compressive strength.

The following conclusions are drawn from this study:

- From the dataset, within an experimental boundary of factors W/C_{mt} (0.3 to 0.45), Q_{CmT} (460 to 500)$kg/m^3$and FA/T_A (0.4 to 0.5), the optimal combination of 0.308 for W/C_{mt}, 493.44kg/m^3 for Q_{CmT} and 0.492 for FA/T_A results to a maximum of 73.31 MPa for 28 days CCS with W/C_{mt} as the major influencing factor. Such findings which requires less number of experimental runs/tests portray RSM to be a significant tool for the optimization of compressive strength of concrete where constraints are observed in experimental boundaries.

- The use of RSM allows for acquiring more information with minimum concrete trial mixture and experimental runs. Additionally, it becomes possible to investigate the interaction effect of constituent factors on the CCS.

- Additionally, from the design in this experimental study where the lack of fit from the ANOVA table appears not significant, this infers that the prediction made from the RSM quadratic model can be considered just as accurate as running additional concrete test experiments within the experiment factor-boundaries, as long as no lurking variables change before additional experiments is run. This portrays the potential of the RSM model as an adequate technique for also predictive studies in the concrete domain.

Prediction for Concrete Compressive Strength of High Performance

This study has examined the use of machine learning techniques to estimate concrete compressive strength (CCS) of high-performance concrete (HPC) as a function of its constituents and mixture proportions, to improve the practice of predicting concrete strength by computational analysis. Concrete data of 28 days with a wide range of explanatory variables were obtained from the literature and unique models created for estimating CCS. A new predictive technique called boosting smooth transition regression trees (BooST) was employed in comparison with other state-of-the-art techniques in predicting CCS. The following conclusions can be deduced from this study:

- With the growing demand for high strength and high-performance concrete, compressive strength evaluation becomes increasingly important. As the first research attempt in using BooST to estimate the concrete compressive strength (CCS) at 28 days, BooST showed improvement in CCS performance prediction from constituent composition and proportioning, which is one of the main contributions of this study. BooST has shown dominance in reduced errors and better model fit over other techniques. This shows BooST can be used as a competent technique in predicting the compressive strength of concrete having a wide range of explanatory variables. Additionally, this suggests the competency of tree-based ensembles in accurate predictions in the concrete domain.

- With unique and no duplication of concrete mixture constituents, training BooST model with the original explanatory variables provides a better prediction of compressive strength at 28 days. This result is shown in Table 4.3

- 4.7 between the first and second sets of models in Table 3.5 and Table 3.6 respectively.

- Furthermore, to assess the sensitivity of concrete and verify the combined effects of the non-dimensional explanatory variables in predicting CCS, the third set of models shows that duplication of explanatory variables in both original form and non-dimensional form contribute not much variation in values as shown between model-6 to model-10.

  Model-3 and model-5 produced the worst predictive results since they contain a limited number of explanatory variables used in predicting CCS. This affirms that concrete is sensitive to the number of explanatory variables used in estimating its compressive strength.

- The BooST, MLP and SVM out-performed NLR and RT, indicating their capability in predictive analysis for concrete data having increased deviation due to the range of values of explanatory (input) values. This addresses the predictive performance of concrete compressive strength in an expanded range of mixture proportion and constituents. Additionally, the findings in these studies yielded can assist engineers and practitioners in identifying critical compositions related to concrete compressive strength performance.

5.3 Future Work

This thesis shows that statistical approaches and machine learning techniques are beneficial, which provides promising modus operandi for concrete compressive strength performance assessments. However, there are also some limitations to these studies. The challenges and the potential future studies for the researches in this thesis are discussed and listed below:

Design of Concrete Experiment for Optimal Compressive Strength

- In using the RSM method, it is difficult to estimate the accuracy of the approximation. That is to say, the magnitude of the approximation errors is difficult since we have interactive concrete factors responsible for approximating the response (CCS). Contributing to this challenge may be due to

the variation in individual constituent and compositions. Hence in future research, the impacts of interactive factors should be studied. Additionally, attention should be given to other conditions that could affect the CCS such as individual properties of concrete constituents, curing procedures as well as working environment/condition of concrete.

- While a range of analysis is executed, which requires being statistically savvy when using RSM for concrete mixture design, a limitation of the method is that it is a local analysis. That is to say, the developed response surface is invalid for regions other than the studied ranges of factors (experimental boundaries) for the mixture design. However, conducting additional experimental tests for regions of interest for concrete mixture design could provide additional validation to the accuracy of the method.

- While it is vital to experiment within high boundaries of concrete of high-performance, it is imperative to explore beyond the selected boundaries of experiments as well as a more significant number of factors/variables used in the mixture design. Such a study may contribute to a higher number of experimental runs; however, it would articulate a proper statistical measure for concrete mixture design.

Prediction for Compressive Strength of Concrete for High Performance

- Since no single experiment on concrete has been recorded in the literature to contain a vast variation in several experimental boundaries, the data used in the modeling was not collected from a single source. Therefore, the database may contain unexpected inaccuracies. For instance, the class of fly ash from some sources were sometimes not reported, and other admixtures may be from different manufacturers, different chemical compositions with little details of absolute values of solid contents. Hence, future study should focus on creating a standard data collection strategy for concrete of high strength and performance.

- The costs as well as environmental impacts of concrete constituents production happens to be an important factor that affects the environment. With

current studies in the literature on climate change, it would be significant to estimate the CCS relative to the production costs of each constituents. Such study would be a significant further work for CCS estimation.

- With the advent of new concrete admixtures for producing high strength and high-performance concrete, future studies can also gather experts (e.g., admixture manufacturers, engineers) opinions on the chemical compositions as well as their quantitative effect on concrete production. Developing models that include newer or several admixtures would allow for a comprehensive assessment of concrete performance. In other words, assessing the compressive strength performance from the composition of concrete with more admixture may be complex but could provide an additional understanding that validates the true non-linear relationship.

- The models that contained additional non-dimensional explanatory variables may be challenging to implement when assessing the true concrete compressive strength performance since there may be duplication of concrete constituents. Therefore, a pilot project can be conducted in a future study to test the feasibility of the proposed models for practical use. Such endevour has potentials to bridge the gap between computational models for concrete and the current practical application.

Bibliography

[1] Y. Fonseca, M. Medeiros, G. Vasconcelos, and A. Veiga, "BooST: Boosting Smooth Trees for Partial Effect Estimation in Nonlinear Regressions," *arXiv:1808.03698v2*, Aug. 2018. [Online]. Available: http://arxiv.org/abs/1808.03698 → pages ix, 44, 45, 46, 47

[2] P. K. Mehta and P.-C. Aïtcin, "Principles underlying production of high-performance concrete," *Cement, concrete and aggregates*, vol. 12, no. 2, pp. 70–78, 1990. → pages x, 10, 11, 12, 13

[3] Y. Edamatsu, N. Nishida, and M. Ouchi, "A rational mix-design method for self-compacting concrete considering interaction between coarse aggregate and mortar particles," in *Proceedings of the first international RILEM symposium on self-compacting concrete, Stockholm. Sweden*, 1999, pp. 309–320. → pages x, 12, 14, 15

[4] C. Hwang and C. Tsai, "The effect of aggregate packing types on engineering properties of self-consolidating concrete," *SCC*, 2005. → pages x, 16

[5] N. Su and B. Miao, "A new method for the mix design of medium strength flowing concrete with low cement content," *Cement and Concrete Composites*, vol. 25, no. 2, pp. 215–222, 2003. → pages x, 12, 16, 17

[6] P. Dinakar and S. Manu, "Concrete mix design for high strength self-compacting concrete using metakaolin," *Materials & Design*, vol. 60, pp. 661–668, 2014. → pages x, 12, 20, 21

[7] A. Habibi and J. Ghomashi, "Development of an optimum mix design method for self-compacting concrete based on experimental results," *Construction and Building Materials*, vol. 168, pp. 113–123, 2018. → pages x, 20, 22

[8] G. F. Kheder and R. S. Al Jadiri, "New method for proportioning self-consolidating concrete based on compressive strength requirements." *ACI Materials Journal*, vol. 107, no. 5, 2010. → pages x, 12, 22, 23

[9] C. R. Gagg, "Cement and concrete as an engineering material: an historic appraisal and case study analysis," *Engineering Failure Analysis*, vol. 40, pp. 114–140, 2014. → page 1

[10] M. Nehdi and A. M. Soliman, "Early-age properties of concrete: overview of fundamental concepts and state-of-the-art research," *Proceedings of the Institution of Civil Engineers-Construction Materials*, vol. 164, no. 2, pp. 57–77, 2011. → page 1

[11] Z. Li, *Advanced concrete technology*. John Wiley & Sons, 2011. → page 1

[12] I. Kett, *Engineered concrete: mix design and test methods*. CRC Press, 2009. → page 2

[13] C. Poon, Z. Shui, L. Lam, H. Fok, and S. Kou, "Influence of moisture states of natural and recycled aggregates on the slump and compressive strength of concrete," *Cement and concrete research*, vol. 34, no. 1, pp. 31–36, 2004. → page 3

[14] A. T. A. Dantas, M. B. Leite, and K. de Jesus Nagahama, "Prediction of compressive strength of concrete containing construction and demolition waste using artificial neural networks," *Construction and Building Materials*, vol. 38, pp. 717–722, 2013. → page 3

[15] H. G. Russell, "Aci defines high-performance concrete," *Concrete international*, vol. 21, no. 2, pp. 56–57, 1999. → page 3

[16] I.-C. Yeh, "Modeling of strength of high-performance concrete using artificial neural networks," *Cement and Concrete research*, vol. 28, no. 12, pp. 1797–1808, 1998. → pages 3, 29, 41, 46, 48, 61, 62

[17] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006. → page 3

[18] Z. Salari, B. Vakhshouri, and S. Nejadi, "Analytical review of the mix design of fiber reinforced high strength self-compacting concrete," *Journal of Building Engineering*, vol. 20, pp. 264–276, 2018. → page 9

[19] F. Deng and H. Smyth, "Contingency-based approach to firm performance in construction: Critical review of empirical research," *Journal of Construction Engineering and Management*, vol. 139, no. 10, p. 04013004, 2013.

[20] R. Silva, J. De Brito, and R. Dhir, "Prediction of the shrinkage behavior of recycled aggregate concrete: A review," *Construction and Building Materials*, vol. 77, pp. 327–339, 2015.

[21] X. X. He and Q. Wang, "Literature review: Mechanical properties of hardened silica fume concrete," in *Advanced Materials Research*, vol. 1008. Trans Tech Publ, 2014, pp. 1357–1362. → page 9

[22] P. Doorn, "Engineering village adds plumx metrics — elsevier engineering village," *Elsevier Engineering Village*, pp. Retrieved March 24, 2018, from https://blog.engineeringvillage.com/content/engineering–village–adds–plumx–metrics, 2018. → page 10

[23] T. T. Le, S. A. Austin, S. Lim, R. A. Buswell, A. G. Gibb, and T. Thorpe, "Mix design and fresh properties for high-performance printing concrete," *Materials and structures*, vol. 45, no. 8, pp. 1221–1232, 2012. → page 10

[24] Y. Malier, *High Performance Concrete: From material to structure*, ser. Modern Concrete Technology. Taylor & Francis, 1992. [Online]. Available: https://books.google.ca/books?id=PKvlI0dRUO0C → page 10

[25] P.-C. Aïtcin, *High performance concrete*. CRC press, 1998. → page 10

[26] M. Barbuta, R.-M. Diaconescu, and M. Harja, "Using neural networks for prediction of properties of polymer concrete with fly ash," *Journal of Materials in Civil Engineering*, vol. 24, no. 5, pp. 523–528, 2011. → pages 10, 28, 29

[27] C. Shi, "Effect of different mineral powders on properties of fresh and hardened self-consolidating concrete," *Special Publication*, vol. 233, pp. 65–76, 2006.

[28] G. Long, D. Feys, K. H. Khayat, and A. Yahia, "Efficiency of waste tire rubber aggregate on the rheological properties and compressive strength of cementitious materials," *Journal of Sustainable Cement-Based Materials*, vol. 3, no. 3-4, pp. 201–211, 2014.

[29] P. Shafigh, M. Z. Jumaat, and H. Mahmud, "Mix design and mechanical properties of oil palm shell lightweight aggregate concrete: a review," *International journal of the physical sciences*, vol. 5, no. 14, pp. 2127–2134, 2010. → page 10

[30] T.-P. Chang, F.-C. Chuang, and H.-C. Lin, "A mix proportioning methodology for high-performance concrete," *Journal of the Chinese Institute of Engineers*, vol. 19, no. 6, pp. 645–655, 1996. → page 10

[31] P. Kumbhar and P. Murnal, "A new mix design method for high performance concrete under tropical conditions," 2014. → page 10

[32] A. Standard, "211.1, standard practice for selecting proportions for normal, heavyweight, and mass concrete," *ACI Manual of Concrete Practice, Part*, vol. 1, pp. 211–1, 1996. → pages 10, 11

[33] R. N. Swamy, "High performance and durability through design," *Special Publication*, vol. 159, pp. 209–230, 1996. → page 10

[34] R. Swamy, "Design for durability and strength through the use of fly ash and slag in concrete," *Special Publication*, vol. 171, pp. 1–72, 1997. → page 10

[35] B. Bharatkumar, R. Narayanan, B. Raghuprasad, and D. Ramachandramurthy, "Mix proportioning of high performance concrete," *Cement and Concrete Composites*, vol. 23, no. 1, pp. 71–80, 2001. → page 11

[36] F. de Larrard, "A method for proportioning high-strength concrete mixtures," *Cement, concrete and aggregates*, vol. 12, no. 1, pp. 47–52, 1990. → pages 12, 14

[37] H. Okamura and K. Ozawa, "Mix design for self-compacting concrete," *Concrete library of JSCE*, vol. 25, no. 6, pp. 107–120, 1995. → pages 12, 14

[38] P.-K. Chang, "An approach to optimizing mix design for properties of high-performance concrete," *Cement and Concrete Research*, vol. 34, no. 4, pp. 623–629, 2004. → pages 12, 15

[39] F. De Larrard and T. Sedran, "Mixture-proportioning of high-performance concrete," *Cement and concrete research*, vol. 32, no. 11, pp. 1699–1704, 2002. → pages 12, 15, 18

[40] P.-L. Ng, A. K.-H. Kwan, and L. G. Li, "Packing and film thickness theories for the mix design of high-performance concrete," *Journal of Zhejiang University-SCIENCE A*, vol. 17, no. 10, pp. 759–781, 2016. → pages 12, 18

[41] M. Simon *et al.*, "Concrete mixture optimization using statistical methods," United States. Federal Highway Administration. Office of Infrastructure , Tech. Rep., 2003. → pages 12, 23

[42] M. Kharazi, L. M. Lye, and A. Hussein, "Designing and optimizing of concrete mix proportion using statistical mixture design methodology," in *CSCE 2013 General Conference*, vol. GEN-71. Canadian Society of Civil Engineering, 2013, pp. 1 – 10. → pages 12, 23

[43] T. Bouziani, "Assessment of fresh properties and compressive strength of self-compacting concrete made with different sand types by mixture design modelling approach," *Construction and Building Materials*, vol. 49, pp. 308–314, 2013. → pages 12, 24

[44] H. Okamura, "Self-compacting high-performance concrete," *Concrete international*, vol. 19, no. 7, pp. 50–54, 1997. → page 14

[45] Y. Edamatsu, T. Sugamata, and M. Ouchi, "A mix-design method for self-compacting concrete based on mortar flow and funnel tests," in *PRO 33: 3rd International RILEM Symposium on Self-Compacting Concrete*, vol. 33. RILEM Publications, 2003, p. 345. → page 14

[46] P. Domone and M. Soutsos, "Approach to the proportioning of high-strength concrete mixes," *Concrete International*, vol. 16, no. 10, pp. 26–31, 1994. → page 15

[47] F. De Larrard, *Concrete mixture proportioning: a scientific approach*. CRC Press, 2014. → page 15

[48] O. Koutnỳ, J. Kratochvíl, J. Švec, and J. Bednárek, "Modelling of packing density for particle composites design," *Procedia Engineering*, vol. 151, pp. 198–205, 2016. → page 15

[49] L. Li and A. Kwan, "Packing density of concrete mix under dry and wet conditions," *Powder technology*, vol. 253, pp. 514–521, 2014. → page 15

[50] C.-L. Hwang, J. Liu, L. Lee, and F. Lin, "Densified mixture design algorithm and early properties of high performance concrete," *Journal of*

the Chinese Institute of Civil and Hydraulic Engineering, vol. 8, no. 2, pp. 217–219, 1996. → page 15

[51] F. de Larrard and T. Sedran, "Optimization of ultra-high-performance concrete by the use of a packing model," *Cement and Concrete Research*, vol. 24, no. 6, pp. 997–1009, 1994. → page 18

[52] G. K. Babu and P. Dinakar, "Strength efficiency of metakaolin in concrete," *Structural concrete*, vol. 7, no. 1, pp. 27–31, 2006. → page 20

[53] DIN1045, "Beton und stahlbeton," *Beton Verlag GMBH. Köln*, 1988. → page 20

[54] EFNARC, "Specification and guidelines for self-compacting concrete, european federation of producers and applicators of specialist products for structures," 2002. → pages 20, 22

[55] ACI Committee 211, *Standard Building Code Requirements for Reinforced Concrete (ACI 211.1-91)*, Farmington Hills MI American Concrete Institute Std. ACI standard 211.1-91, 1991. → page 22

[56] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons, 2016. → pages 23, 24, 36, 37, 38, 40, 41, 42, 62, 63, 64

[57] K. Ellis, R. Silvestrini, B. Varela, N. Alharbi, and R. Hailstone, "Modeling setting time and compressive strength in sodium carbonate activated blast furnace slag mortars using statistical mixture design," *Cement and Concrete Composites*, vol. 74, pp. 1–6, 2016. → page 24

[58] B. A. Young, A. Hall, L. Pilon, P. Gupta, and G. Sant, "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*, vol. 115, pp. 379–388, 2019. → page 26

[59] M. I. Khan, "Predicting properties of high performance concrete containing composite cementitious materials using artificial neural networks," *Automation in Construction*, vol. 22, pp. 516–524, 2012. → pages 26, 28

[60] Z.-H. Duan, S.-C. Kou, and C.-S. Poon, "Prediction of compressive strength of recycled aggregate concrete using artificial neural networks,"

Construction and Building Materials, vol. 40, pp. 1200–1206, 2013. → page 26

[61] H. Wang, C. Qi, H. Farzam, and J. Turici, "Interaction of materials used in concrete," *Concrete International*, vol. 28, no. 4, pp. 47–52, 2006. → page 27

[62] U. Anyaoha, X. Peng, and Z. Liu, "Concrete performance prediction using boosting smooth transition regression trees (boost)," in *Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XIII*, vol. 10971. International Society for Optics and Photonics, 2019, p. 109710I. → page 27

[63] P. Ramadoss, "Modeling for the evaluation of strength and toughness of high-performance fiber reinforced concrete," *Journal of Engineering Science and Technology*, vol. 7, no. 3, pp. 280–291, 2012. → page 27

[64] S. Subaşı, A. Beycioğlu, E. Sancak, and İ. Şahin, "Rule-based mamdani type fuzzy logic model for the prediction of compressive strength of silica fume included concrete using non-destructive test results," *Neural Computing and Applications*, vol. 22, no. 6, pp. 1133–1139, 2013.

[65] A. Khashman and P. Akpinar, "Non-destructive prediction of concrete compressive strength using neural networks," *Procedia Computer Science*, vol. 108, pp. 2358–2362, 2017. → pages 27, 28

[66] S. Bhanja and B. Sengupta, "Investigations on the compressive strength of silica fume concrete using statistical methods," *Cement and Concrete Research*, vol. 32, no. 9, pp. 1391–1394, 2002. → page 27

[67] M. F. M. Zain and S. M. Abd, "Multiple regression model for compressive strength prediction of high performance concrete," *Journal of applied sciences*, vol. 9, no. 1, pp. 155–160, 2009. → page 27

[68] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees. wadsworth int," *Group*, vol. 37, no. 15, pp. 237–251, 1984. → pages 27, 30

[69] J.-S. Chou and A.-D. Pham, "Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength," *Construction and Building Materials*, vol. 49, pp. 554–563, 2013. → page 27

[70] P. Chopra, R. K. Sharma, and M. Kumar, "Prediction of compressive strength of concrete using artificial neural network and genetic programming," *Advances in Materials Science and Engineering*, vol. 2016, 2016. → page 28

[71] F. Khademi and S. M. Jamal, "Predicting the 28 days compressive strength of concrete using artificial neural network," *i-managers Journal on Civil Engineering*, vol. 6, no. 2, 2016.

[72] J. Noorzaei, S. Hakim, M. Jaafar, and W. Thanoon, "Development of artificial neural networks for predicting concrete compressive strength," *International Journal of engineering and Technology*, vol. 4, no. 2, pp. 141–153, 2007. → page 28

[73] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of microbiological methods*, vol. 43, no. 1, pp. 3–31, 2000. → pages 28, 29, 30

[74] N. Deshpande, S. Londhe, and S. Kulkarni, "Modeling compressive strength of recycled aggregate concrete by artificial neural network, model tree and non-linear regression," *International Journal of Sustainable Built Environment*, vol. 3, no. 2, pp. 187–198, 2014. → page 28

[75] S. Gupta, "Support vector machines based modelling of concrete strength," *World Academy of Science, Engineering and Technology*, vol. 36, pp. 305–311, 2007. → page 28

[76] K. Yan and C. Shi, "Prediction of elastic modulus of normal and high strength concrete by support vector machine," *Construction and building materials*, vol. 24, no. 8, pp. 1479–1485, 2010. → page 28

[77] H. Chore and M. Joshi, "Multiple regression models for prediction of compressive strength of concrete comprising industrial waste products," *Indian Concrete Journal*, vol. 89, pp. 33–46, Sep. 2015. → page 29

[78] H. Liang and W. Song, "Improved estimation in multiple linear regression models with measurement error and general constraint," *Journal of multivariate analysis*, vol. 100, no. 4, pp. 726–741, 2009. → page 29

[79] J. Sobhani, M. Najimi, A. R. Pourkhorshidi, and T. Parhizkar, "Prediction of the compressive strength of no-slump concrete: A comparative study of regression, neural network and anfis models," *Construction and Building Materials*, vol. 24, no. 5, pp. 709–718, 2010. → pages 29, 46, 48, 52

[80] V. Vapnik, S. E. Golowich, and A. J. Smola, "Support vector method for function approximation, regression estimation and signal processing," in *Advances in neural information processing systems*, 1997, pp. 281–287. → page 32

[81] S.-H. An, U.-Y. Park, K.-I. Kang, M.-Y. Cho, and H.-H. Cho, "Application of support vector machines in assessing conceptual cost estimates," *Journal of Computing in Civil Engineering*, vol. 21, no. 4, pp. 259–264, 2007. → page 32

[82] A. Çevik, A. E. Kurtoğlu, M. Bilgehan, M. E. Gülşan, and H. M. Albegmprli, "Support vector machines in structural engineering: a review," *Journal of Civil Engineering and Management*, vol. 21, no. 3, pp. 261–281, 2015. → page 32

[83] V. Vapnik, "Statistical learning theory wiley-interscience," *New York*, 1998. → page 32

[84] D. H. Wolpert, W. G. Macready *et al.*, "No free lunch theorems for optimization," *IEEE transactions on evolutionary computation*, vol. 1, no. 1, pp. 67–82, 1997. → page 33

[85] H. Güllü and H. İ. Fedakar, "Response surface methodology for optimization of stabilizer dosage rates of marginal sand stabilized with sludge ash and fiber based on ucs performances," *KSCE Journal of Civil Engineering*, vol. 21, no. 5, pp. 1717–1727, 2017. → page 36

[86] I.-C. Yeh, "Analysis of strength of concrete using design of experiments and neural networks," *Journal of Materials in Civil Engineering*, vol. 18, no. 4, pp. 597–604, 2006. → page 36

[87] S. Ahmad and S. A. Alghamdi, "A statistical approach to optimizing concrete mixture design," *The Scientific World Journal*, vol. 2014, 2014. → pages 36, 37, 62

[88] B. Lejano and J. Gagan, "Optimization of compressive strength of concrete with pig-hair fibers as fiber reinforcement and green mussel shells as partial cement substitute." *International Journal of GEOMATE*, vol. 12, pp. 37–44, 2017. → page 36

[89] M.Y. oordin, V.C. Venkatesh, S, Sharif, S. Elting and A. Abdullah, "Application of response surface methodology in describing the performance of coated carbide tools when turning aisi 1045 steel." *Journal*

of Materials Processing Technology, vol. 145, DOI 10.1016/S0924-0136(03)00861-6, pp. 46–58, 2004. → page 37

[90] N. B. Siraj, A. R. Fayek, and A. A. Tsehayae, "Development and optimization of artificial intelligence-based concrete compressive strength predictive models," *International Journal of Structural and Civil Engineering Research*, vol. 5, no. 3, pp. 156–167, 2016. → page 37

[91] ACI Committee E-701, Materials for Concrete Construction, "Cementitious materials for concrete," American Concrete Institute, 2001.

[92] W. Meng, M. Valipour, and K. H. Khayat, "Optimization and performance of cost-effective ultra-high performance concrete," *Materials and structures*, vol. 50, no. 1, p. 29, 2017. → page 62

[93] A. Neville and P.-C. Aitcin, "High performance concretean overview," *Materials and structures*, vol. 31, no. 2, pp. 111–117, 1998. → page 37

[94] I.-C. Yeh, "Generalization of strength versus water–cementitious ratio relationship to age," *Cement and Concrete Research*, vol. 36, no. 10, pp. 1865–1873, 2006.

[95] A. M. Neville *et al.*, *Properties of concrete*. longman London, 1995, vol. 4.

[96] B. Şimşek, Y. T. İç, and E. H. Şimşek, "A rsm-based multi-response optimization application for determining optimal mix proportions of standard ready-mixed concrete," *Arabian Journal for Science and Engineering*, vol. 41, no. 4, pp. 1435–1450, 2016.

[97] S. H. Kosmatka and M. L. Wilson, *Design and control of concrete mixtures*. Portland Cement Assoc., 2011. → pages 37, 62

[98] M. Lichman, "UCI Machine Learning Repository," *Irvine, CA: University of California, School of Information and Computer Science.*, p. [http://archive.ics.uci.edu/ml], 2013. → page 37

[99] G. E. Box and D. W. Behnken, "Some new three level designs for the study of quantitative variables," *Technometrics*, vol. 2, no. 4, pp. 455–475, 1960. → page 38

[100] R. V. Lenth *et al.*, "Response-surface methods in r, using rsm," *Journal of Statistical Software*, vol. 32, no. 7, pp. 1–17, 2009. → pages 38, 41, 65

[101] J. Lawson, *Design and Analysis of Experiments with R*. Chapman and Hall/CRC, 2014. → pages 38, 40, 61, 65

[102] J. J. Borkowski, "Spherical prediction-variance properties of central composite and boxbehnken designs," *Technometrics*, vol. 37, no. 4, pp. 399–410, 1995. → pages 39, 65

[103] J. C. Da Rosa, A. Veiga, and M. C. Medeiros, "Tree-structured smooth transition regression models," *Computational Statistics & Data Analysis*, vol. 52, no. 5, pp. 2469–2488, 2008. → pages 44, 46

[104] W. Dai, T. Tong, and M. G. Genton, "Optimal estimation of derivatives in nonparametric regression," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 5700–5724, 2016. → page 45

[105] P. L. Bühlmann, "Consistency for $l_2$boosting and matching pursuit with trees and tree-type basis functions," in *Research report/Seminar für Statistik, Eidgenössische Technische Hochschule (ETH)*, vol. 109. Seminar für Statistik, Eidgenössische Technische Hochschule (ETH), 2002. → page 45

[106] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001. → page 45

[107] T. Zhang, B. Yu *et al.*, "Boosting with early stopping: Convergence and consistency," *The Annals of Statistics*, vol. 33, no. 4, pp. 1538–1579, 2005. → page 46

[108] I. Yeh *et al.*, "Modeling slump of concrete with fly ash and superplasticizer," *Computers and Concrete*, vol. 5, no. 6, pp. 559–572, 2008. → pages 46, 48

[109] C. Videla and C. Gaedicke, "Modeling portland blast-furnace slag cement high-performance concrete," *Materials Journal*, vol. 101, no. 5, pp. 365–375, 2004. → page 48

[110] J.-K. Kim, S. H. Han, Y. D. Park, and J. H. Noh, "Material properties of self-flowing concrete," *Journal of Materials in Civil Engineering*, vol. 10, no. 4, pp. 244–249, 1998. → page 48

[111] L. Lam, Y. Wong, and C. S. Poon, "Effect of fly ash and silica fume on compressive and fracture behaviors of concrete," *Cement and Concrete research*, vol. 28, no. 2, pp. 271–283, 1998. → page 48

[112] H. Y. Leung, J. Kim, A. Nadeem, J. Jaganathan, and M. Anwar, "Sorptivity of self-compacting concrete containing fly ash and silica fume," *Construction and Building Materials*, vol. 113, pp. 369–375, 2016. → page 48

[113] M. Liu, "Self-compacting concrete with different levels of pulverized fuel ash," *Construction and Building Materials*, vol. 24, no. 7, pp. 1245–1252, 2010. → page 48

[114] M. M. Johari, J. Brooks, S. Kabir, and P. Rivard, "Influence of supplementary cementitious materials on engineering properties of high strength concrete," *Construction and Building Materials*, vol. 25, no. 5, pp. 2639–2648, 2011. → page 48

[115] M. C. Nepomuceno, L. Pereira-de Oliveira, and S. M. R. Lopes, "Methodology for the mix design of self-compacting concrete using different mineral additions in binary blends of powders," *Construction and Building Materials*, vol. 64, pp. 82–94, 2014. → page 48

[116] A. Oner and S. Akyuz, "An experimental study on optimum usage of ggbs for the compressive strength of concrete," *Cement and Concrete Composites*, vol. 29, no. 6, pp. 505–514, 2007. → page 48

[117] N. Bouzoubaa and M. Lachemi, "Self-compacting concrete incorporating high volumes of class f fly ash: Preliminary results," *Cement and concrete research*, vol. 31, no. 3, pp. 413–420, 2001. → page 48

[118] A. Ghezal and K. H. Khayat, "Optimizing self-consolidating concrete with limestone filler by using statistical factorial design methods," *Materials Journal*, vol. 99, no. 3, pp. 264–272, 2002. → page 48

[119] V. K. Bui, Y. Akkaya, and S. P. Shah, "Rheological model for self-consolidating concrete," *Materials Journal*, vol. 99, no. 6, pp. 549–559, 2002. → page 48

[120] M. Sonebi, "Medium strength self-compacting concrete containing fly ash: Modelling using factorial experimental plans," *Cement and Concrete research*, vol. 34, no. 7, pp. 1199–1208, 2004. → page 48

[121] R. Patel, K. Hossain, M. Shehata, N. Bouzoubaa, and M. Lachemi, "Development of statistical models for mixture design of high-volume fly ash self-consolidating concrete," *Materials Journal*, vol. 101, no. 4, pp. 294–302, 2004. → page 48

[122] W. Wongkeo, P. Thongsanitgarn, A. Ngamjarurojana, and A. Chaipanich, "Compressive strength and chloride resistance of self-compacting concrete containing high level fly ash and silica fume," *Materials & Design*, vol. 64, pp. 261–269, 2014. → pages 46, 48

[123] ASTM C 595, "Standard specification for blended hydraulic cements," *Annual book of ASTM standards*, 2001. → page 48

[124] ASTM C 150, "Standard specification for portland cement," *Annual book of ASTM standards*, 2004. → page 48

[125] ASTM C 618-99, "Standard specification for coal fly ash and raw or calcined natural pozzolan for use as a mineral admixture in concrete," *Annual book of ASTM standards*, 1999. → page 48

[126] ASTM C 1240-99, "Standard specification for silica fume for use as a mineral admixture in hydraulic-cement concrete, mortar, and grout," *Annual book of ASTM standards*, 1999. → page 48

[127] British Standard EN 197-1., "Cement–part 1: Composition, specifications and conformity criteria for common cements," *British Standards Institution*, 2000. → page 48

[128] European Standard EN 197-1., "Cement–part 1: Composition, specifications and conformity criteria for common cements," *Comité Européen De Normalisation CEN, Brussels*, 2000. → page 48

[129] ASTM C 192, "Standard practice for making and curing concrete test specimens in the laboratory," *Annual book of ASTM standards*, 2000. → page 48

[130] ASTM C 989, "Standard specification for ground granulated blast-furnace slag for use in concrete and mortars," *Annual book of ASTM standards*, 1994. → page 48

[131] J. Kasperkiewicz, J. Racz, and A. Dubrawski, "Hpc strength prediction using artificial neural network," *Journal of Computing in Civil Engineering*, vol. 9, no. 4, pp. 279–284, 1995. → page 49

[132] R. Siddique, P. Aggarwal, and Y. Aggarwal, "Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks," *Advances in engineering software*, vol. 42, no. 10, pp. 780–786, 2011. → page 52

[133] D.-S. Yang, S.-K. Park, and J.-H. Lee, "A prediction on mix proportion factor and strength of concrete using neural network," *KSCE Journal of Civil Engineering*, vol. 7, no. 5, pp. 525–536, 2003.

[134] J. Namyong, Y. Sangchun, and C. Hongbum, "Prediction of compressive strength of in-situ concrete based on mixture proportions," *Journal of Asian Architecture and Building Engineering*, vol. 3, no. 1, pp. 9–16, 2004. → page 52

[135] T. Hegazy, P. Fazio, and O. Moselhi, "Developing practical neural network applications using back-propagation," *Computer-Aided Civil and Infrastructure Engineering*, vol. 9, no. 2, pp. 145–159, 1994. → page 54

[136] S. Arlot, A. Celisse *et al.*, "A survey of cross-validation procedures for model selection," *Statistics surveys*, vol. 4, pp. 40–79, 2010. → page 54

[137] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 569–575, 2009. → page 54

[138] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2. Montreal, Canada, 1995, pp. 1137–1145. → page 54

[139] E. Güneyisi, M. Gesoğlu, Z. Algın, and K. Mermerdaş, "Optimization of concrete mixture with hybrid blends of metakaolin and fly ash using response surface method," *Composites Part B: Engineering*, vol. 60, pp. 707–715, 2014. → page 60

[140] M. Zhang, V. Shim, G. Lu, and C. Chew, "Resistance of high-strength concrete to projectile impact," *International Journal of Impact Engineering*, vol. 31, no. 7, pp. 825–841, 2005. → page 62

[141] V. Živica, "Effects of the very low water/cement ratio," *Construction and building materials*, vol. 23, no. 12, pp. 3579–3582, 2009. → page 62

[142] D. C. Montgomery, *Design and analysis of experiments*. John wiley & sons, 2017. → pages 63, 65

Appendix A

Appendix

A.1 Figures

As shown in Figures A.1 to A.10, the more the observed data points congregates in a straight line about the diagonal, the better the performance is judged. The figures presented here further shows the predictive accuracy for the techniques as presented in Figure 4.5 to Figure 4.8. Figure A.1 visualizes the predictability of the first set of model where no duplication of explanatory variables is experienced. The graphical plot indicates that the observed data points of BooST congregates more in a straight line than the rest of the techniques.

Figure A.2 visualizes the predictability of the model–2 in the second set of models where transformed states of the explanatory variables in terms of mix ratio helps to understand the sensitivity of each technique in predicting CCS. The graphical plot indicates that the observed data points of the advanced techniques (BooST, SVM and MLP) congregates more in a straight line than NLR and RT techniques. The same can be observed in the rest of the models in the second set (model–3, model–4 and model–5) as depicted in Figure A.3, Figure A.4 and Figure A.5.

Lastly, the representation in the third set of models where duplication of explanatory variables in mix ratio form helps in further sensitivity analysis in estimating CCS can be depicted in Figure A.6 to Figure A.10. Figure A.6 depicts the predictability of the model–6 where the first duplication of explanatory variables in mix ratio form is observed in predicting CCS. The graphical plot indicates that

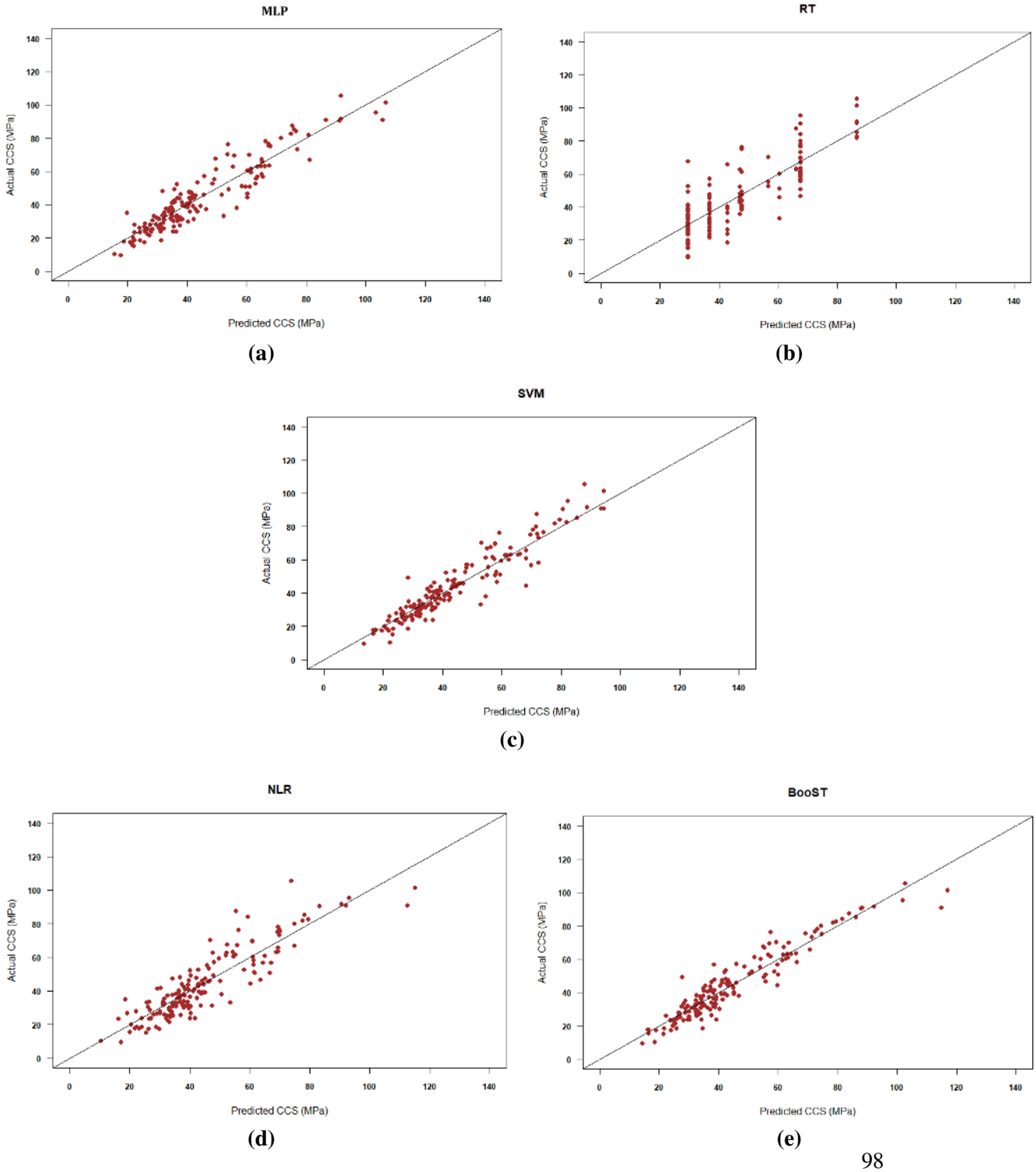


Figure A.1. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 1 **(b)** RT 1 **(c)** SVM 1 **(d)** NLR 1 **(e)** BooST 1.

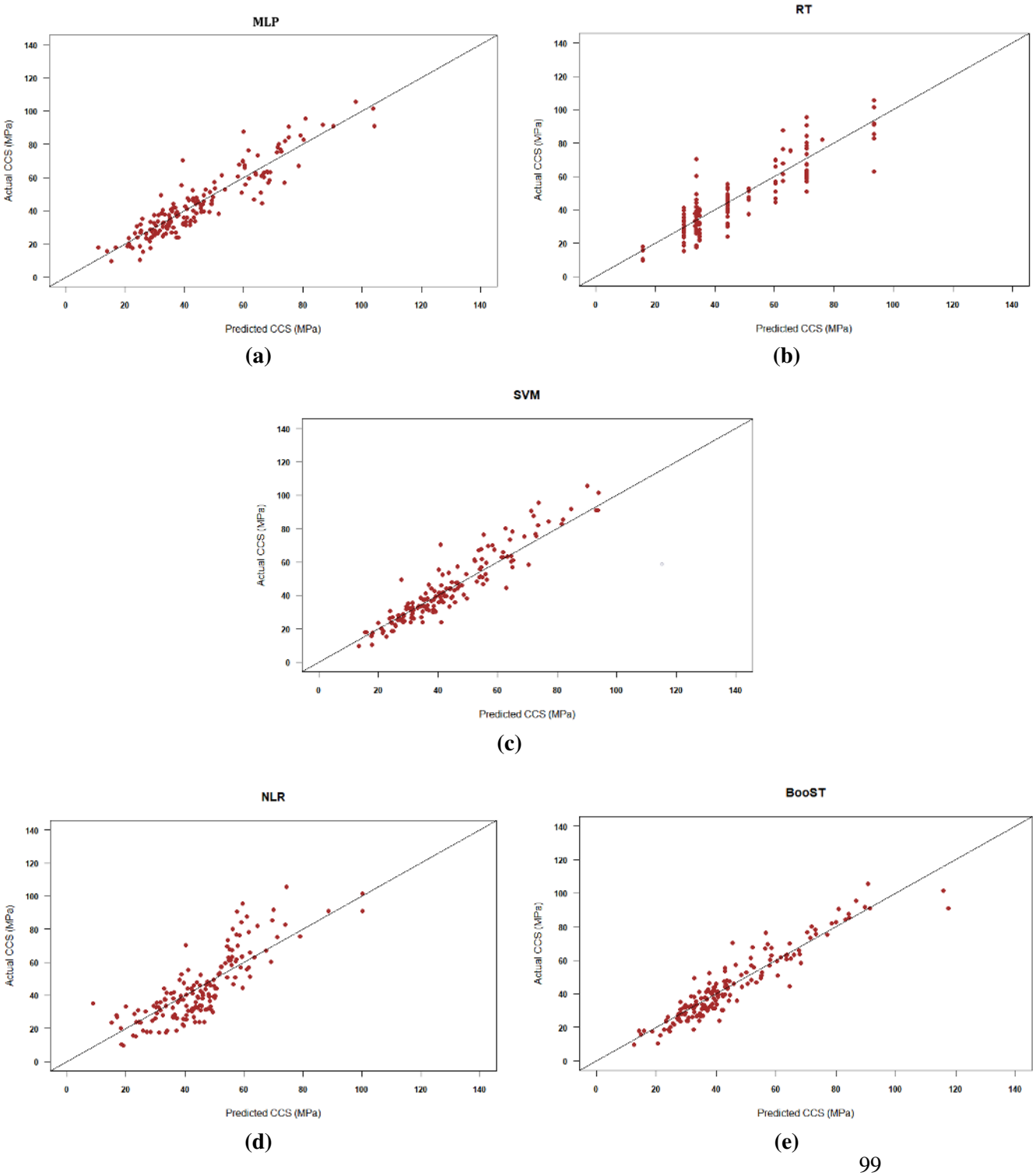


Figure A.2. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 2 **(b)** RT 2 **(c)** SVM 2 **(d)** NLR 2 **(e)** BooST 2.

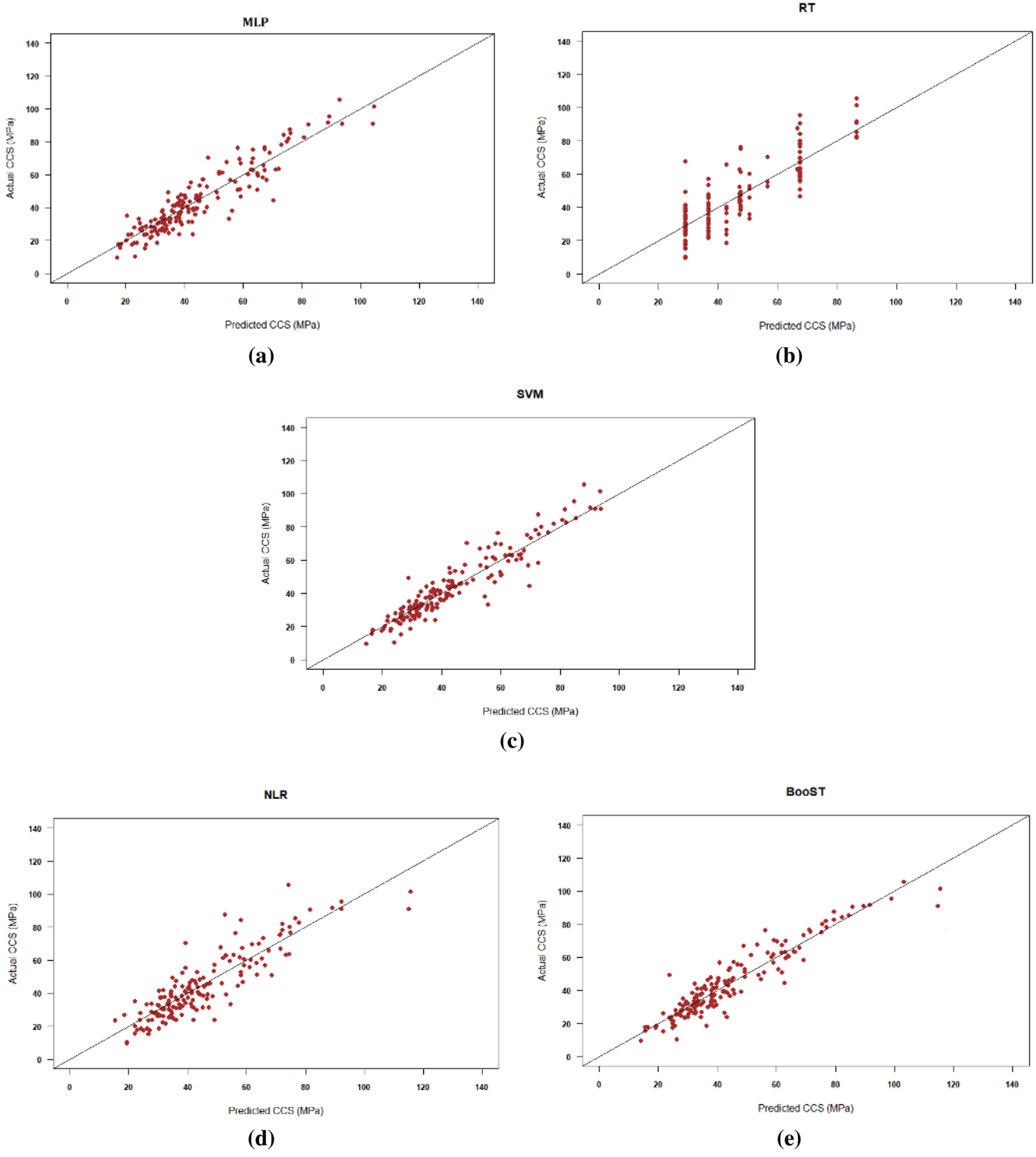


Figure A.3. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 3 **(b)** RT 3 **(c)** SVM 3 **(d)** NLR 3 **(e)** BooST 3.

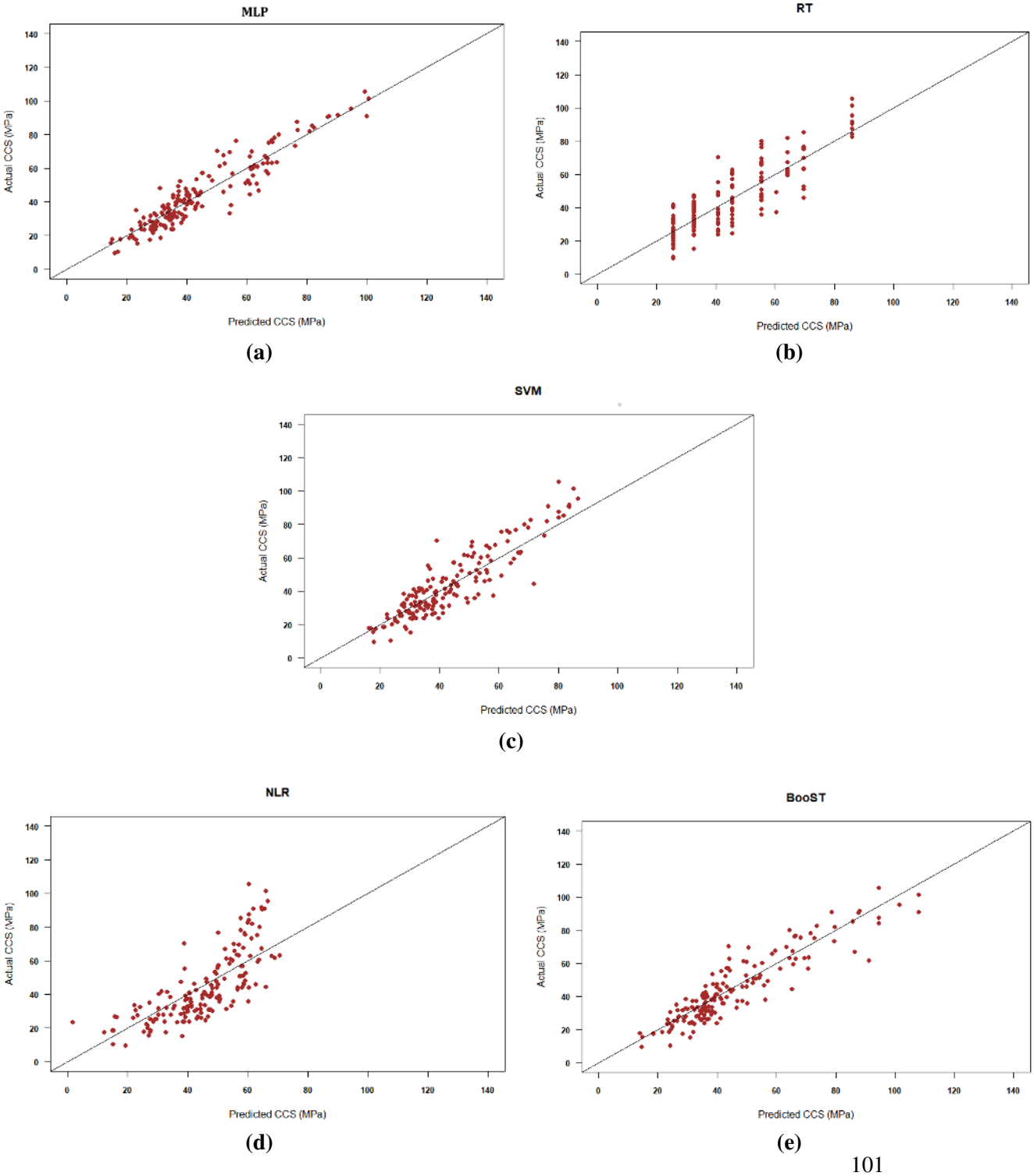


Figure A.4. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 4 **(b)** RT 4 **(c)** SVM 4 **(d)** NLR 4 **(e)** BooST 4.

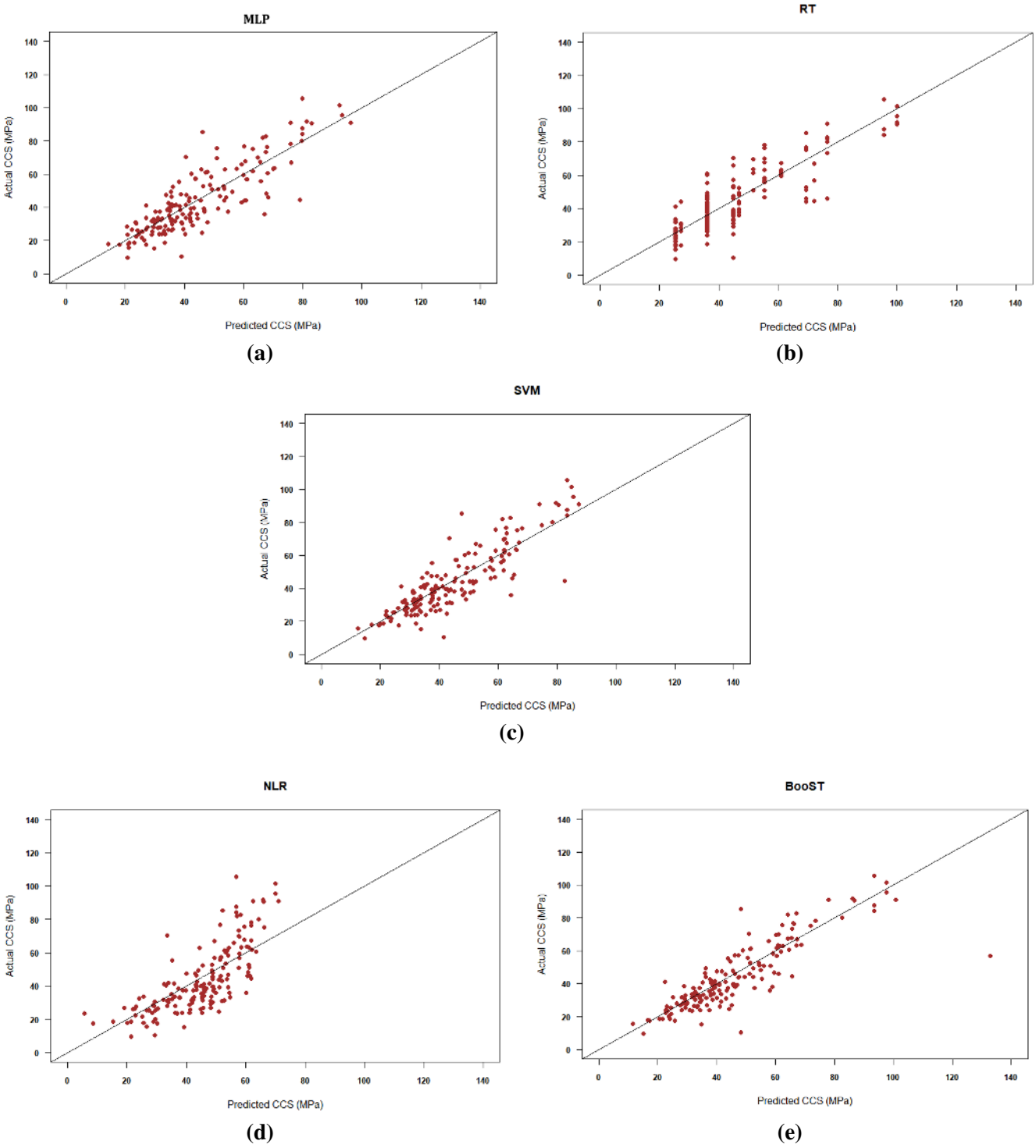


Figure A.5. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 5 **(b)** RT 5 **(c)** SVM 5 **(d)** NLR 5 **(e)** BooST 5.

the observed data points of the advanced techniques (BooST, SVM and MLP) congregates more in a straight line than NLR and RT techniques. The same can be observed in the rest of the models in the third set (model–7, model–8, model–9 and model–10) as depicted in Figure A.7, Figure A.8, Figure A.9 and Figure A.10. However, the plots from the BooST models align more diagonally that the rest of the models.

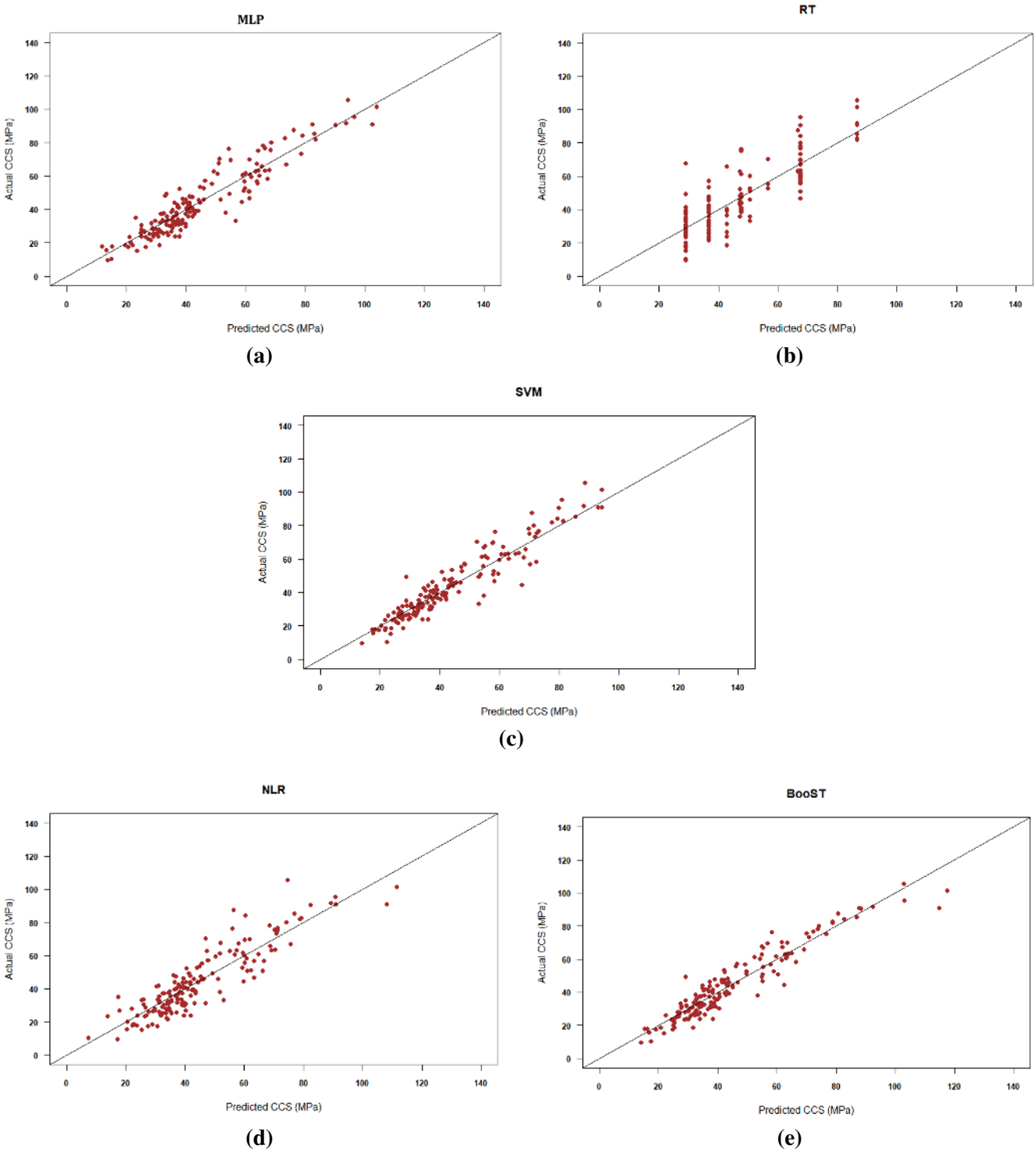


Figure A.6. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 6 **(b)** RT 6 **(c)** SVM 6 **(d)** NLR 6 **(e)** BooST 6.

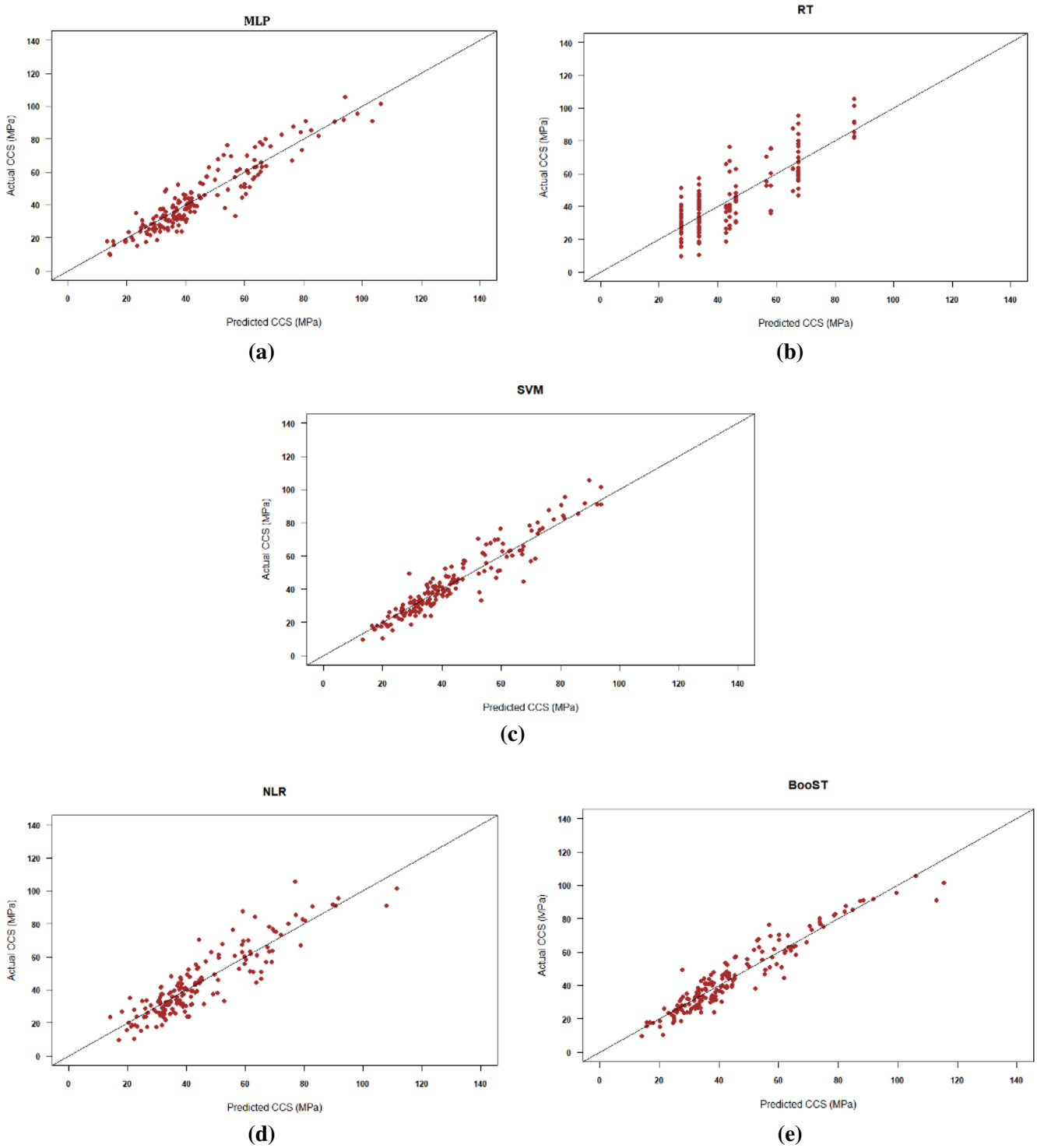


Figure A.7. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 7 **(b)** RT 7 **(c)** SVM 7 **(d)** NLR 7 **(e)** BooST 7.

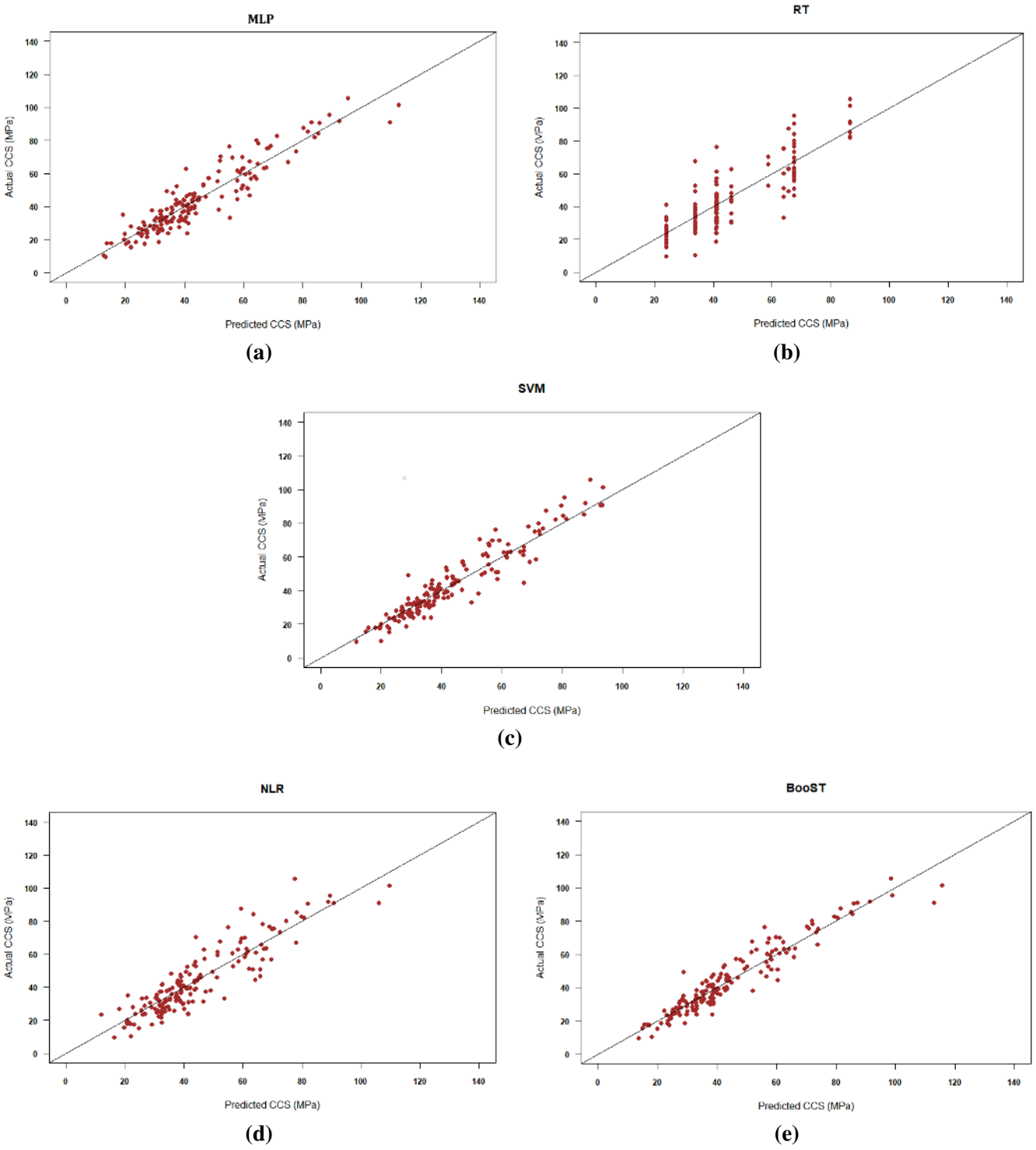


Figure A.8. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 8 **(b)** RT 8 **(c)** SVM 8 **(d)** NLR 8 **(e)** BooST 8.

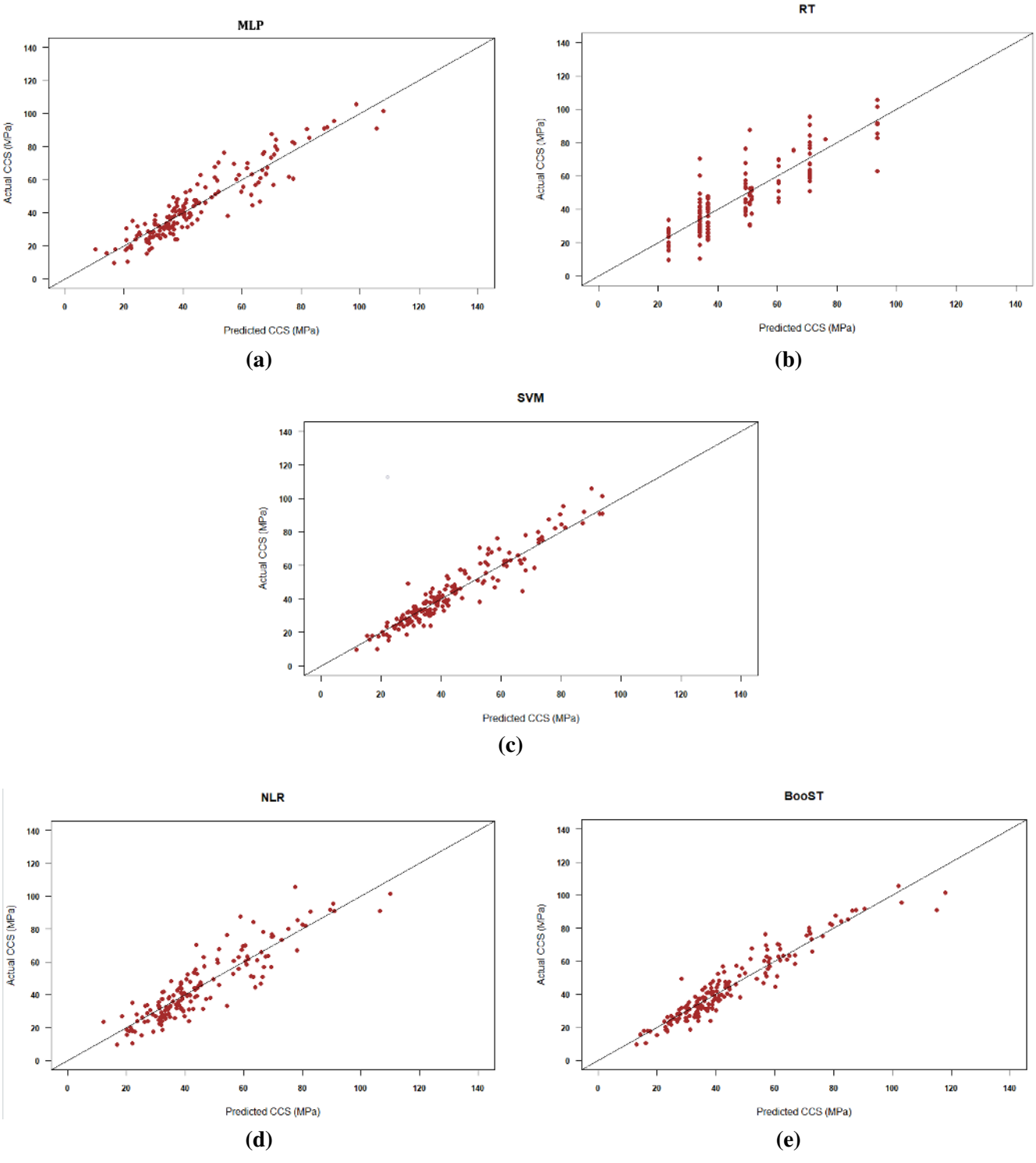


Figure A.9. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 9 **(b)** RT 9 **(c)** SVM 9 **(d)** NLR 9 **(e)** BooST 9.

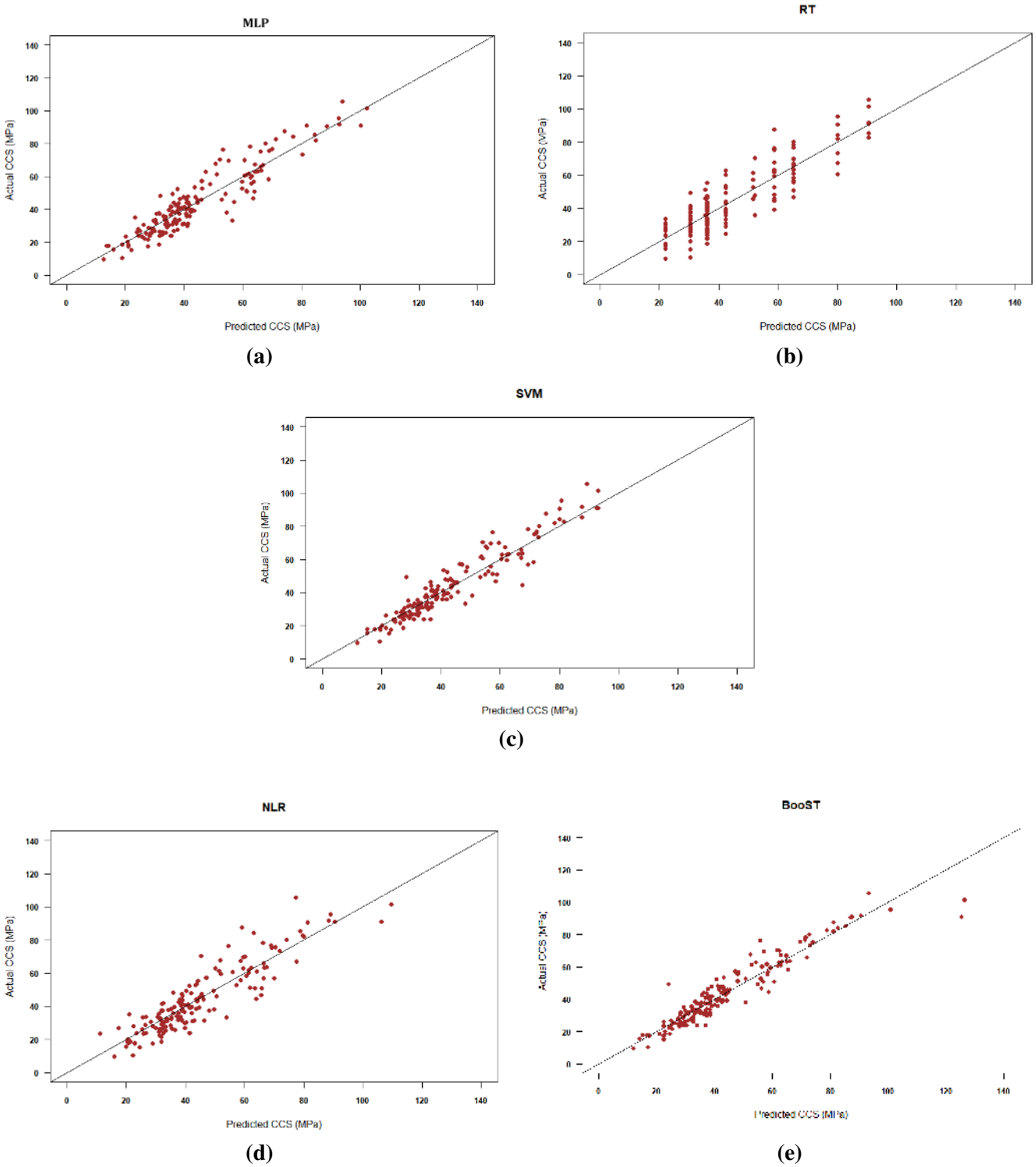


Figure A.10. Actual CCS against predicted CCS plots for each technique in the first set of predictive models **(a)** MLP 10 **(b)** RT 10 **(c)** SVM 10 **(d)** NLR 10 **(e)** BooST 10.