Experimental Investigations and Machine Learning-Based Predictive Modeling of the

Chemo-mechanical Characteristics of Ultra-High Performance Binders

by

Emily Ford

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2020 by the
Graduate Supervisory Committee:

Narayanan Neithalath, Chair
Subramaniam Rajan
Barzin Mobasher
Nikhilesh Chawla
Christian G. Hoover
Kailas Maneparambil

ARIZONA STATE UNIVERSITY

December 2020

ABSTRACT

Ultra High Performance (UHP) cementitious binders are a class of cement-based materials with high strength and ductility, designed for use in precast bridge connections, bridge superstructures, high load-bearing structural members like columns, and in structural repair and strengthening. This dissertation aims to elucidate the chemo-mechanical relationships in complex UHP binders to facilitate better microstructure-based design of these materials and develop machine learning (ML) models to predict their scale-relevant properties from microstructural information.

To establish the connection between micromechanical properties and constitutive materials, nanoindentation and scanning electron microscopy experiments are performed on several cementitious pastes. Following Bayesian statistical clustering, mixed reaction products with scattered nanomechanical properties are observed, attributable to the low degree of reaction of the constituent particles, enhanced particle packing, and very low water-to-binder ratio of UHP binders. Relating the phase chemistry to the micromechanical properties, the chemical intensity ratios of Ca/Si and Al/Si are found to be important parameters influencing the incorporation of Al into the C-S-H gel.

ML algorithms for classification of cementitious phases are found to require only the intensities of Ca, Si, and Al as inputs to generate accurate predictions for more homogeneous cement pastes. When applied to more complex UHP systems, the overlapping chemical intensities in the three dominant phases – Ultra High Stiffness (UHS), unreacted cementitious replacements, and clinker – led to ML models misidentifying these three phases. Similarly, a reduced amount of data available on the hard and stiff UHS phases prevents accurate ML regression predictions of the

microstructural phase stiffness using only chemical information. The use of generic

virtual two-phase microstructures coupled with finite element analysis is also adopted to

train MLs to predict composite mechanical properties. This approach applied to three

different representations of composite materials produces accurate predictions, thus

providing an avenue for image-based microstructural characterization of multi-phase

composites such UHP binders. This thesis provides insights into the microstructure of the

complex, heterogeneous UHP binders and the utilization of big-data methods such as ML

to predict their properties. These results are expected to provide means for rational, first-

principles design of UHP mixtures.

DEDICATION

I dedicate my dissertation thesis to my wonderful family, especially Grandma Sara Beth

Ford, Grandma Marlene Armentrout, my little brother Joey, my mom Jandi, and my dad

Henry Ford for their support and inspiration. Without you, this would not be possible.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

v

LIST OF TABLES

LIST OF FIGURES

xxi

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Concrete as a construction material has existed for thousands of years, with historic sites surviving to this day serving as testament to its longevity and versatility as a building material. Yet, there are still many unknowns in the chemical reaction processes, kinetics, and the resulting microstructures which ultimately influence concrete's mechanical properties. It is common knowledge that microstructural studies of building materials such as steel [1] or timber [2] have led to advancements in their production and quality. The large body of work available on the fundamental materials-structure-processing-property relationships of concrete have led to the use of many unconventional materials (such as waste or recycled materials from other industries), the refinement of mixture proportions, and the enhancement of tools for material and property interrogation, all of which have consequently improved design and construction practices. While great strides are being made with the development of new materials and processes, much remains to be done to advance the understanding and prediction of the mechanical behavior of cementitious binders and their relationship to microstructural and chemical characteristics. This is especially true for complex multi-phase binders such as high- and ultra-high performance cement systems, which is the focus of this thesis. In conjunction with fundamental materials-science based approaches, this work also ventures into big-data analysis using novel techniques such as machine learning to generalize and expedite the predictive effort.

A specialized concrete mixture that demonstrates desirable properties such as high strength, ductility, and durability, yet minimizing the use of greenhouse gas producing Ordinary Portland Cement (OPC) is Ultra-High Performance Concrete (UHPC). UHPC has been utilized in the construction of high strength infrastructural components, bridge repair [3] [4], slender structural elements, as well as rehabilitation of pre-existing reinforced concrete or steel structures [5]. Featuring multiple cementitious replacement materials with controlled particle size distributions to ensure dense packing, UHPC is a heterogeneous system where the cementitious materials are incompletely reacted (e.g., low degrees of hydration), owing to a low water-to-binder ratio (w/b). The cost of UHPC is significantly higher than that of conventional concrete and varies from $545/$m^3$ to $3,000/$m^3$ [6] [7] depending on the constitutive materials. At the same time, the compressive, tensile, and flexural strengths of UHPC, as well as its durability, are highly dependent on the type and amount of source materials, their packing, and their degree of reaction [6] [3]. An important objective of this study is to understand the influence of these factors on the fundamental nano-and-microstructural features of UHPCs, so that efficient microstructure-based materials design can be implemented for the family of these complex material systems.

A large amount of experimental data, both with respect to properties and microstructures, is generated in studies dealing with concrete. The traditional approach is to scientifically analyze the data and arrive at interpretations of material behavior, or to use the data for property performance prediction. With the advent of tools that examine big data and use efficient computing to derive patterns and predictions, it is instructive to use them in the field of cement-based materials, which predominantly relies on experimentation. Thus,

this study also explores the use of machine learning techniques in the context of UHPCs and other multi-phase cementitious systems.

Machine learning is an artificial intelligence technique that uses large data sets to develop prediction tools for a specific problem, which otherwise is tedious or even impossible to solve. Significant advances are happening in the topic of machine learning related to fields such as health and medicine, security, economics, robotics, etc. Past applications of machine learning to materials science include the quantification, classification, optimization, evolution, and reconstruction of microstructures, mechanical properties, and material performance along the process-structure-property-performance chain [8]. In cementitious building materials, machine learning has been applied to predicting compressive strength given the mixture design [9, 10, 11, 12, 13], optimizing concrete mixture proportions under multiple constraints to satisfy desired strength, cost, and slump [14], and reconstructing cementitious phases from prior knowledge of microstructural phases and nanoscale mechanical-chemical mapping [15, 16]. To further the understanding of the nano- and micro-scale mechanical behaviors of cementitious binders, this study will focus on the ability of machine learning to utilize real and/or virtual microstructural and chemical data sets to predict material properties.

1.2 RESEARCH OBJECTIVES

This research focuses, first, on fundamentally understanding the inter-relationships between the micromechanical, microchemical, and bulk mechanical properties of economical, non-proprietary UHP cementitious binders developed in a previous work at ASU. Next, the large datasets made during this study, supplemented with datasets from other well-known studies, are utilized in conjunction with customized machine learning

algorithms to classify microstructures and/or predict the mechanical properties of complex heterogeneous systems. Finally, the ability of machine learners to interpret microstructural images and predict the mechanical properties for any two-phase composite system, including UHP mortars, is explored. The major objectives are listed below:

1. To link the micromechanical properties of UHP binders to their constitutive materials – which is a function of the type and amount of binder materials (e.g., cement, limestone, fly ash, silica fume) and the extent of their reaction.

2. To evaluate the fundamental microstructural phase chemistry in UHP binder systems and examine the chemistry-property relationships at the nano-/micro-scale.

3. To develop machine learning algorithms capable of classifying the complex material microstructure into distinct cementitious phases, given the chemical intensity and/or the micromechanical features.

4. To develop machine learning regression algorithms capable of predicting the micromechanical properties of different phases in complex heterogeneous binders based on information from microstructural chemical maps (which can then be upscaled using analytical homogenization or numerical tools) to reduce the reliance on highly expensive, sophisticated, and time-consuming experiments.

5. To combine microstructural image-based machine learning techniques and finite element analysis to predict the mechanical properties of a generic two-phase composite, as a means of generalizing this approach for multi-phase media.

1.3 RESEARCH APPROACH

The first task of this study deals with a thorough micromechanical investigation of the mechanical properties (hardness and stiffness) of conventional OPC and UHP cementitious binders after 30 and 90 days of hydration. Grid nanoindentation, where thousands of nanoscale indents are made on polished cementitious binder surfaces to account for the statistical heterogeneity in the sample, is performed. Nanoindentation coupled with Bayesian statistical analysis help reveal the mechanical clustering of distinct phases in the complex, heterogeneous microstructure. These clusters are chemically identified by synthesizing data from X-ray diffraction (XRD) and thermogravimetric analysis (TGA), as well as an understanding of the hydration products of cementitious systems. The UHP mixtures are comprised of several cementitious replacement materials and a low w/b that produces phases with higher hardness and stiffness than the OPC paste owing, in part, to the retention of a significant amount of unhydrated starting minerals. Furthermore, formation of an Ultra High Stiffness (UHS) hydration phase is also preferred in low w/b mixtures, the origins of which is probed in this work through experiments and packing density calculations of the resultant C-S-H gel. Elastic homogenization approaches are performed to ensure that the nanomechanical data can reliably be upscaled into bulk material properties.

Once the mechanical properties of the different microstructural phases in UHP binders are determined, the next phase of the work focuses on developing qualitative relationships between the micromechanical properties and cementitious phase chemistry. To accomplish this goal, nanoindentation measurements of stiffness and hardness are paired with chemical maps from scanning electron microscopy (SEM) coupled with

energy-dispersive X-ray spectroscopy (EDS) analysis. The chemical mapping is carried out in a qualitative manner rather than a quantitative manner. Bayesian statistical clustering is repeated with this additional microchemical information, providing new insights into the chemistry and micromechanics of UHP phases. The chemical intensity ratios, including the Ca/Si and Al/Si ratios, are used to infer the changes in the reaction product (C-(A)-S-H gel) chemistry. The relative dependence of Al incorporation in the gel based on the gel chemistry is explored in detail. The relationship between stiffness and hardness of the reaction product phases in the UHP binders, and their dependence on the reaction product chemistry is elucidated.

The next phase of this thesis deals with synthesizing the large amount of data generated in this work, as well as related works in recent years, towards useful, generalizable models for material performance. Given the large quantity of data gathered on nanoindentation stiffness, nanoindentation hardness, microstructural images, and chemical maps, machine learning algorithms are deemed ideal to develop classification and regression models based on these data sets. Classification algorithms are generated in this work using the elemental intensity chemical maps and the micromechanical data such that each indent can be attributed to a microstructural phase. The fundamental premise is that if microstructural chemical maps of complex binders such as UHPC are available (which are easy to obtain using a generic scanning electron microscopes outfitted with an EDS system), then the correlation between phase chemistry and phase nanomechanical properties can be accurately identified. Such phase identification through machine learning would eliminate or minimize the need for complex nanomechanical experimentation. Phase labels for training data are obtained from the microchemical

6

mapping, and the final classification algorithms are tested on data not used in model generation to determine the degree to which they could predict the cementitious phase, without knowing the original composition of the paste or the degree of reaction of the system (i.e., age). For verification purposes, the classification algorithms are also applied to the data for more conventional cementitious mixtures.

Further, regression machine learning techniques are employed to predict the nanoindentation stiffness from the elemental intensities obtained from the SEM and EDS maps. For relatively less complex systems such as plain OPC pastes, pastes containing smaller amounts of reactive cement replacement materials, and well-hydrated systems, chemical intensity maps coupled with regression machine learning provide an easy and reliable methodology to predict the phase stiffnesses. On the other hand, the stiffnesses of the UHP paste phases are difficult to model using any of the machine learning methods. The difficulty of regression machine learners to predict the stiffness based solely on the elemental intensities for UHP pastes stems from the microstructural complexity and lack of a large amount of high-quality training data. Specifically, it is noticed that the prediction quality is better for the major reaction product (C-S-H), but a large scatter in the data for the high modulus phases and the mixed reactant/product phases reduced the overall predictive ability.

While the previous explorations of classification and regression are based on discrete data obtained from microstructural images, the final component of this thesis forays into using microstructural images themselves, coupled with finite element analysis, as training data for the machine learning algorithms. Here, generic two-phase composites – e.g., a unidirectional carbon fiber composite, a UHP cement paste-sand mortar, and a cement

7

paste with steel particle inclusions, with available experimental data are considered. Random images are generated representing different volume fractions, size distributions, and shapes of the inclusions. These images are used as inputs into a linear elastic finite element analysis to predict the elastic properties and stresses in the constituent phases of the composites. The results of validated finite element analysis are utilized as the ground truth to train the machine learning models. Given other two-phase microstructures that the model has not yet encountered, the algorithm enables adequate prediction of the composite mechanical properties. Figure 1-1 outlines the evolution of this study and summarizes the study's approach to achieving the objectives.

Objectives 1-2
Approach: Nanoindentation, SEM/EDS, and Bayesian statistical clustering to connect the micromechanical and microchemical properties of UHP cementitious binders.

Objectives 3-4
Approach: Develop machine learning models to predict the cementitious phase and mechanical properties of UHPC cementitious binders.

Objective 5
Approach: Train machine learners to predict the mechanical properties of any two-phase composite microstructural image to generalize image-based micromechanical property prediction.

Figure 1-1: Flowchart summarizing the research approach of this study and how the approach will address each of the objectives.

1.4 DISSERTATION OUTLINE

The core of this dissertation consists of five research papers (Chapters 3-7) that have been published, submitted, or are being prepared for submission. The overall organization of the thesis is shown below.

Chapter 1 provides a brief introduction about the topic of the study and presents the background and significance of cementitious materials and machine learning. The research objectives are identified, and a chapter outline is presented.

Chapter 2 presents a thorough literature review covering the UHPC materials and methods used for their microstructural characterization. The basics of machine learning and its methods are also introduced here, though explicit details on classification and prediction can be found in the corresponding chapters.

Chapter 3 demonstrates the use of experimental nanoindentation results along with Bayesian statistical clustering to understand the micromechanical response of individual phases in UHPC microstructures. The heterogeneous nature of UHP binders is first explored solely using information available from the micromechanical testing. Data from this study is used in conjunction with homogenized models to predict the elastic response of the UHP binders.

Chapter 4 evaluates the UHP binders from a combined micromechanical and microchemical viewpoint, utilizing nanoindentation, scanning electron microscopy, and energy dispersive X-ray analysis. Bayesian clustering is again performed with both the nanoindentation stiffness and relative chemical intensity ratios at every indentation point. Resulting clusters are compared to understand the role of different cementitious

replacement materials and reaction progress on the phase chemistry, which is related to the micromechanical properties.

Chapter 5 introduces machine learning classifiers that can be used to identify the major cementitious phase at each indentation point in a microstructure, given microchemical and/or micromechanical information. Data from several different cementitious microstructures are individually used to train the classifiers and establish the relationships between mixture complexity and accurate phase identification.

Chapter 6 showcases machine learning regression models that are used to predict the microstructural stiffness of individual points within cementitious binders given the relative chemical intensity from SEM and EDS. Machine learning models are developed from several different cementitious mixtures to relate the heterogeneity to the ability of the algorithms to learn from the data. Suggestions to improve the machine learning predictions of complex microstructures such as UHP pastes are explained.

Chapter 7 explores the use of generic two-phase microstructural images, coupled with finite element analysis, to train machine learning regression algorithms to predict mechanical properties. Machine learning models capable of interpreting two-phase microstructural images with varying inclusion volume fractions, size distributions, and shapes are generated. The application of dimensionality reduction techniques on the correlation functions used as inputs to image-based machine learning regression models are discussed.

Chapter 8 summarizes the conclusions of this study and recommends related topics for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter is divided into three main sections. The first section discusses ultra-high

performance (UHP) binders and the different cement replacement materials used, along

with their effects on the microstructural development of UHP binders. The second section

deals with the sample preparation methods and materials characterization tools utilized in

this study. Lastly, the principles and application of machine learning methods as applied

to materials characterization is discussed.

2.2 ULTRA-HIGH PERFORMANCE CONCRETE AND CEMENT

    REPLACEMENT MATERIALS

2.2.1   Ultra-High Performance Concrete (UHPC)

Ultra-High Performance Concrete (UHPC) is a specialized concrete with a low water-to-

binder ratio (w/b) and enhanced particle packing, resulting in low porosity, high

resistance to cracking, high compressive and tensile strengths, high ductility, and

enhanced durability. Compared to High Performance Concrete (HPC), or concrete only

produced with Ordinary Portland Cement (OPC), UHPCs enable production of

infrastructural components at lower life-cycle costs owing to their high strength and

durability. With high longevity leading to financial savings in terms of reduced

maintenance and longer service life, UHPCs are considered as one of the most important

advances in the materials field of structural engineering. UHPC is defined by the United

States Federal Highway Administration (FHWA) as a cementitious composite with

compressive strength in excess of 150 MPa (22 ksi) and tensile strength in excess of 5

MPa (0.72 ksi) [3] [4]. Due to a low w/b and high particle density, UHPC has greater durability compared to HPC or other types of conventional concrete. For a desired strength, UHPC can also be used to produce structural elements with reduced cross-sectional areas, and therefore self-weight. Typically, UHPC is produced with a w/b between 0.15-0.0.25 and dense particle packing through the use of multiple cement replacement materials [4]. With proper mixing techniques, the low w/b helps in reducing the porosity and often eliminates the need for air entrainment [4] as UHPC has sufficient tensile strength to resist the stresses induced by freezing-and-thawing.

Applications of UHPC are well-documented, and include the construction of high-strength columns and bridge decks, strengthening of pre-existing reinforced concrete or steel structures [5], as well as field-cast connections between bridge elements [3]. Three key challenges to the implementation of UHPC into common construction practice include the high initial cost, the specialized mixture design needs, and the equipment requirements for installation. A cubic yard of UHPC is estimated to cost 13-17 times more than that of conventional concrete in North America [4] [17] due to the stringent performance requirements that demand expensive materials and processing. Specialized tools and considerations for the installation of UHPC include high energy mixers to properly mix the constituent materials in such low w/b ratios, the consistent distribution and orientation of fibers, as well as rigid curing requirements to maximize the strength and durability [17]. Several studies have addressed UHPC mixture design from the viewpoint of maximizing particle-packing in order to produce the greatest gains in durability and strength [18] [19] [20]. Efforts to reduce the cost [6] and increase the sustainability of UHPC through the incorporation of readily available materials such as

river sand to substitute for quartz sand [21], pozzolanic industrial wastes like silica

fume/microsilica from silicon production, fly ash from coal burning power plants, or

mine tailings [22] have yielded UHPCs with high compressive strength [23].

Incorporating so many different materials to form a cohesive cementitious binder

introduces specific challenges. The major consideration when producing UHPC is to

ensure mechanical homogeneity, maximum particle packing density and minimum size of

flaws [17]. The following sections address the cementitious replacement materials

included in the UHP binders used in this study, and provides an overview of the effects of

their incorporation into the microstructure and the properties of the resulting composite.

2.2.2   Fly Ash

Fly ash is an amorphous material produced as a by-product of coal-burning power

generation plants. ASTM C618-19 identifies two main classes of fly ash: Class C with

greater than 18% calcium oxide leading to cementitious and pozzolanic properties, and

Class F with less than 18% calcium oxide with only pozzolanic properties [24]. Different

types of fly ash are generated by differences in burning equipment and types of coal [24].

Cementitious materials react with water to form hard and stiff products such as calcium

silicate hydrate (C-S-H) and calcium hydroxide (CH). A pozzolanic material reacts with

cement hydration products, especially calcium hydroxide (CH), to produce additional C-

S-H gel with a decreased Ca/Si ratio [25] [26] [27]. It is well known that reducing the

Ca/Si ratio increases the elastic modulus, strength, and hardness of the C-S-H gel [25]

[28] [29]. In general, Class F fly ash can be used to replace 10-30% of cement by mass

13

while Class C fly ash can be used in higher proportions [4]. Both types of fly ash are indicated in the ternary diagram of CaO-Al$_2$O$_3$-SiO$_2$ in Figure 2-1 [30].
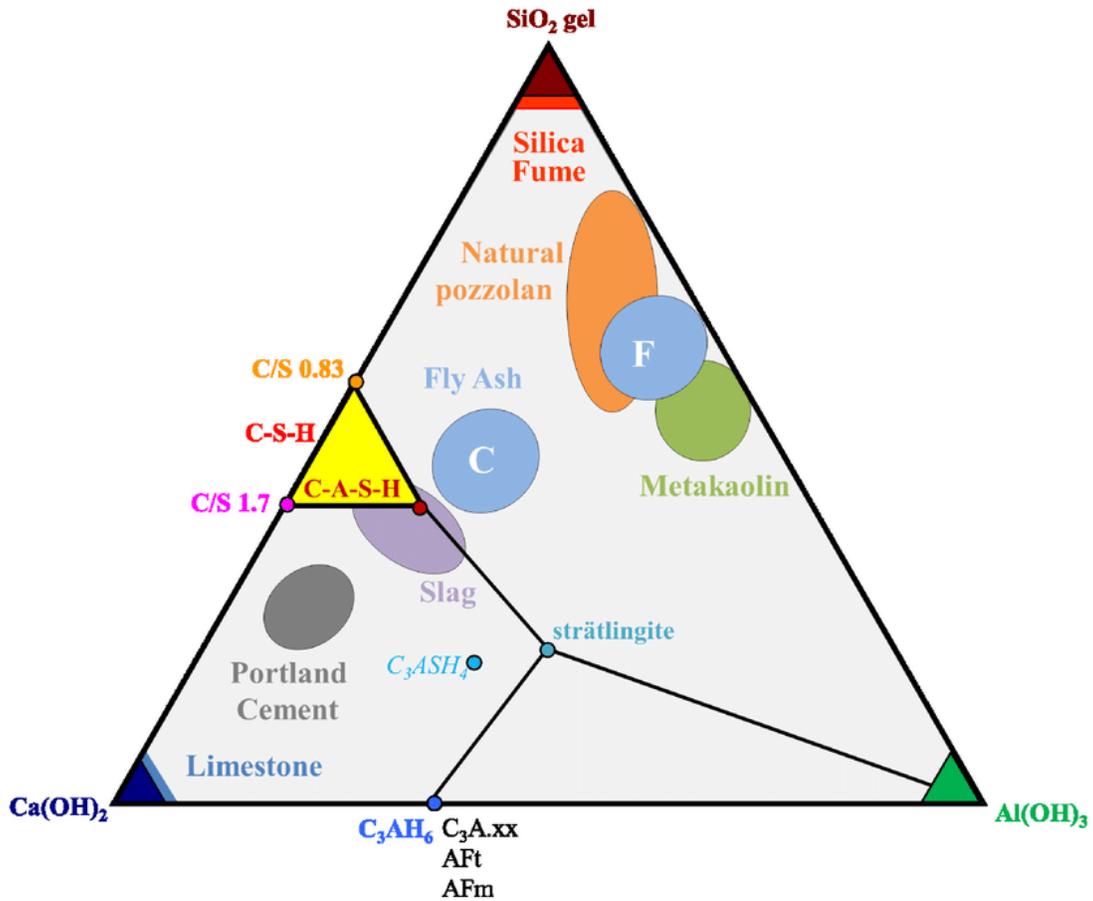
Figure 2-1: CaO-Al$_2$O$_3$-SiO$_2$ ternary diagram of cementitious materials and hydrate phases from the work [31] adapted from [27].

When incorporated into cementitious mixtures, fly ash reduces the heat of hydration [32], delays setting [23] [32] [4], and acts as a filler at early ages. Fly ash has a greater effect on the long-term compressive strength due to its delayed reaction and pozzolanic activity [32]. Although fly ash demonstrates large variability in its chemical composition due to types and sources of coal and processing, the components of glassy (amorphous silica) phase (SiO2, Al$_2$O$_3$, and Fe$_2$O$_3$) have a high correlation with the pozzolanic activity and

14

consequently the compressive strength [32]. The aluminates in fly ash also participate in other cementitious reactions such as its reaction with limestone to form carboaluminate phases [33] [34] [35].

2.2.3   Microsilica

Microsilica, or silica fume, is generated as a by-product of silicon metal manufacturing and is comprised almost entirely of SiO_2. The microsilica particles are very fine, an entire magnitude smaller than that of OPC [4], such that its incorporation densifies the microstructure, but also causes potential agglomeration and decreases the workability [4] [19]. Microsilica is a highly reactive with a small particle size, and thus large surface area, which serves as nucleation sites for high pozzolanic activity [27] [19]. These nucleation sites outside of clinker grains generate early pozzolanic reactions to consume CH and produce additional C-S-H [27]. The addition of microsilica surface area for hydration is used to off-set the retardation effects of fly ash or superplasticizers needed to achieve functional workability in UHP mixtures [19]. Scanning electron microscope (SEM) images of microsilica are shown in Figure 2-2 to demonstrate their size and morphology. The refined microstructural packing from microsilica incorporation leads to improved moisture and ionic transport resistance of concrete [23]; however issues such as additional air entrapment due to the increased viscosity when combining components needs to be carefully considered [19].
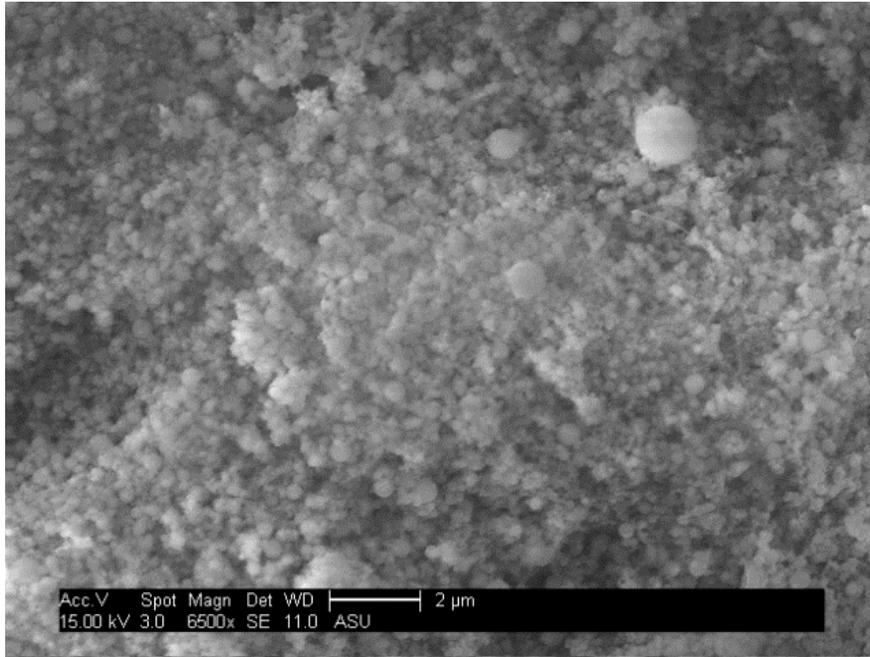
(a)
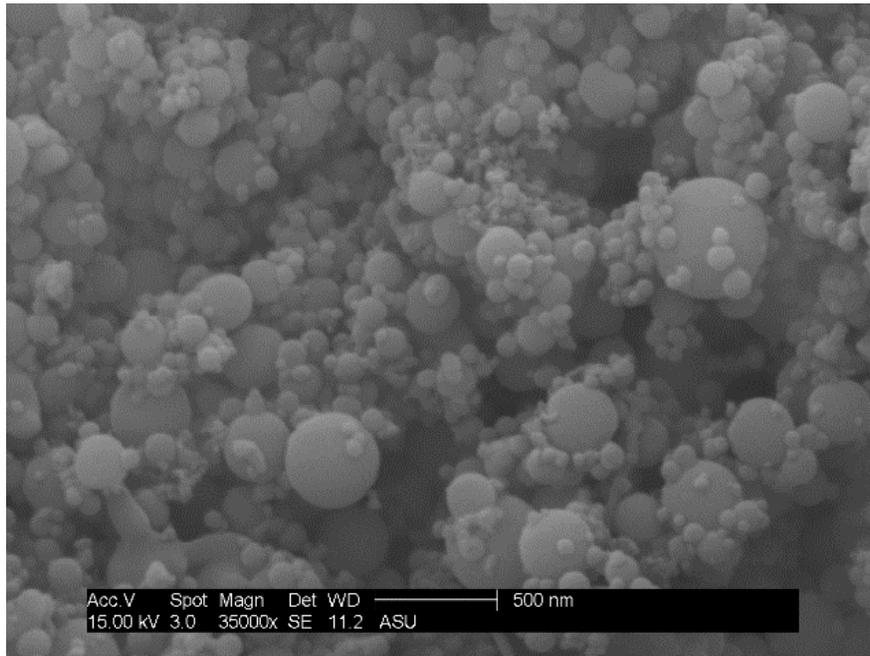
(b)

Figure 2-2: Scanning electron microscope (SEM) images in secondary electron (SE) mode of microsilica (silica fume) at: (a) 6500x magnification and (b) 35000x magnification.

16

2.2.4 Limestone

Limestone is an abundant mineral comprised primarily of $CaCO_3$ and is commonly employed as an "inert filler" in UHPCs (though they demonstrate minor reactivity under favorable chemical conditions) to reduce the OPC content, reduce capillary porosity, and improve workability. Use of limestone with a fineness similar to or lower than the average OPC particle size has been shown to improve workability [36] [37]. At higher dosages (in excess of 15% by mass of cement), limestone has a dilution effect that can reduce the strength of the cementitious binders [35]. The term "inert filler" only applies to limestone in cementitious systems lacking sufficient alumina to react with the calcium carbonate [35] [33]. Limestone is known to react with hydratable aluminate phases from reactants such as tricalcium aluminate (C_3A) in OPC, and especially pozzolanic admixtures like metakaolin and fly ash [35], to produce stable carboaluminate phases such as calcium monosulfoaluminate, calcium hemicarboaluminate, and calcium monocarboaluminate [36] [35]. The carboaluminate phases are reported to possess similar stiffness as limestone or microsilica particles [38, 39] and consumes portlandite (CH) [33] [35], thereby leading to an increase in the overall compressive strength and stiffness of the cementitious mixture. The effects of limestone as a filler or microstructural enhancer in UHP binders is highly dependent on the particle fineness and its proportion relative to hydratable aluminates [40].

2.2.5 Superplasticizers

Superplasticizers are the chemical admixture used to dramatically increase the flowability of cementitious mixes. A superplasticizer is an individual or mix of polymers which are absorbed into the surface of particles and carry a negative surface charge [4]. This

17

negative charge around the surface of particles repels water molecules and therefore increases workability. Adsorption of admixtures onto the particle surfaces depends on the composition and surface are of particles [4]. Typically, superplasticizer is added by ~5% by mass of binder. Careful consideration and note of how much superplasticizer is added must be made when developing and comparing cementitious mixture designs.

2.3 METHODS

2.3.1 Materials, Mixing, and Curing

The binder contents in the UHP pastes, addressed in detail in each individual chapter, are based on previous studies at ASU [18, 41] that proportioned a family of UHP binders based on optimal particle packing, rheology, and substantial clinker factor reduction. The mixing procedure for UHP pastes followed that used in previous authors [18] and ASTM C1738, using a high-speed shear mixer and blending all dry powders prior to wet mixing. The procedure is briefly summarized as follows: (i) adding all water and superplasticizer to the mixer, (ii) mixing at 4000 rpm for ~30 seconds and carefully adding the blended dry powders to the mixer, (iii) mixing at 12,000 rpm for ~30 seconds, (iv) allowing the paste to rest for 2 minutes, and (v) final mixing at 12,000 rpm for ~90 seconds. Following mixing, the paste was poured into plastic test tubes 2.5 cm (1 inch) in diameter, sealed, and allowed to cure for the desired time period.

2.3.2 Mercury Intrusion Porosimetry (MIP)

The pore structure of the pastes was evaluated using Mercury Intrusion Porosimetry (MIP). In this method, crushed paste samples are filled with mercury at high pressures to quantify the size and volume of pores. For MIP experiments, small pieces of the paste were weighed and placed in the low-pressure chamber of the porosimeter (Quantachrome

Instruments Pore Master) pictured in Figure 2-3. The sample was filled with mercury

starting from ambient pressure to 345 kPa (60 psi). The sample was then placed in the

high-pressure chamber and the applied pressure increased to 414 MPa (60,000 psi). The

pore diameter, d, as a function of the intrusion pressure was obtained from the Washburn

equation [42] as:

$$d = \frac{-4\sigma\cos(\theta)}{\Delta P} \qquad (1)$$

where ΔP is the difference in the pressure between successive steps (MPa), θ is the

contact angle between mercury and the cylindrical pore, taken as 130° in this study, and σ

is the surface tension between mercury and the pore walls, taken as 485 mN/m [43] [44]

[45]. The critical pore size, indicative of the percolating pore size of the material, was

obtained from the peak of the differential pore volume curve. The porosity of the sample,

φ, was determined from the cumulative volume of mercury intruded and the bulk density

of the sample, obtained from helium pycnometry.

Figure 2-3: Quantachrome Instruments Pore Master mercury intrusion porosimeter
utilized in this study.

2.3.3 Thermogravimetric Analysis (TGA)

Thermogravimetric analysis is the technique of heating a powdered sample at a controlled rate to measure the change in mass as the various components decompose at different temperature ranges. The mass loss profile plots the change in mass with respect to the temperature to identify the presence and relative mass of reaction products. TGA was performed on cement paste samples to primarily determine the calcium hydroxide (CH) and calcium carbonate contents [46]. The tests were performed in an inert N_2 environment at a gas flow rate of 20 ml/s. The samples were heated from ambient temperature to 900^oC at a heating rate of 15^oC/min in a Perkin Elmer simultaneous thermal analyzer (STA 6000) pictured in Figure 2-4.
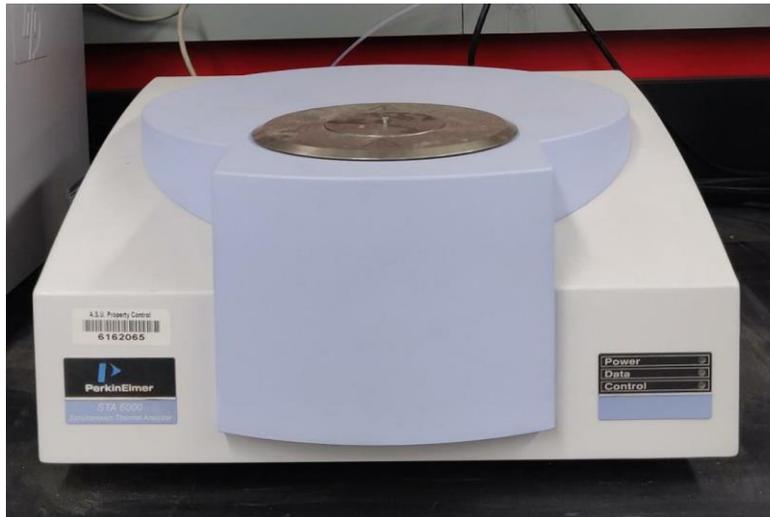
Figure 2-4: Perkin Elmer simultaneous thermal analyzer (STA 6000) utilized in this study.

2.3.4 X-Ray Diffraction (XRD)

X-Ray Diffraction (XRD) is the study of the crystalline structure of materials using the angle at which incoming X-rays are diffracted. XRD spectra were obtained using a

Siemens D-5000 Powder X-ray Diffractometer utilizing a Cobalt (Co) Kα radiation with

2 mm and 0.2 mm slits to identify the crystalline phases in the samples, pictured in Figure

2-5. Following previous studies of cement pastes, the scanning angle (2θ) range used was

from 10$^\text{o}$ to 80$^\text{o}$ with a step size of 0.02$^\text{o}$ and 1 second per step [47, 48, 49], with MDI

Jade 9 software used for peak identification.

Figure 2-5: Siemens D-5000 Powder X-ray Diffractometer utilized in this study.

2.3.5 Nanoindentation

Nanoindentation is a testing method to characterize the nanomechanical response of

materials or material phases. Here, a tip of known geometry penetrates the surface of a

sample and tracks the applied load and depth of penetration into the sample. In studies of

cementitious materials, grids of nanoindentation points were used to characterize the

hardness and linear elastic stiffness of cement hydration products [50] [51] [52] with

phase identification determined via statistical nanoindentation [50] [51]. Figure 2-6(a) is

a photograph of the Ultra Nanoindentation Tester (UNHT^3; Anton Paar) utilized in

testing and Figure 2-6(b) is a close-up of nanoindentations made on the surface of a

polished UHP binder by the diamond tipped Berkovich indenter.

(a)

20μm
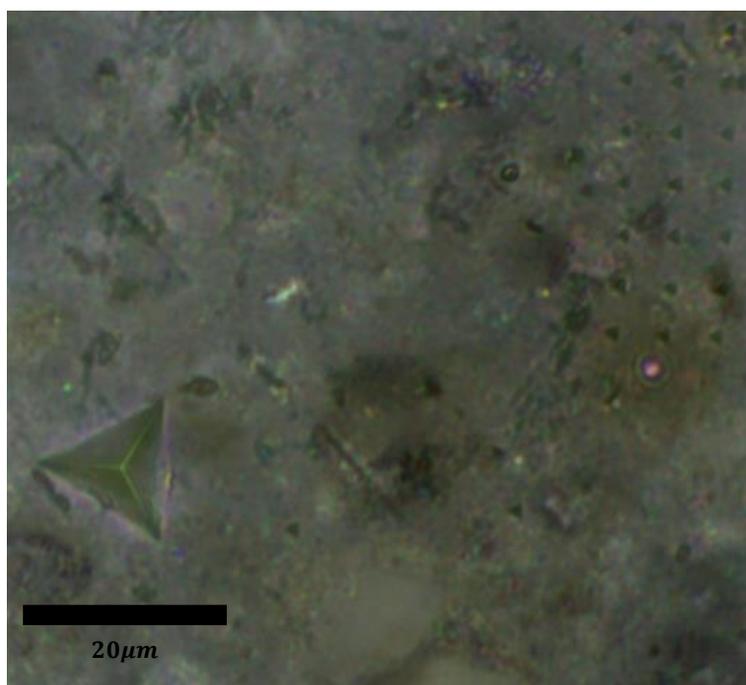
(b)

Figure 2-6: (a) Anton Paar Ultra Nanoindentation Tester (UNHT^3) utilized in this study and (b) 200x magnification microscope image of a microindent (lower left) and part of a nanoindentation grid (top right) on a polished UHP paste surface.

Use of nanoindentation to characterize the hardness and Young's modulus of cement

hydration phases has been performed on a variety of cementitious mixtures. Several

studies have focused on the identification and characterization of hydration products of

OPC [52] [53] [54] [55]. Nanoindentation has also been used to characterize the changes

in mechanical properties of OPC pastes undergoing heat treatment [56]. Similarly,

nanoindentation has also been used to characterize the hydration products of systems

proportioned using different w/b [57]. Nanoindentation has also been used in studies

incorporating different mineral admixtures that partially replace cement. For example,

cement pastes with fly ash replacement of cement at 20%, 30%, and 60% by mass were

studied [58] [59] [60]. Nanoindentation of cements using slag as a 50% and a 60%

replacement by mass were addressed as well [60] [61]. Silica fume replacement of

cement by mass ranging from 8% to 32% was also studied using nanoindentation in

previous studies [59] [62] [63].

In most of these studies, the four most commonly identified hydration phases via

nanoindentation were low density (LD) C-S-H (also called outer product C-S-H), high

density (HD) C-S-H (also called inner product C-S-H), ultra-high density C-S-H mixed

with CH, and unhydrated cement clinker. Nanoindentation has also been used to

characterize the constituents of Portland cement powder including clinker, alite, belite,

tricalcium aluminate, and calcium aluminoferrite [64]. The hardness and Young's

Modulus of unreacted fly ash particles in blended cement pastes have also been

investigated by nanoindentation [58] [59].

Indentations are made by applying a load to the Berkovich three-sided pyramid

nanoindenter tip with a known Poisson's ratio and Young's Modulus at a constant rate

24

until a determined maximum applied load or maximum displacement into the sample is achieved [50] [51]. The load is held for a certain time, then unloaded at a constant rate. The load applied to the tip and the subsequent displacement into the sample is plotted, such as shown in Figure 2-7.
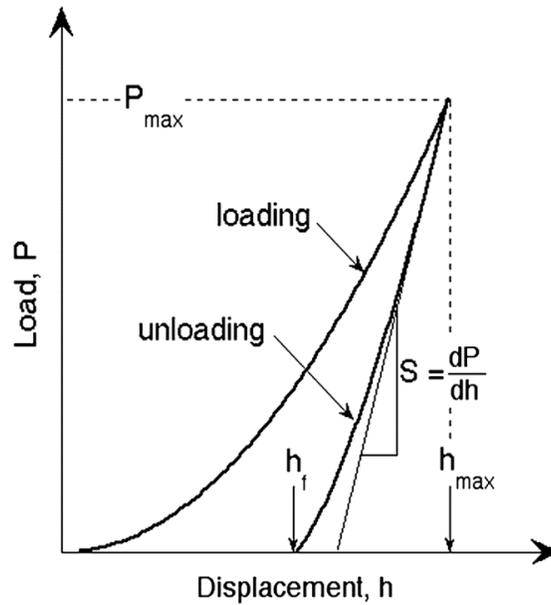
Figure 2-7: Ideal force-displacement curve generated during indentation, taken from the work [65].

According to Oliver and Pharr, the linear unloading portion of the force-displacement curve alongside the contact area of the tip to the sample surface is used to calculate the hardness (H) and the effective Young's Modulus (E_{eff}) [65] [66]. The hardness is calculated as:

$$H = \frac{P_{max}}{A} \tag{2}$$

Where P_{max} is the maximum load applied and A is the contact area of the indenter into the surface, not to be confused with the residual area left after the indenter has been removed from the surface. The effective linear elastic Young's Modulus is found as:

25

$$E_{eff} = \frac{\sqrt{\pi}}{2} * \frac{S}{\sqrt{A}} \qquad (3)$$

Where S is slope of the unloading curve [65] [66]. The effective Young's Modulus is related to the known Poisson's ratio and Young's modulus of the tip, as well as an assumed value for the Poisson's ratio of the material.

$$\frac{1}{E_{eff}} = \frac{1 - v^2}{E} + \frac{1 - v_i^2}{E_i} \qquad (4)$$

Where v is the Poisson's ratio of the sample, v_i is the Poisson's ratio of the nanoindenter, and E_i is the Young's Modulus of the nanoindenter [65] [66].

Oliver and Pharr developed the numerical method to estimate the contact area of the tip A as:

$$A = \sum_{n=0}^{8} C_n (h_c)^{2-n} \qquad (5)$$

Where C_n are the constants by curve-fitting and h_c is the depth of the indent into the material [65] [66]. The total depth of the indent, h, is equal to:

$$h = h_c + h_s \qquad (6)$$

Where h_c is the depth of the indent into the material and h_s is the surface displacement until the edge of the indent.

2.3.6 Scanning Electron Microscopy (SEM) and Energy Dispersive X-Ray Spectroscopy (EDS)

Scanning electron microscopy (SEM) and energy-dispersive X-ray spectroscopy (EDS) are two common materials characterization techniques based on subjecting the sample to high-energy electrons and detecting the resulting signals. In the case of SEM possible modes include secondary electron (SE) detection as well as back-scattered electron (BSE) detection. Secondary electrons result from the inelastic scattering of the primary and

backscattered electron beams and are used to generate detailed surface topographical

images. Backscattered electrons are generated from the elastic scattering events between

the source beam and the sample, resulting in a greater energy and larger interaction

volume, but lowered spatial resolution compared to SE [67]. An important feature of BSE

imaging is the brightness contrast, which is proportional to the average atomic number of

the phase [67] [68]. The backscatter coefficient, a measure of the backscattered electron

fraction, is estimated to have a cubic relationship to the pure element atomic mass [67].

For multi-element phases, the backscatter coefficient may be estimated using the mass

fractions of each element to determine the contrast between constituents. In cements, the

contrast between alite and belite is strong, while the contrast between belite and

tricalcium aluminate is too low to distinguish between the two phases [67] [68].

In energy dispersive X-ray spectroscopy (EDS) analysis, the source electron beam is

impinged on the sample and the characteristic X-rays resulting from the radiative

electronic transition within the sample are detected. The energy of the X-rays returning

from the surface are mapped, generating a set of statistical curves within each pixel [68].

These curves are well-studied, and their relative heights and energies are known to

correspond to different elements. Elements can be quantified from EDS images using

standards of known chemical composition; otherwise the analysis is qualitative in nature

and represents the relative abundance of each element throughout the sample. Several

studies have utilized nanoindentation paired with qualitative and/or qualitative SEM-EDS

analysis in order to establish the chemical composition of each location within the

nanoindentation grid [51] [59] [69] [70] [71]. The technique is destructive to the surface

of the sample; therefore after scanning over the grids they are effectively destroyed, and

27

further scanning will reflect the composition below that of the interaction volume. The

interaction volume and escape depth of the signal X-rays are dependent on the

microscope parameters as well as the sample density and composition [69] [72]. It has

been shown through Monte-Carlo simulations that, for both clinker and C-S-H phases,

most of the characteristic X-rays escaping the material using a 15 keV accelerating

voltage at a 15 mm working distance are generated within a depth of 2 μm [69] [72].

Continuous spectrum background (signal noise) is inherently present in EDS mapping,

leading to low elemental intensities even in spots where the element is not present [68].

This necessitates an understanding of the relative elemental intensities that can occur in

cement-based materials to accurately interpret the EDS maps. In terms of sample

preparation, cement is a non-conductive material, and thus a conductive coating such as

carbon or gold-palladium must be applied to the sample to avoid charge build-up from

the electron source beam that would skew the picture quality and EDS results [73].

Photos of the Nanoimage TableTop - SNE-4500M Plus SEM used during

experimentation are shown in Figure 2-8(a) with a picture of the sample stage during

calibration shown in Figure 2-8(b).

(a)

(b)

Figure 2-8: (a) Nanoimage TableTop - SNE-4500M Plus SEM utilized during experimentation and (b) image of SEM sample stage during calibration with cement paste sample loaded.

2.3.7 Sample Preparation for Nanoindentation and SEM/EDS

Sample preparation is key to ensure that the nanoindentation results follow the analysis assumption that the sample surface is perfectly smooth [50] [51], and that the electrons could successfully reflect back to the detector. The typical procedure for sample preparation commonly reported in several studies was followed in this study as well. First, the cement paste sample was cut to a certain height so that the sample was easy to polish and would fit into the testing equipment. Next the sample surface was sanded by hand or by polishing machine. Polishing was completed in ascending grit size from 400 grit to 1200 grit, then 9 μm down to 1 μm using a combination of silicon carbide papers and diamond particles suspended in oil [50]. Between grit sizes, the samples were ultrasonicated in a bath of isopropyl alcohol (IPA) to clean any lingering cement particles that could potentially scratch the surface in subsequence polishing. Figure 2-9(a) features the glassy, polished surface of a UHP binder sample post-polishing. Cement hydration was suspended by storing the sample in IPA at all times [74]. Storage in IPA ensures that the cement sample did not continue to hydrate, which would increase the surface roughness and change the volume fraction of hydration products in the sample. Minimizing the topography of the sample surface and maintaining a slope less than or equal to 3% in all directions will prevent damage to the nanoindenter tip and fulfill the analysis assumption of a perfectly smooth surface. Evaluation of the sample surface was performed qualitatively by visually noting the reflectivity and the presence of scratches on the surface. The more reflective the surface and the fewer the visible scratches, the more acceptable the surface is to place a nanoindentation grid. Quantitively, the root mean square roughness number (RMS) was determined using atomic force microscopy

(AFM). An important caveat is that RMS is highly dependent on the size of the area

being scanned, such that the RMS of a square of side length 10 μm will be very different

from the RMS of a square of side length 5 μm scanned over the same area [51].

Using microscopes with 5x, 20x, and 50x magnifications, the location of the

nanoindentation grid placement was determined. It was necessary to create physical

indicators on the sample surface, such as a scratch, to indicate the grid location. Without

a proper map of measurements and landmarks on the surface, locating the

nanoindentation points again in the SEM/EDS would be nearly impossible. Finally, to

prevent static charge on the surface of the non-conductive cementitious surface, AuPd

sputter coating was applied prior to scanning in the SEM/EDS. Figure 2-9(b) features a

typical cement paste sample after polishing, grid locating, and sputter coating.
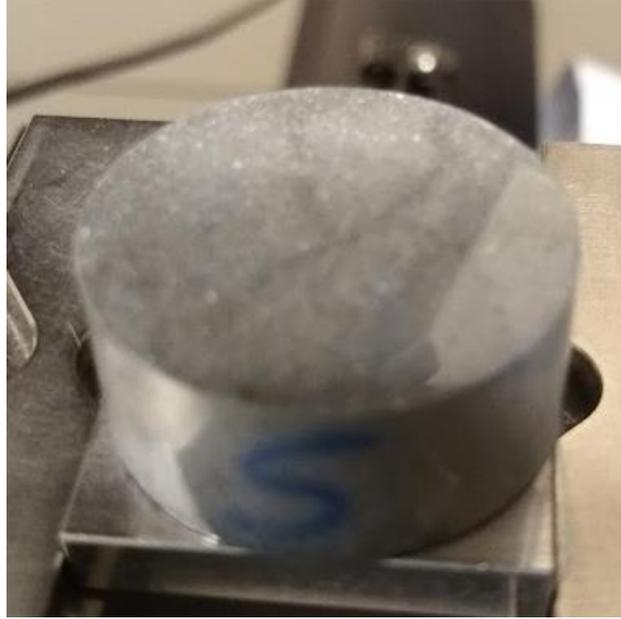
(a)
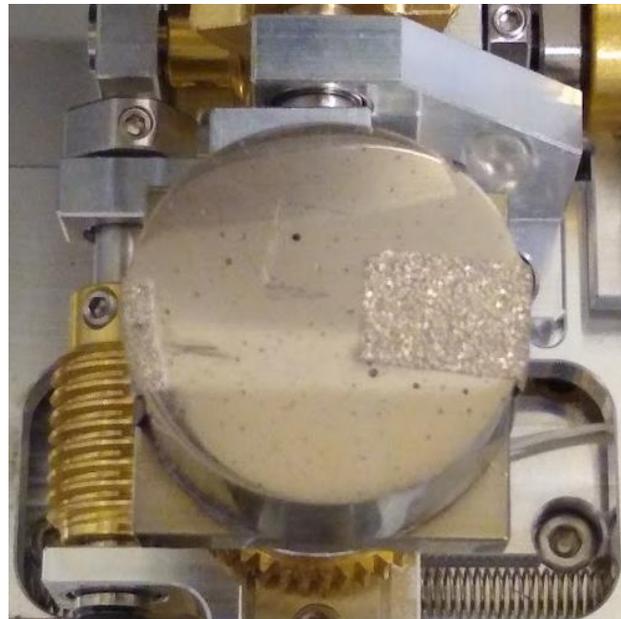
(b)

Figure 2-9: UHP paste sample: (a) after polishing at 1200 grit and ultrasonication, and (b) after marking grid location, carbon taping to the sample holder, and AuPd sputter coating.

2.3.8 Atomic Force Microscopy (AFM)

Following surface preparation procedures, surface roughness was measured using Atomic Force Microscopy (AFM). AFM is a surface measurement technique most commonly used to map the topology or calculate the interatomic forces of a sample [75]. In AFM, a laser is reflected off a cantilever tip that is in physical contact with the sample (contact mode) or vibrating at a set frequency just above the sample surface (tapping mode). Given the mechanical properties of the cantilever and the feedback from the laser, a 3D topographical map of the surface can be generated to measure the average roughness of the sample surface. Past cementitious studies have employed AFM prior to nanoindentation [50, 76, 77] to quantify the root mean square roughness (S_q) and ensure that depth of the indentation will be at least 5 times greater than S_q [78]. This roughness criterion, first proposed in [76], assures that the Hertzian contact (two smooth spherical surfaces with different radii and elastic constants in contact) assumed by nanoindentation analysis holds true [65] [66]. AFM in dynamic (tapping) mode was performed in this study using a Nanosurf Flex AFM inside of vibration dampening container, pictured in Figure 2-10(a). The pyramidal silicon tip, pictured on a UHP sample surface in Figure 2-10(b), was calibrated using the Sader method [79].
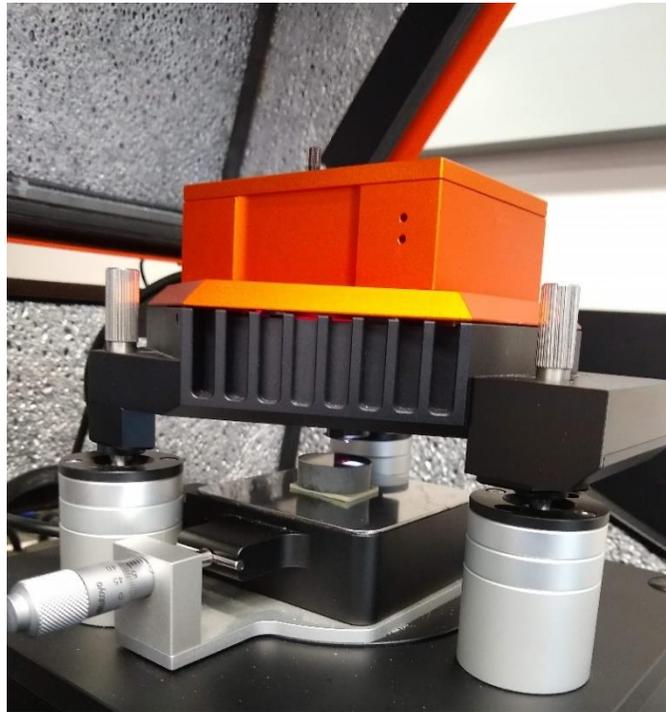
(a)
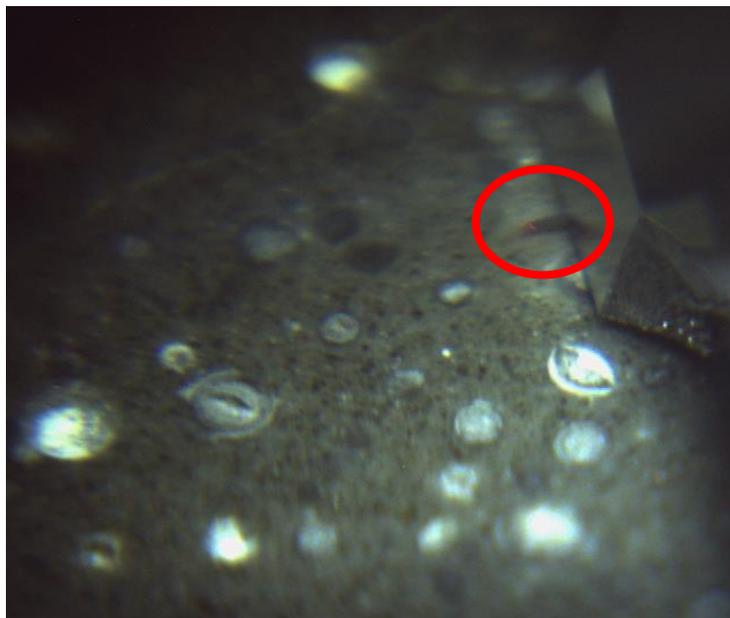
(b)

Figure 2-10: (a) Nanosurf Flex AFM with a cement paste sample (b) Microscope side-view of the AFM tip (circled in red) on a UHP paste polished surface at 10x magnification (field view is 5 x 3 mm).

2.4 MACHINE LEARNING

2.4.1 Background

Machine learning (ML) is a set of mathematical and computational techniques to analyze a large set of data, gaining new insights into patterns, predicting outcomes, and optimizing design parameters [8]. Empirically based, ML depends on a large set of reliable data instead of modeling underlying physical mechanisms to make predictions and decisions. In cases where large datasets are not available or are costly to acquire, the machine learning technique of transfer learning can be used to train an algorithm originally on a source set with a sufficiently large number of data points, and later transfer the algorithm to a target set of data with less points [80] [81]. Machine learning approaches are divided into three board categories depending on the type of information available and the desired output. Supervised ML is when both input and output data is available and the algorithm is being trained to make a prediction based on all the available information [82]. Unsupervised learning is used when the goal is to discover patterns and relationships between the input variables, often taking the form of clustering data into groups [15]. Finally, reinforcement learning is a ML operating in a dynamic environment seeking to perform operations that maximize a "reward" signal, reinforcing that behavior [83]. Examples of RL applications include autonomous vehicles learning how to drive or spelling auto-correction dictionaries.

Material science investigations into the process-structure-property-performance chain using machine learning dates to the early 2000s through the quantification, classification, optimization, evolution, and reconstruction of microstructures and mechanical properties [8] [84]. In recent years, materials-based machine learning efforts have focused on

linking images to microstructural properties, specifically through the generation and sampling of 2D [85] and 3D representative volume elements [86]. In the cementitious sciences, a number of recent studies have addressed the applications of machine learning as a viable prediction tool for the compressive strength of concrete given the input material and mixture proportions [11] [87] [88], including cements incorporating fly ash [9] [82] [89]. Machine learning has not been adequately applied to innovative concrete types such as UHPCs, self-healing concrete, geopolymers, etc. [90], which this study aims to accomplish through selected, focused utilization of ML to UHP binders. Some of the most common machine learning techniques available to study materials mechanics include support vector regression (SVR) [9, 14, 10], gaussian process regression (GPR) [12, 91], decision tree/forest ensembles [11] [87] [9], and artificial neural networks [11] [9] [82] [89]. The focus of this study will be on ensemble and ANN machine learning models, whose applications and limitations are detailed below.

2.4.2 Forest Ensembles

Forest ensembles are based on the architecture of a decision tree and follow a sequence of input feature splits to predict outputs [87] [9]. A basic form of forest ensemble is the Random Forest (RF), in which the best split of the data into branches and nodes is determined by considering all of the input features and checking a criterion, such as mean-squared error, to select the most discriminative threshold [92, 87]. Each individual decision tree in the RF ensemble does not use the entire set of training data, but a bootstrap sample made from subsets of the training data with replacement [87]. Another forest ensemble is the Extra Trees (ET) regressor in which the splits are drawn at random for each input feature and the best split, as measured by the chosen criteria, is selected as

36

the splitting rule [92, 87]. In the ET regression model, the entire dataset is incorporated

into each individual tree [92]. The prediction results of the individual trees are averaged

to produce the output prediction in the RF and ET regressions. Figure 2-11 features

flowcharts depicting the differences between the random forest (Figure 2-11(a)) and extra

tree (Figure 2-11(b)) ensemble methods. In a Gradient Boosted Tree (GBT) ensemble, an

initial tree is trained with the entire data set using all available features. All subsequent

trees in the forest are trained to minimize the residual (least squares error) between the

predicted and actual values of the previous tree via steepest gradient descent [9, 92]. The

final prediction is calculated as the weighted sum of the predictions of each tree, where

for each tree beyond the first, the prediction is multiplied by the learning rate, with

typical values between 0.01 and 0.1 [9, 92]. A specialized form of the GBT is Extreme

Gradient Boosted (XGB) tree [93]. XGB performs shrinkage and column subsampling

techniques to prevent overfitting between boosted trees and additionally offers scalability

through parallel tree boosting (efficient computing regardless of data size) [93].
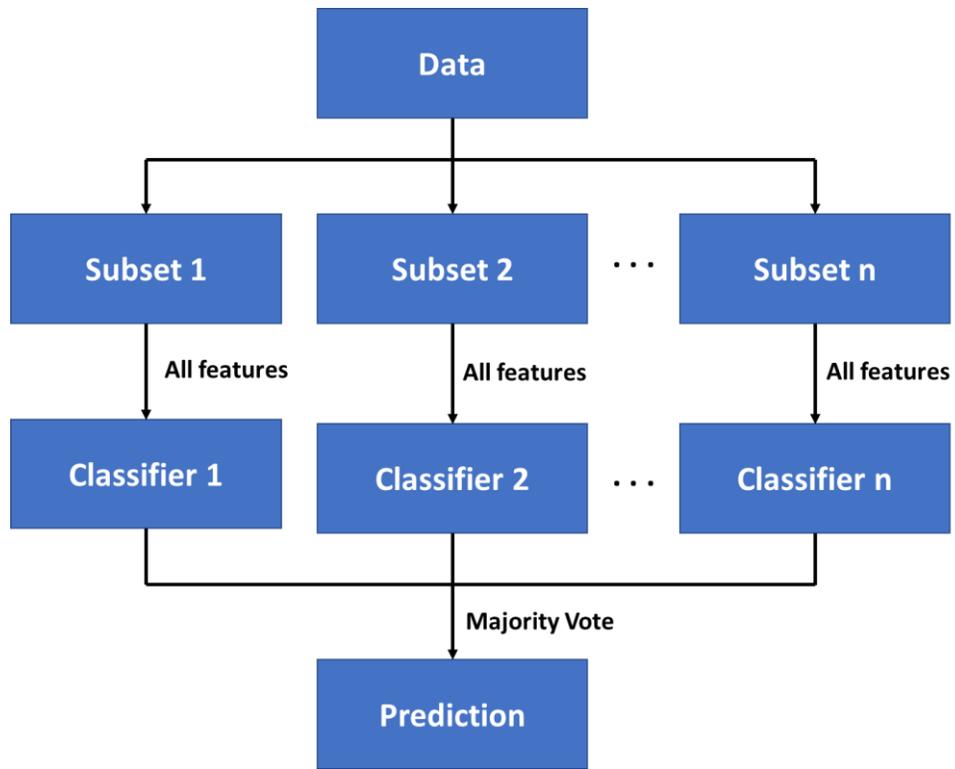
(a)
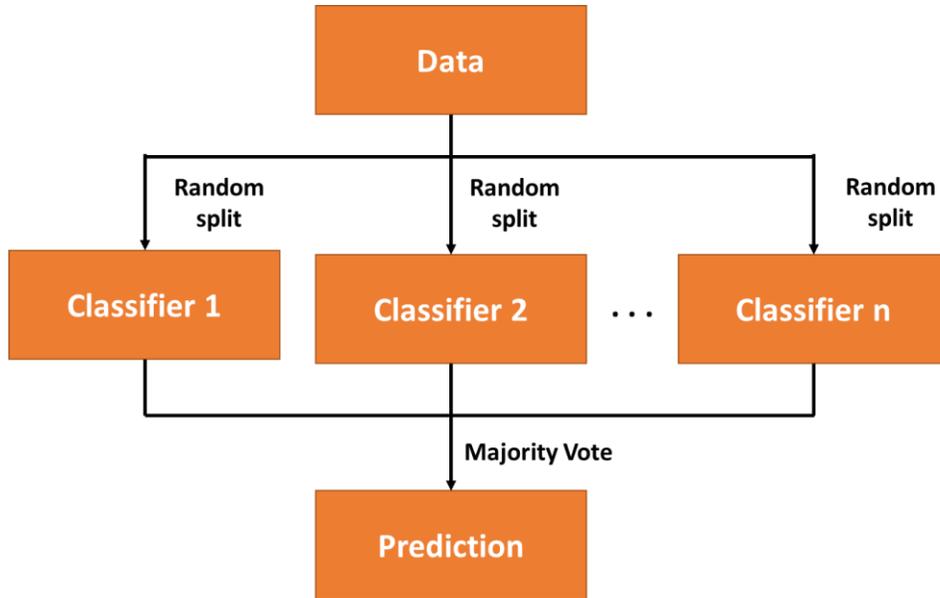
(b)

Figure 2-11: Flowcharts of: (a) Random forest (RF) tree ensemble, and (b) Extra Trees (ET) forest ensemble. Based on an illustration from [94].

2.4.3 Artificial Neural Networks (ANNs)

Artificial neural networks (ANN) mimic synaptic response in vertebrate brains and are organized into an input layer, hidden layer(s), and an output layer [95]. ANNs are some of the most popular among researchers for their ability to map complex non-linear relationships between multiple variables [8]. Between layers, neurons are fully connected to every other neuron. Each connection has a unique weight, w, relating the two neurons. The number of input layer neurons is equal to the number of inputs; the number of output layer neurons is equal to the number of outputs; and there can be any number of neurons in the hidden layer(s). An example schematic of a full-connected ANN is shown in Figure 2-12. The ANN can be shallow with only a few hidden layers, or deep with many layers. Generally, networks with > 10 hidden layers are considered deep networks [83]. The value within each neuron of the hidden layer(s) and output layer depends on the previous neurons, the weights, and the chosen activation function [95].

Utilization of the gradient of the previous iteration to train the weights of the ANN is known as backpropagation. Backpropagation is the backbone of weight optimization schemes, such as the RMSprop optimizer with an adaptive learning rate formula [96]. This learning rate is used to automatically adjust the initial learning rate for each parameter when generating the updated weights.

As indicated in the equations above, neural networks feature many fitting parameters that allow them to predict nonlinear interactions. A disadvantage of neural networks is the potential for over-fitting the data, or training the weights to precisely match the training data set and render the algorithm unable to accurately predict results of the test data set [9]. To minimize over-fitting, a dropout rate can be incorporated into the ANN [97]. A

dropout rate is a hyperparameter probability that any neuron and its connections will be temporarily excluded from the network [97]. While testing, the entire neural network is used, but the connection weights are multiplied by the dropout rate to combine the effect of the thinned-out training networks. More details on how ANNs are implemented to specific problems of interest can be found in Chapters 5, 6, and 7.
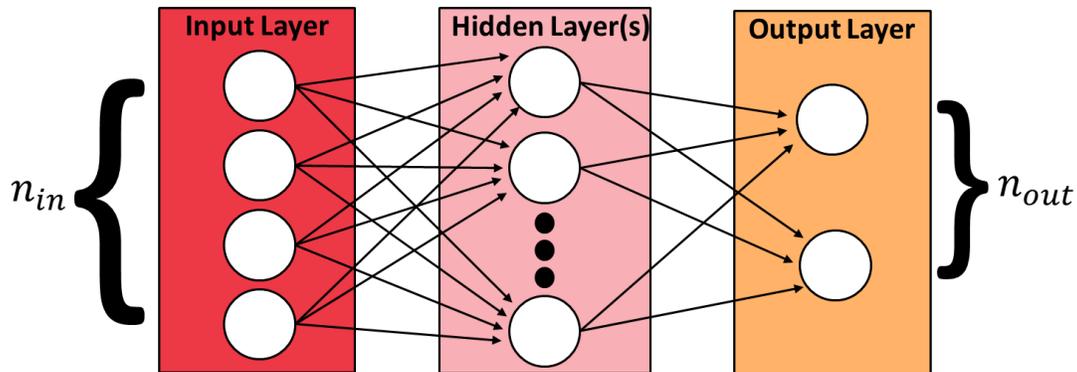
Figure 2-12: Schematic of a typical fully-connected feed-forward artificial neural network (ANN).

2.5 SUMMARY

This chapter reviewed key studies and state of the art work on ultra-high performance (UHP) binders, cementitious material microstructural characterization techniques, and machine learning (ML) algorithms for understanding the process-structure-property-performance chain in materials science. The different cement replacement materials used in this study, along with their effects on the microstructural development of UHP binders, were addressed. Sample preparation methods and equipment were discussed. Lastly, the principles and application of machine learning methods as applied to materials characterization were discussed.

CHAPTER 3[1]

ELUCIDATING THE NANO-MECHANICAL BEHAVIOR OF MULTI-

COMPONENT BINDERS FOR ULTRA-HIGH PERFORMANCE CONCRETE

3.1 INTRODUCTION

Ultra-high performance concrete (UHPC) is becoming more prominent in infrastructural

construction because of its wide-ranging benefits: high compressive and flexural strengths

result in smaller cross-sections and material savings; improved ductility enables efficient

structural design; and enhanced durability provides beneficial attributes to service-life and

sustainability [18]. UHPC binders employ a very low water-to-binder ratio (w/b) and a

high volume of reactive powders to ensure efficient space filling and porosity reduction

[18]. The attributes that contribute to a denser and stiffer microstructure are responsible for

its enhanced strength and durability. The demand for a denser microstructure has

traditionally resulted in the use of very high amounts of cementing materials and fine fillers

(including highly reactive supplementary cementing materials such as microsilica and

metakaolin, and expensive fine fillers like ground quartz) [19] [98] [99] [20], thereby

increasing the cost of UHPC and impacting large-scale adoption. Several recent efforts

have focused on tailoring the UHPC microstructure with commonly available cement

replacement materials and fine fillers in order to make it more cost-effective, sustainable,

and durable [21] [100] [36] [41] [101]. The use of coarse aggregates is not common in

UHPC, but recent studies have reported the development of UHPC with coarse aggregates

to counter issues relating to volumetric stability, given the high paste content in UHPC [41]

---

[1] This chapter is derived from the publication: E. L. Ford, A. Arora, B. Mobasher, C. G. Hoover and N. Neithalath, "Elucidating the nano-mechanical behavior of multi-component binders for ultra-high performance concrete," *Construction and Building Materials,* vol. 243, p. 118214, 2020.

[102] [103]. The understanding that physical particle packing effects of binding materials and fillers/aggregates are as important (or even more important) as the ultimate degrees of reactivity has resulted in scientific binder selection methods that comprehensively account for microstructural packing using a stochastic packing model with periodic boundary conditions [18]. This ensures a dense microstructure even at a low water content that does not favor optimal reaction product formation. The microstructure of UHPCs thus formed are highly heterogeneous, with several reaction products and unreacted particles. The UHP binders that are discussed in this study, which have been shown to be significantly cheaper compared to proprietary mixtures and other available alternatives [41], utilize cement replacement levels of 30% or 50% (by mass), with the replacement materials being fly ash, microsilica, and fine limestone powder. This leads to a highly heterogeneous, multi-phase microstructure.

Statistical nanoindentation is a well-accepted technique to probe microvolumes of cementitious materials to determine the intrinsic mechanical properties of the different phases present and their volume fractions [21] [59] [50], including UHPCs [51]. The use of micromechanics-based models on nanoindentation data coupled with colloid-based models for morphological arrangement of C-S-H have also been used to identify the local packing densities (and thus the porosities) of the different C-S-H phases [57] [104]. Nanoindentation coupled with chemical mapping has also been used to relate the chemical constitution of the phases to their mechanical properties [69] [70] [105]. The use of nanoindentation to determine the influence of cement replacement materials such as fly ash, slag, and silica fume on the nanoscale response of pastes has also been reported [58] [59] [60] [61] [62] [63]. In this paper, statistical nanoindentation is employed on two UHPC

pastes with high volume cement replacement by multiple, commonly available materials. The presence of multiple/mixed phases attributable to different starting materials with differing reactivities render the matrix highly heterogeneous and complex to analyze. Thus, a comprehensive understanding of the micro-/nano-structure of such UHPCs require elucidation of the following aspects, which is carried out in detail in this paper: (i) the influence of a combination of very low w/b and high volume of cement replacement materials (including reactive fillers) on the nanomechanical properties of the constituent phases and their evolution, (ii) the presence of multiple mixed phases including an ultra-high stiffness (UHS) phase that is observed to be dominant in low w/b systems, and (iii) upscaling of the elastic modulus from nanomechanical properties and the influence of material heterogeneity on predictive capability.

3.2 EXPERIMENTAL PROGRAM

3.2.1   Materials and Mixtures

Type I/II ordinary Portland cement (OPC) conforming to ASTM C 150, Class F fly ash conforming to ASTM C 618, limestone powder conforming to ASTM C 568, and microsilica conforming to ASTM C 1240 were utilized in the preparation of ultra-high performance cementitious pastes. Limestone powders with two different median particle sizes (1.5 µm and 3.0 µm) were used to ensure efficient packing. The chemical composition of the source materials is shown in Table 3-1, and their particle size distributions, extracted from laser particle size analysis, in

Figure 3-1. A polycarboxylate ether (PCE)-based superplasticizer with a solids content of 43% was used to ensure workability of the very low w/b pastes.

Table 3-1: Chemical composition and specific gravity of the powder materials used in this study.

Components of the binder	Chemical composition (% by mass)							Specific Gravity
	SiO_2	Al_2O_3	Fe_2O_3	CaO	MgO	SO_3	LOI	(g/cm^3)
OPC	19.60	4.09	3.39	63.21	3.37	3.17	2.54	3.15
Fly Ash (F)	58.40	23.80	4.19	7.32	1.11	3.04	2.13	2.24
Microsilica (M)	> 90.0	-	-	< 1.0	-	-	-	2.18 [59]
Limestone (L), 1.5	> 97% $CaCO_3$							2.7 [106]
Limestone (L), 3								
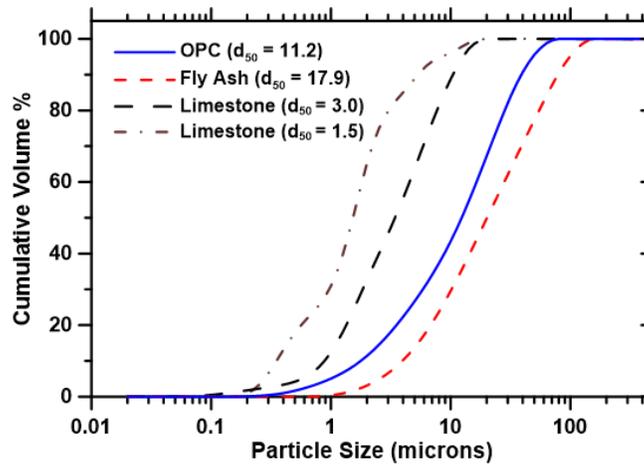
Figure 3-1: Particle size distribution (PSD) of the starting materials. The median size in microns is shown in parentheses. The PSD of microsilica is not shown, but the median size is $< 0.5 \ \mu m$ in a well-dispersed state.

Three different paste mixtures – two UHP pastes (w/b ~ 0.20) and a companion OPC paste (w/c = 0.40) were prepared. The binder contents in the UHP pastes were based on previous work of the authors [18] [41] that proportioned a family of UHPCs based on optimal particle packing, rheology, and substantial clinker factor reduction. One quaternary binder with 17.5% fly ash, 7.5% silica fume, and 5% limestone (all mass-based) replacing OPC for an overall cement replacement level of 30% (termed as $F_{17.5}M_{7.5}L_5$), and a ternary binder with 20% microsilica and 30% limestone (mass-based) replacing cement (termed as

$M_{20}L_{30}$), for an overall cement replacement level of 50%, were adopted. Paste samples were prepared and cured under moist conditions until their respective testing durations.

3.2.2   Sample Preparation for Nanomechanical Studies

The cylindrical paste samples were cut into ~12.5 mm thick discs using a Bruker IsoMet 1000 saw with a diamond wafer blade after 30 ± 3 days or 90 ± 5 days of moist curing. Isopropyl alcohol (IPA) was used as a coolant for the saw to avoid further hydrating the cement samples [74]. Samples were sanded and polished successively using Buehler CarbiMet silicon carbide abrasive paper of grit sizes 50, 18.3, 10.6, 9, 3, and 1 μm to ensure desired smoothness for the nanoindentation testing [50] [51]. Between sanding at each grit size, the samples were ultrasonicated in a bath of IPA for 5 minutes to dislodge any trapped debris from the pores and microstructure. Samples were polished until mirror reflectivity was achieved. Following the final ultrasonication, the samples were stored in IPA until testing [74].

3.2.3   Nanoindentation

Nanoindentation grids were placed on the sample surface using an Ultra Nanoindentation Tester (UNHT[3]; Anton Paar) in areas that appeared to be smooth and free of visible surface defects when viewed under the microscope at 5x, 20x, and 50x magnifications. Each sample had at least 1250 indents, split among multiple grids in different locations, to ensure that the heterogeneous nature of the microstructure was adequately captured. The tests were performed in the force control mode with a maximum displacement cutoff of 250 nm (0.25 μm). This depth corresponded to an interaction volume idealized as a hemisphere with a radius of about 4 to 10 times the maximum cutoff, or 1.0 μm to 2.5 μm according to the estimates in [69] [107]. The distance between points in the grid was chosen as 5 μm to

45

ensure that the indents minimally influenced each other [50]. The linear loading profile had a loading and unloading rate of 12 mN/min with a pause for 5 seconds at the peak load when the maximum displacement cutoff was reached.

The hardness (*H*) and the effective Young's modulus (*M*), were extracted from each of the indents according to Oliver and Pharr method [65] [66]. The effective modulus (M) is a function of the elastic moduli and Poisson's ratios of the sample and the tip as shown below:

$$\frac{1}{M} = \frac{1 - v^2}{E} + \frac{1 - v_i^2}{E_i} \tag{7}$$

v is the Poisson's ratio of the sample, v_i is the Poisson's ratio of the diamond tip, which is equal to 0.07, and E_i is the Young's Modulus of the tip, equal to 1141 GPa [65] [66]. v of different phases was assigned after identifying them in statistical clustering [39]: clinker as 0.31, HD and LD C-S-H as 0.25, UHS phase as 0.29 (the average of C-S-H and CH), and mixed phases as 0.27 (between reactants and products). As part of post-processing, abnormal load-depth curves representing surface pores or partial material collapse were removed from the data set as described in [59].

3.2.4   Pore Structure and Thermogravimetric Analysis

The pore structure of the pastes was evaluated using Mercury Intrusion Porosimetry (MIP). For MIP experiments, small pieces of the paste were weighed and placed in the low-pressure chamber of the porosimeter (Quantachrome Instruments Pore Master). The sample was filled with mercury starting from ambient pressure to 345 kPa (60 psi). The sample was then placed in the high-pressure chamber and the applied pressure increased to

414 MPa (60,000 psi). The pore diameter, d, as a function of the intrusion pressure was obtained from the Washburn equation [42] as:

$$d = \frac{-4\sigma \cos(\theta)}{\Delta P} \qquad (8)$$

where ΔP is the difference in the pressure between successive steps (MPa), θ is the contact angle between mercury and the cylindrical pore, taken as 130° in this study, and σ is the surface tension between mercury and the pore walls, taken as 485 mN/m [43] [44] [45]. The critical pore size, indicative of the percolating pore size of the material, was obtained from the peak of the differential pore volume curve. The porosity of the sample, φ, was determined from the cumulative volume of mercury intruded and bulk density of the samples.

Thermogravimetric analysis was performed on the paste samples to determine the calcium hydroxide (CH) and calcium carbonate contents. The tests were performed in an inert N_2 environment at a gas flow rate of 20 ml/s. The samples were heated from ambient temperature to 900°C at a heating rate of 15°C/min.

3.3 RESULTS AND DISCUSSIONS

3.3.1 Pore structure and CH contents of UHP pastes

The cumulative volume of mercury intruded and the differential pore volumes as a function of pore diameter are plotted in Figure 3-2(a) and (b) for the OPC and UHP pastes evaluated in this study. The pore volumes and critical pore sizes for the OPC and UHP pastes after 30 days obtained from mercury intrusion porosimetry are shown in Table 3-2. The 30-day porosities of the UHP pastes, calculated from the total volume of mercury intruded and the cement paste bulk density of about 2 g/cm^3, was found to be in the 14-16% range, which

is about half of that of the 0.40 w/c control cement paste. The critical pore sizes are also lower for the UHP pastes, owing to the lower w/b and significantly improved particle packing through the use of ultrafine materials.

Table 3-2: Pore volume and critical pore sizes of UHP pastes after 30 days of hydration

Mixture	MIP Porosity (%)	Critical pore size (μm)
OPC (w/c = 0.40)	28.22	0.026
$F_{17.5}M_{7.5}L_5$	14.36	0.019
$M_{20}L_{30}$	15.21	0.014
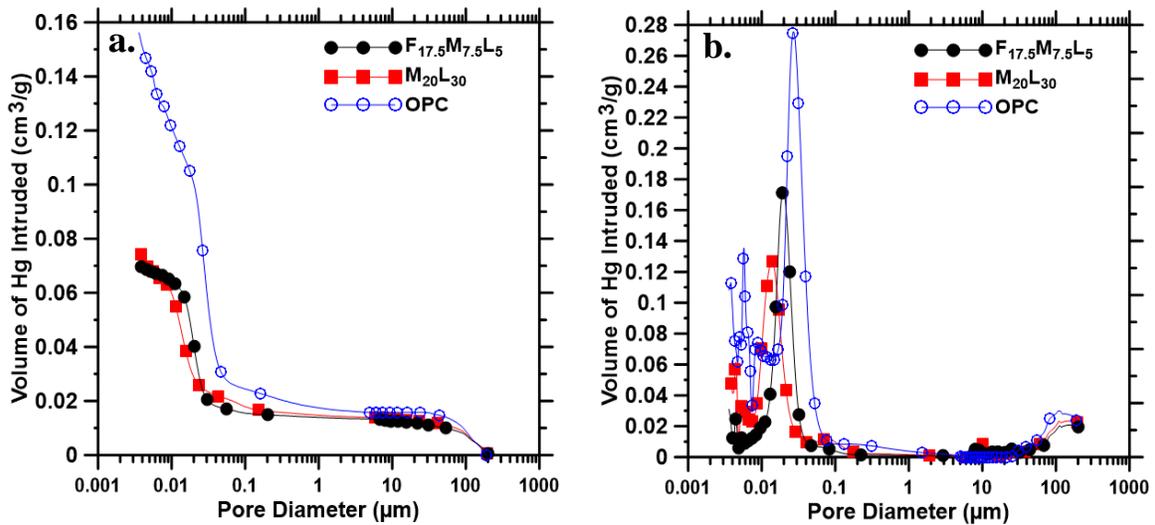
Figure 3-2: (a) Cumulative and (b) Differential pore volume curves after 30 days of hydration for the OPC, $F_{17.5}M_{7.5}L_5$, and $M_{20}L_{30}$ pastes

Thermogravimetric (TG) and differential thermogravimetric (DTG) traces of the OPC and UHP pastes after 30 days of hydration are shown in Figure 3-3. The major peak in the 100-120°C range corresponds to the loss of evaporable water, the one in the 430-500°C range corresponds to the dehydration of calcium hydroxide, and the one in the 750-850°C range corresponds to the decarbonation of calcium carbonate [18] [36] [46]. The carbonate peaks accurately identify the amount of added limestone in the pastes. Even with very high

dosages of cement replacement materials (especially for the $M_{20}L_{30}$ paste), there was remnant CH in the pastes, once again primarily attributed to the low w/b in these systems. The volume fractions of CH were found to be 0.14, 0.058, and 0.047 for the OPC, $F_{17.5}M_{7.5}L_5$, and $M_{20}L_{30}$ pastes respectively.
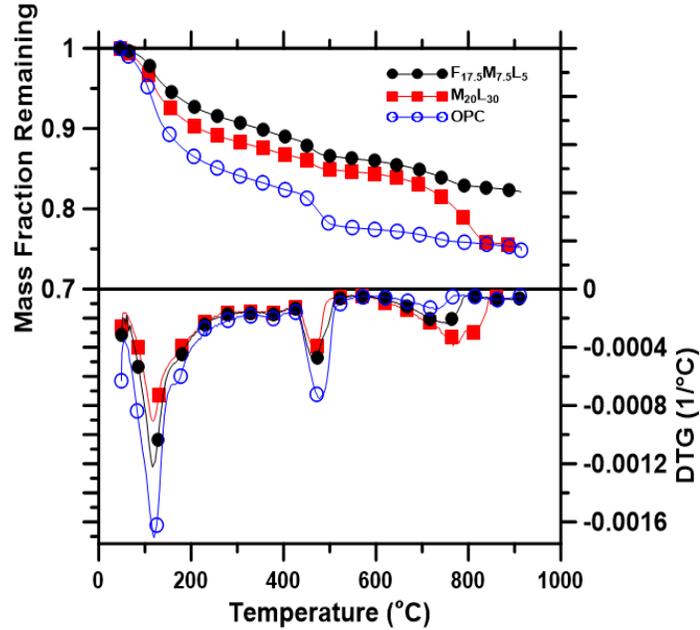
Figure 3-3: TGA and DTG traces of the OPC and UHP pastes at 30 days

3.3.2 Frequency Histogram of Elastic Properties from Nanoindentation and Deconvolution Methods

A representative frequency histogram for elastic moduli of different phases in the 30-day cured $F_{17.5}M_{7.5}L_5$ paste is shown in Figure 3-4(a). If there exists n phases in the microstructure with each phase occupying a volume fraction of f_i (i = 1…n) such that $\sum_{i=1}^{n} f_i = 1$, the properties of each phase can be approximated by a Gaussian distribution with a probability density function (PDF) given as:

$$PDF = \sum_{i=1}^{n} f_i \, \psi_i \tag{9}$$

49

Here, ψ_i is the property of interest of the phase. It has been shown that the same statistical nanoindentation results can be fit using different numbers of phases and volume fractions [108]. In this study, a Bayesian Information Criterion (BIC) with negative log likelihood method was implemented for statistical deconvolution. Minimizing the BIC determines the best number of phases to include [109]:

$$BIC = 2 * NlogL + p * \log(num) \tag{10}$$

Here, num is the number of indentation points, p is the number of identifying parameters available at each indentation point (e.g., M and H), and NlogL is the maximum negative log likelihood property, which is defined as:

$$NlogL = -\max(\log(\textstyle\prod_n PDF(n_i))) \tag{11}$$

Here, n_i represents the distribution parameters, in the case of a Gaussian distribution the mean and standard deviation, that are iterated to maximize the likelihood function. The maximum negative log likelihood estimation was used to find the Gaussian normal probability distribution functions that best represent the experimental data. The smaller the resulting BIC, the better the data was modeled by the tested number of phases. For each input data set, the number of phases that resulted in the smallest BIC varied between 3 and 5. For example, for the 30-day hydrated $F_{17.5}M_{7.5}L_5$ paste, the use of five phases resulted in the smallest BIC. The elastic modulus histogram overlaid with the optimal five Gaussian distributions generated from the BIC method for the 30-day hydrated $F_{17.5}M_{7.5}L_5$ paste are shown in Figure 3-4(b). Occasionally the clustering algorithm yielded clusters that were embedded inside of one another, or unreasonably split the high hardness-stiffness values into multiple clusters. In such cases, the clusters were manually grouped together, and the

new average stiffness and hardness determined. An iterative Gaussian fitting algorithm [110] [111] was also used to statistically deconvolute the data by pre-defining the number of phases, however the BIC-negative log likelihood method produced smaller errors and thus was chosen to analyze the indentation data in this study.
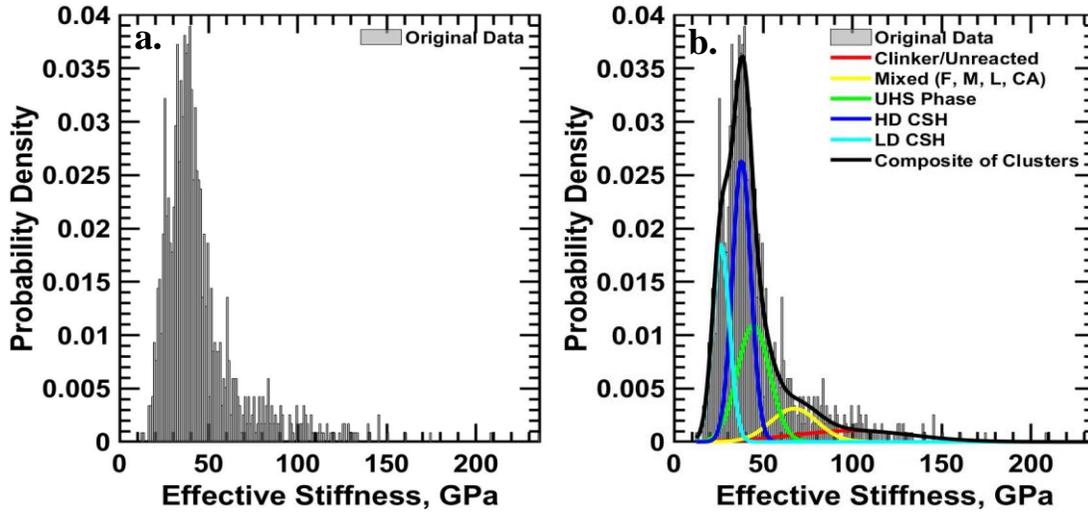
Figure 3-4: (a) Frequency histogram of stiffness in the $F_{17.5}M_{7.5}L_5$ paste after 30 days, and (b) fits of BIC with maximum log likelihood scheme with 5 phases on the experimental indentation moduli

3.3.3 Modulus-hardness Clusters from Nanoindentation and Insights into Nanomechanical Response

Based on the results of the previous section, M-H clusters obtained from BIC and negative log likelihood statistical analysis for the companion OPC paste and the different UHP pastes are presented here. Table 3-3 lists the mean and standard deviation for the hardness and stiffness of each cluster identified as well as their corresponding volume fraction. Due to the presence of multiple binders and fillers used in these UHP pastes and a very low w/b, the matrices exhibit significant heterogeneity. In general, the reaction products in the UHP pastes also have (M, H) ≤ (65.0, 3.0) as reported elsewhere on cementitious systems [63]

51

[56] [52] even though the distribution generally tends towards higher values. This could be attributed to several reasons, including the increased density of reaction products and the presence of multiple stiffer phases that remain unreacted under conditions of low w/b. These attributes are is discussed in detail in the forthcoming sections.

Figure 3-5 shows the M-H clusters for the 30-day and 90-day hydrated companion OPC paste (w/c = 0.40). As expected, there are four phases, corresponding to low density (LD) C-S-H, high density (HD) C-S-H, CH, and unhydrated clinker grains. The volume fractions of individual phases are also shown in the stacked bar next to the cluster plot. With increasing curing duration, the unhydrated cement fraction reduces in volume along with a slight increase in the HD C-S-H phase fraction.
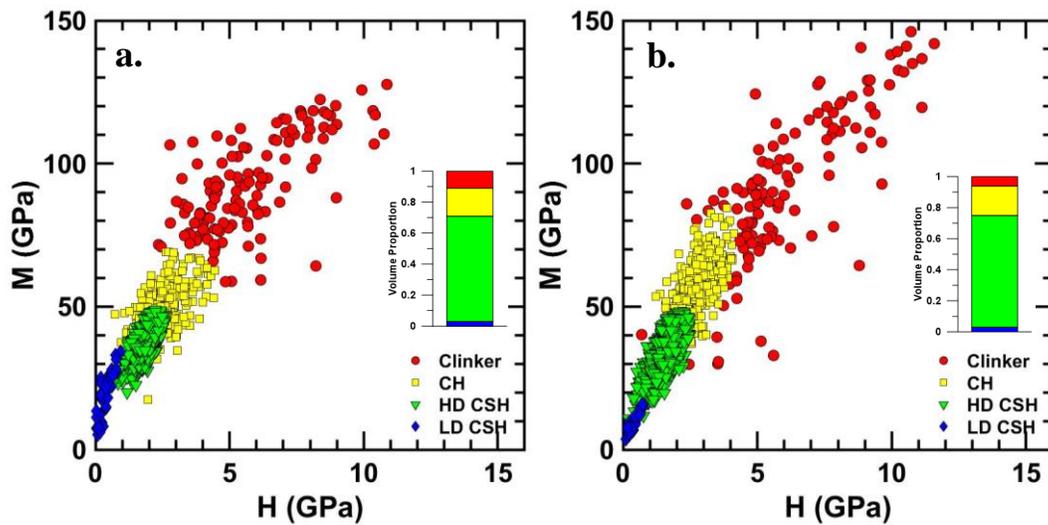
Figure 3-5: M-H cluster for companion OPC paste after: (a) 30 days, and (b) 90 days of curing

Table 3-3: Elastic properties (M, H) and volume fractions (f) of the different phases in the UHP pastes. F, M, L, and CA denotes fly ash, microsilica, limestone, and carboaluminates respectively.

UHP paste	Phase	Extracted values from nanoindentation results						Reference values	
		M (GPa)		H (GPa)		f		M (GPa)	H (GPa)
		30d	90d	30d	90d	30d	90d		
OPC	LD C-S-H	19.65±8.18	9.10±3.37**	0.40±0.27	0.39±0.18	0.03	0.03	18.2 [112] 22.89 [53]	0.45 [112] 0.93 [53]
	HD C-S-H	36.85±5.05	33.37±6.54	1.80±0.33	1.55±0.38	0.68	0.72	29.1 [112] 33.65 [62]	1.36 [50] 1.04 [62]
	CH	49.52±9.32	56.15±9.94	2.54±0.76	2.69±0.60	0.18	0.19	39.77-44.89 [113]	1.31 [112]
	Clinker	90.48±18.63	87.44±28.50	5.62±2.03	5.72±2.45	0.11	0.06	100.3 [55] 122.2 [53]	7.86 [55] 6.67 [53]
$F_{17.5}M_{7.5}L_5$	LD C-S-H	26.32±4.78	-	1.39±0.36	-	0.22	-	18.2 [112] 22.89 [53]	0.45 [112] 0.93 [53]
	HD C-S-H	37.89±5.20	-	1.66±0.28	-	0.34	-	29.1 [112] 33.65 [62]	1.36 [50] 1.04 [62]
	UHS Mixed (F, M, L, CA)	44.47±9.01 67.19±13.06	49.79±8.69 62.73±10.25	2.15±0.59 3.97±1.06	2.60±0.59 3.84±0.78	0.25 0.10	0.17 0.62	42.8 [57] 75.1 (F) [58] 72.8 (M) [39] 83.8 (L) [114] 51-80 (CA) [38]	1.43 [57] 8.47 (F) [58] 6.0 (M) [65] 1.98 (L) [115]
	Clinker/ Unreacted	100.21±32.80	101.40±28.14	8.12±3.20	9.38±2.68	0.09	0.21	100.3 [55] 122.2 [53]	7.86 [55] 6.67 [53]
$M_{20}L_{30}$	UHS/C-S-H	44.11±5.06	40.31±6.99	1.85±0.26	1.85±0.48	0.59	0.77	42.8 [57]	1.43 [57]
	Mixed (M, L)	64.10±4.97 (M) 66.29±8.36 (L)	57.82±11.08	4.48±0.37 (M) 3.73±0.49 (L)	2.75±0.79	0.14 0.15	0.16	72.8 (M) [39] 83.8 (L) [114]	6.0 (M) [65] 1.98 (L) [115]
	Clinker/ Unreacted	87.42±25.13	96.29±25.46	5.83±2.81	6.21±2.62	0.12	0.07	100.3 [55] 122.2 [53]	7.86 [55] 6.67 [53]

**Could be an experimental anomaly where the load-displacement curves for a few indents predominantly on pores were not identified or removed by the algorithm.

3.3.4   Fly ash-limestone-microsilica UHP Paste

Figure 3-6 (a) and (b) depict the M-H cluster data obtained from the nanoindentation experiments for the 30- and 90-day cured quaternary blend UHP pastes containing fly ash, limestone, and microsilica. The 30-day cured paste demonstrates the presence of five distinct phases based on M and H, while there are only three phases in the 90-day cured paste. Most of the experimental indentation moduli of the different solid phases fall in the range of 20-130 GPa, as is commonly reported for cementitious materials [51] [56] [52] [63]. The hardness values of the constituent phases are somewhat higher than those reported for conventional cementitious materials, the reasons for which can be found in the discussions below.

For the 30-day cured paste, at the lower end of the modulus spectrum, three different phases are observed, with mean values of M at 26.3 GPa, 37.9 GPa, and 44.5 GPa respectively. Based on published literature [50] [51] [53], these can be assigned to low-density (LD) C-S-H, high-density (HD) C-S-H, and an ultra-high stiffness (UHS) phase respectively. In general, the material properties of LD and HD C-S-H are reported to be independent of the mixture proportions [50] [51] [53], as was noticed in this study also. The UHS phase (M ~ 45-48 GPa, H ~ 2.0 GPa) is reported in literature either as C-S-H with a very high packing density (~0.84) [57] [116] or as an intimate nanocomposite of C-S-H and nanoscale CH [69] especially in low w/b binders such as the ones used in this work (see a later section for a more detailed analysis of the UHS phase). The fact that fine limestone filler (d_{50} of 1.5 μm or 3 μm) and microsilica ($d_{50} <$ 0.5 μm in a well-dispersed state) fills the interstitial spaces between the cement and fly ash particles in these mixtures likely enhances the formation of such a composite C-S-H with nanoscale CH phase due to the significant

54

reduction in available capillary spaces that would otherwise enable CH precipitation as microcrystals. Indeed, the lower the w/b, the lower the amount of available water to react, thereby lowering the C/S ratio of C-S-H gel and increasing the elastic modulus and hardness of C-S-H [25] [28]. Similar to the UHS phase, CH from cement hydration has an indentation modulus in the 40 to 45 GPa range [39] [113]. As observed from the DTG curves in Figure 3-3, CH was present in this UHP paste, however, the volume fraction of CH observed was around 0.05. This is significantly lower than the volume fraction of UHS phase, and thus the assumption that the UHS phase is a composite of C-S-H and CH is likely valid. Specific chemical information at each indentation point can help accurate identification of the composition of UHS phases in multi-component binder systems, which is the focus of an ongoing work.

The mixed phase is a combination of multiple phases with a mean M of 67.2 ± 13.1 GPa. In highly packed systems with multiple blends and a variety of fine particulates that react at varying degrees, it is not uncommon to encounter such stiff and hard phases [69]. The carboaluminates formed through the reaction between carbonates from limestone and alumina from fly ash [117] [118] [119] [35] have a reported moduli of 51-80 GPa [38]. Based on previous work, about 20% of limestone reacts to form carboaluminates in the presence of alumina from fly ash [33]. Unreacted fly ash particles, microsilica, and limestone used as a filler in UHP pastes have moduli in the 75-80 GPa range [58] [39] [114]. It is also possible that some phase interfaces are also captured in the mixed phase. Thus, the mixed phase could consist of unreacted fly ash, limestone, microsilica, carboaluminates and/or the interfaces between them. Such a possibility, that the phases detected in the M-H

plots are a composite response of multiple phases, has been put forth in [21] [70] [62] [110] also, and is more significant for multi-component blends displaying higher heterogeneity. The grouping at the high end of the M and H spectrum can be attributed to unhydrated clinker. The mean stiffness and hardness of this phase are higher than those commonly reported for clinker grains, and the spread also is much larger (compared to the spread for the OPC pastes shown in Figure 3-5). This could be because there are several high hardness/modulus phases in fly ash (such as hematite and mullite; XRD spectra revealed the presence of mullite (M ≈ 220 GPa) in this fly ash). It has also been shown that accurate identification of high modulus phases in a matrix requires a modulus mismatch ratio (the ratio between the stiffness of the matrix (M_m) and the modulus of the indented microstructure (M_i)), M_m/M_i, between 0.2 and 5 for indentation depths smaller than or equal to 10% of the characteristic length scale of the microstructure [21] [110] [63]. In many conventional cementitious pastes with "softer" matrices, this condition might not be satisfied, and high modulus ratio phases could be ignored [21]. A stiff UHP matrix allows the identification of these phases, resulting in values of mean indentation modulus and hardness of the unreacted phases that are higher than sometimes reported.

After 90 days of hydration, the LD and HD C-S-H phases are not detected in this paste. Instead, a UHS phase is mainly observed with a mean modulus of 49.8 GPa. Enhanced hydration in very low w/b systems is reported to favor formation of an UHS phase [57], which is in line with the observation here. The pozzolanic reaction of fly ash/microsilica that results in C-S-H gel with a lower C/S ratio could also contribute to this observation. Lowering the C/S ratio of C-S-H gel, which is a result of increased silica polymerization, is noted to increase the elastic modulus and hardness [25] [28] [120].

56

Between the 30- and 90-day cured specimens, the number of clusters dropped from five to three, with a large volume fraction being occupied by the mixed phases. As reaction progresses in such multi-component blends through pozzolanic reaction of fly ash and silica fume and the preferential deposition of products closer to reacting surfaces in densely packed systems, this can be expected. A closer look at the M-H cluster of the 30- and 90-day cured mixtures also suggests that for M < 65 GPa, about 10%-25% of the reaction products show H > 3 GPa, indicating that a multi-phase response is indeed acquired. The increased volume fraction of this phase at later ages can also be attributed to this reason. The cluster with (M, H) of (101.4, 9.38) can be attributed to unreacted clinker and other stiffer/harder phases. The hardness of different phases in the fly ash-based UHP matrix discussed above are higher than those reported for conventional cementitious pastes, which is contributed by the lower w/b in the UHP paste mixtures and corroborated by a stronger matrix [21] [63] [110], since hardness correlates directly with compressive strength.
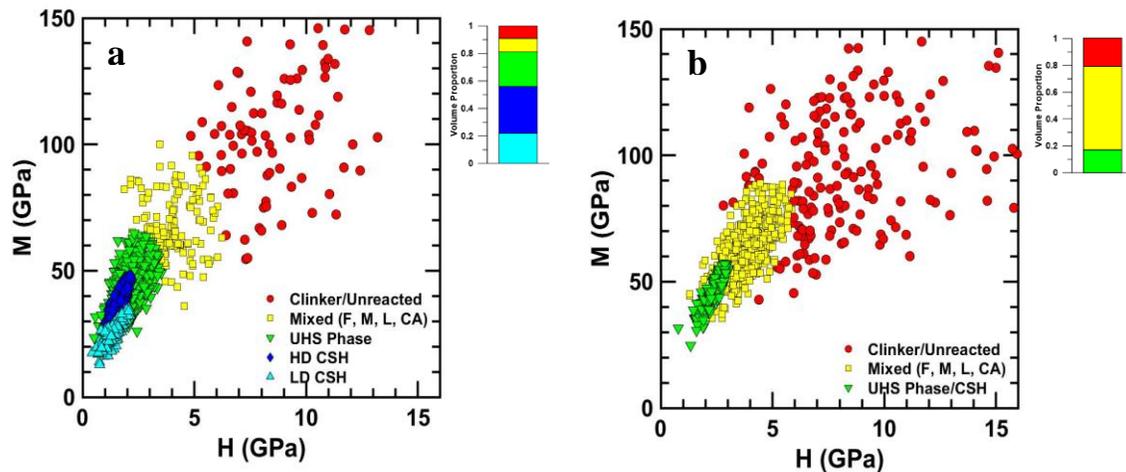
Figure 3-6: M-H clusters of fly ash-limestone-microsilica UHP paste after: (a) 30 days, and (b) 90 days of reaction. F, M, L, and CA denotes fly ash, microsilica, limestone, and carboaluminates. Since this is likely a mixed phase, with accurate identification difficult, the foregoing abbreviation is used.

3.3.5 Microsilica-limestone UHP Paste

The reaction products in microsilica-limestone UHP paste (with a 50% cement replacement) are fundamentally different from that of the fly ash-based system discussed earlier. The M-H clusters of this paste after 30 and 90 days of hydration are shown in Figure 3-7(a) and (b) respectively. A UHS phase is noticed at both ages, with M ~ 42±6 GPa, and H ~ 2.0 GPa, and the LD/HD C-S-H phases are absent as independent phases. It is known that the incorporation of microsilica or nanosilica increases the volume fraction of high-modulus C-S-H phases due to enhanced pozzolanic reaction and consequent reduction of the C/S ratio [121]. The high microsilica content and the high overall cement replacement level could have contributed to denser reaction products even at earlier ages, resulting in high M and H values, and the high volume fraction of this phase.

In the 30-day paste an interesting observation is that there are two distinct clusters with similar M values (~65 GPa), but different H values. The lower hardness phase called Mixed (L) phase likely corresponds to the mixed phase of limestone powder (which constitutes 30% of the starting materials) and some hydrates, while the higher hardness phase called Mixed (M) can be attributed to the mixed phase of unreacted microsilica and some hydrates. Limestone and microsilica have similar moduli but different hardness as shown in

3. These clusters could have been also combined without any appreciable loss of
accuracy. The size of limestone powder (d_{50} of 1.5 to 3 µm) and microsilica (d_{50} < 1 µm)
lends credence to the fact that the indentation measurements attributed to them are likely
to be those of mixed phases containing these materials and the hydrated products. After
90 days of hydration, the UHS phase has a volume fraction roughly equal to the sum of
the volume fractions of the UHS phase and the Mixed (M) phase in the 30-day sample.
This is an indication of the reaction of microsilica to further form low C/S C-S-H gel.
Lowering of C/S is reported to result in a pronounced increase in modulus and hardness
[25] [28], which can also be noticed in Figure 3-7 (b). The unhydrated cement phase
reduces in volume fraction from 30 to 90 days of hydration as expected. The Mixed (L,
M) phase with indentation modulus of ~60 GPa, attributed to limestone powder and some
hydrates, is also present in this system at a similar volume fraction as Mixed (L) in the
30-day paste. This points to further consumption of microsilica with time.
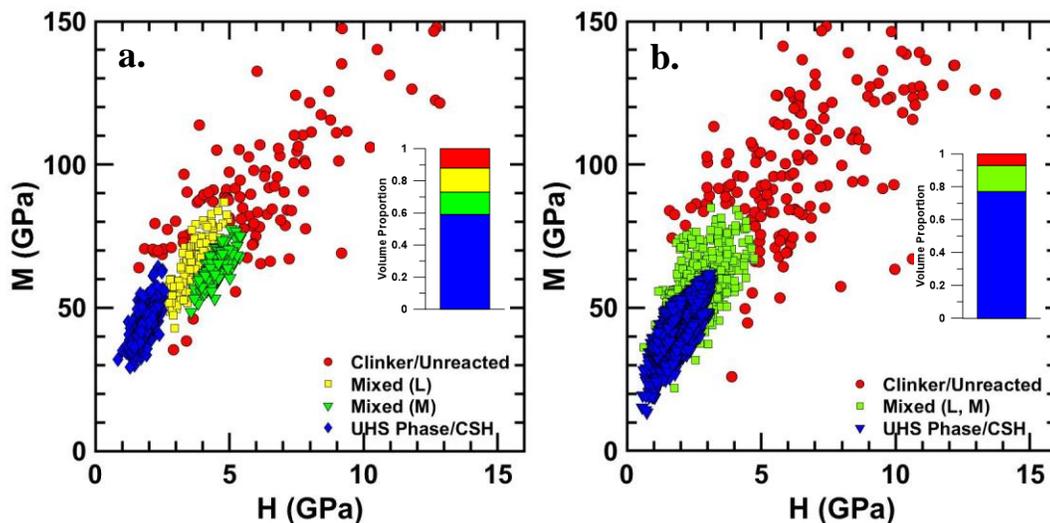
Figure 3-7: M-H clusters of microsilica-limestone UHP paste after: (a) 30 days, and (b)
90 days.

59

3.3.6 Multi-Step Homogenization for Macroscale Elastic Property Prediction from Nanoscale Response

The use of nanomechanical properties of individual phases in homogenization models to predict the macroscale response of composites has been carried out for normal and high-performance concretes [50] [122]. The analytical homogenization models are based on Eshelby's solution of a strain localization tensor for ellipsoidal inclusions embedded in a matrix [123] [124]. This method has been applied regularly in homogenization schemes of cementitious materials [123] [124] [125] [126] [127]. The shear and bulk moduli of the individual phases are obtained from their elastic moduli and Poisson's ratios [123] [126] . In this study the bulk and shear moduli were determined based on the nanoindentation cluster data, converting the indentation modulus to Young's modulus as shown in Equation 3 with the following Poisson's ratios assigned to each cluster based on the major component [39]: clinker as 0.31, HD and LD C-S-H as 0.25, and UHS phase as 0.29 (the average of C-S-H and CH). For the mixed phase where every indent is a composite response across multiple possible phases, the Poisson's ratio was taken as 0.27, which lies between the Poisson's ratios of the reactants and hydration products. The fine and coarse aggregate were considered to have a Poisson's ratio of 0.25 [128]. The homogenized bulk and shear moduli for two-phase materials can be quantified from the individual phase properties as shown in Equations 8 and 9.

$$K_{hom} = \frac{\sum_i f_i K_i \left(1 + A_r \left(\frac{K_i}{K_r} - 1\right)\right)^{-1}}{\sum_i f_i \left(1 + A_r \left(\frac{K_i}{K_r} - 1\right)\right)^{-1}} \qquad (12)$$

$$G_{hom} = \frac{\sum_i f_i G_i \left(1 + B_r \left(\frac{G_i}{G_r} - 1\right)\right)^{-1}}{\sum_i f_i \left(1 + B_r \left(\frac{G_i}{G_r} - 1\right)\right)^{-1}} \tag{13}$$

where f_i is the volume fraction of each inclusion phase, K_r and G_r represent the bulk moduli and shear moduli of the reference medium (the matrix), K_i and G_i represent the bulk moduli and shear moduli of the inclusion phases. The coefficients A_r and B_r are given as:

$$A_r = \frac{3K_r}{3K_r + 4G_r} \tag{14}$$

$$B_r = \frac{6K_r + 12G_r}{15K_r + 20G_r} \tag{15}$$

The double-inclusion method [129] [130] based on Eshelby's solution consists of an ellipsoidal inclusion of stiffness $\tilde{S}_{I-1}$ embedded in another ellipsoidal matrix of stiffness $\tilde{S}_{I-2}$, which is further embedded in an infinitely extended homogeneous medium of stiffness $\tilde{S}$. The double-inclusion method is capable of accounting for the inclusion-inclusion interactions in addition to the inclusion-matrix interactions considered in the Mori-Tanaka approach, and has also be implemented for cementitious materials [125] [131]. The average elastic moduli can be calculated as:

$$\bar{S} = \left[\tilde{S}^{-1} + f_1 \alpha + f_2 \beta\right]^{-1} \tag{16}$$

In the above equation, f_1 and f_2 are volume fractions of two inclusion phases, α and β are functions of f_1, f_2, and the Eshelby's tensors $\tilde{S}$, $\tilde{S}_{I-1}$, and $\tilde{S}_{I-2}$. Detailed derivation and analysis procedure are described in [129] [131]. For both Mori-Tanaka and double-inclusion methods, a multi-level nested homogenization approach is implemented as described below.

Previous studies have used a combination of techniques including MIP, X-ray tomography, and electron microscopy to identify the pore volumes that are of interest in predicting macroscale response from microscale observations [122] [125] [132]. In this work, considering the maximum indentation depth of 250 nm and a probed volume with a radius of ~1.5-2.5 μm, it is assumed that the effect of pores smaller than 1 μm would be accounted for in the deconvoluted stiffness of the phases. Hence, the pore volume corresponding to d > 1 μm from MIP is used in the homogenization process for the standalone void phase. The volume fractions of phases obtained from statistical nanoindentation, which is used in the homogenization process, are shown in Figure 3-8.
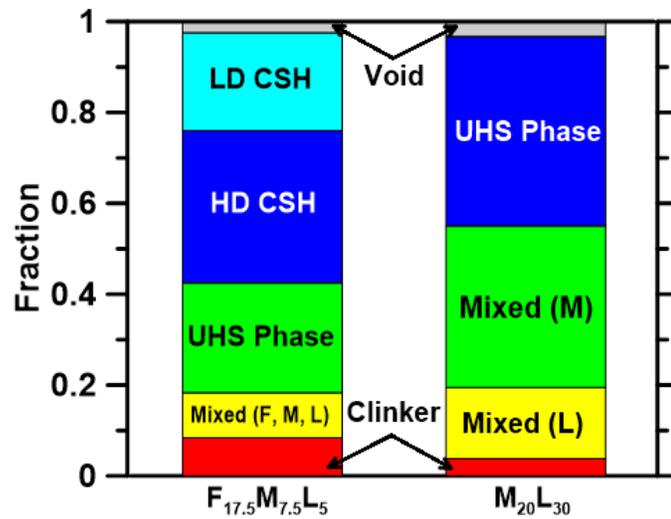
Figure 3-8: Components of the 30 day old pastes used for homogenization

Utilizing the Mori-Tanaka scheme, two different homogenization approaches were used. In the first case, a two-step homogenization method was used. Here, the cementitious phase with the highest volume fraction was used as the matrix (HD CSH for $F_{17.5}M_{7.5}L_5$ and UHS Phase for $M_{20}L_{30}$), and the other cementitious phases and voids as inclusions. In the multi-step Mori-Tanaka method, the order of homogenization was varied based on the modulus of the different phases. The phase with the highest volume fraction was chosen as the

matrix, and the voids were added as an inclusion. In the subsequent steps, the phase with the lowest modulus was homogenized with the resultant of the previous step. The remaining process was carried out similarly. For the double inclusion model, the homogenization is not capable of directly accounting for the voids. Hence, the first step in this process was the homogenization of voids with the reaction product with the highest volume fraction using the Mori-Tanaka method. Next, the other identified cementitious phases were used as double inclusions in the previously homogenized matrix. In the third step, both coarse and fine aggregates were used as inclusions. The homogenization approaches are schematically represented in Figure 3-9.
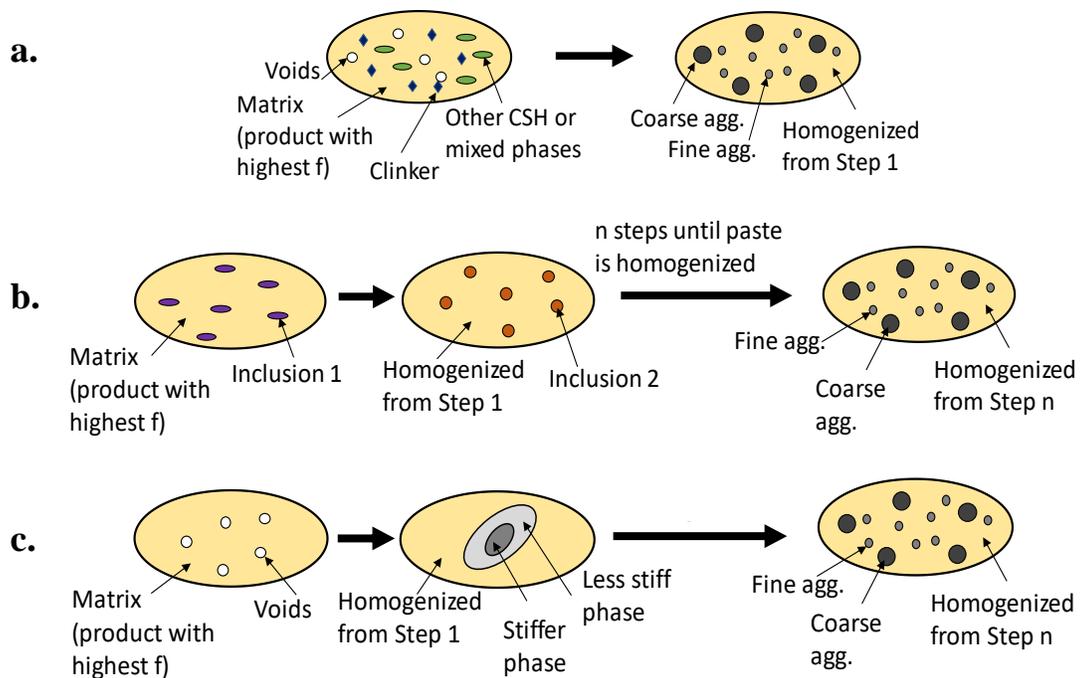
Figure 3-9: Schematic representation of the homogenization schemes: (a) two-step Mori-Tanaka scheme, (b) multi-step Mori-Tanaka scheme, and (c) double inclusion method

Table 3-4 reports the stiffness results from the analytical homogenization process for both the paste and UHPC. The elastic modulus determined from quasi-static, radial strain

63

controlled compression tests from a recent study of 28-day cured UHPC by the authors [41] are also provided. The homogenized elastic modulus of the fly ash-based UHPC is similar to that of the experimental value from compression tests, but the microsilica-based mixture shows a higher predicted value. The higher predicted stiffness of the microsilica-based paste can be attributed to the presence of large volume fractions of UHS phase as well as the mixed phases observed from the cluster plots. It is hypothesized that the high volume of microsilica used in this paste was less well-dispersed in a low w/b system, despite extended mixing and the use of a high volume of superplasticizer, as reported elsewhere [133]. The reduced dispersion of microsilica reduces its degree of reaction, and consequently, high stiffness mixed phases are present instead of lower density C-S-H phases, which results in an over-prediction of the paste modulus. When aggregates are incorporated into UHPC, the shearing action of aggregates on the paste de-agglomerates many of these larger associations. UHPC mixtures are mixed in a high-shear mixer for a longer duration [41], and hence more reaction products of lower stiffness are formed at the expense of higher stiffness agglomerated phases. Thus the experimental stiffness of the concrete mixture is lower than the upscaled value obtained through analytical homogenization. Ongoing chemical species mapping of indentation locations is expected to shed more light into this effect.

Table 3-4: Analytical homogenization results for the 30-day cured UHPC mixtures

| UHP Paste | Method of homogenization | Elastic modulus from homogenization | | Elastic modulus of UHPC from experiments (GPa) [41] |
		E of UHP Paste (GPa)	E of UHPC (GPa)	
$F_{17.5}M_{7.5}L_5$	Mori-Tanaka (Two step)	36.96	46.56	47.53
	Mori-Tanaka (Multi-step)	37.17	44.38	
	Double inclusion	36.78	44.09	
$M_{20}L_{30}$	Mori-Tanaka (Two step)	46.05	51.18	43.1
	Mori-Tanaka (Multi-step)	45.68	50.90	
	Double inclusion	45.67	50.89	

3.3.7 The Nature of the UHS Phase in UHP Pastes

Previous studies have identified the UHS phase as a very high packing density (~0.84) C-S-H phase [57] [116] or as a nanocomposite of HD C-S-H and nanoscale CH [69] [134]. In very low w/b pastes where the particles are efficiently packed through the use of ultrafine materials, there is deficiency of water and space for the formation of CH microcrystals outside the gel pores. In such highly confined reaction zones, nanosized crystals are preferred since there is limited space between the reactants, and the supersaturation in the liquid is high [69].

Assuming that C-S-H is nanogranular in nature as described in [112] [135] [136] [137], and that the UHS clusters identified through nanoindentation contain no other products or interfaces with similar mechanical characteristics, the phase packing density (η) can determined based on its relationship with the indentation modulus and hardness described as follows [57] [104]:

$$\frac{M}{m_s} = \prod_M(v_s, \eta, \eta_0) \tag{17}$$

$$\frac{H}{h_s} = \prod_H(\alpha_s, \eta, \eta_0, \theta) \tag{18}$$

where M is the indentation modulus, η is the packing density, m_s is the asymptotic elastic modulus $= M(\eta \to 1)$, v_s is the Poisson's ratio of the solid, η_0 is the solid percolation threshold, H is the indentation hardness, h_s is the asymptotic hardness $= H(\eta \to 1)$ of the cohesive-friction solid that obeys the Druker-Prager criterion, α_s is the solid friction coefficient, and θ is the indenter cone angle.

$$m_s = \frac{E_s}{1-v_s^2} \tag{19}$$

$$h_s = c_s \times A(1 + B\alpha_s + (C\alpha_s)^3 + (D\alpha_s)^{10}) \tag{20}$$

where E_s is the plane-stress elastic modulus and c_s is the cohesion of the solid. The variables A, B, C, and D are the fitted parameters corresponding to a simulated indentation experiment on a granular, isotropic porous solid with a conical indenter with equivalent cone angle to a Berkovich tip [138] [139]. For a Berkovich indenter with a cone angle of 70.32°, A = 4.76438, B = 2.5934, C = 2.1860, and D = 1.6777 [138]. The solid properties m_s, c_s, v_s, α_s, and the packing density η can be back-calculated from the experimental M and H values [57] [104].

The HD C-S-H in the OPC- and fly ash-based mixtures demonstrate packing densities of 0.76±0.03 and 0.72±0.03 respectively (note that the microsilica-based paste does not show a HD C-S-H phase) based on the above-mentioned granular model, which corresponds to FCC or HCP packing ($\eta \approx 0.74$) [104] [54]. The packing densities reported are average values from a large number of indents. The volume fraction of C-S-H globules in HD C-S-H is approximately equal to the packing density of HD C-S-H, ~0.75, therefore a gel porosity of $\phi_{HD-CSH} \approx 0.25$ can be used for HD C-S-H for all the pastes. If the CH

66

nanocrystals are formed within the HD C-S-H structure, they result in an effective reduction in the HD C-S-H porosity ($\phi_{HD\text{-}CSH}$) by an amount equal to the CH volume fraction (f_{CH}). Thus, the residual gel porosity of the UHS phase ($\phi_{residual}$) can be defined as:

$$\phi_{residual} = \phi_{HD-CSH} - f_{CH} \tag{21}$$

Note that it is being implicitly assumed here that the local constraints force the formation of nanoscale CH and not microcrystals as in conventional pastes. The theoretical maximum CH fraction that can precipitate as nanocrystals in HD C-S-H to form the UHS phase is equal to $\phi_{HD\text{-}CSH}$, even though it could generally be lower. The volume fraction of CH, f_{CH}, in each of the indentation volumes cannot be accurately determined without chemical mapping of the indentation locations using energy dispersive X-ray analysis. A Mori-Tanaka homogenization scheme is used to determine the effective modulus of the UHS phase consisting of HD C-S-H and nanoscale CH. The homogenized, theoretical UHS phase moduli are plotted as a function of residual gel porosity in Figure 3-10, shown as solid lines. For a residual porosity of zero, f_{CH} is 0.25 (i.e., the composite is 75% HD C-S-H and 25% nano CH), whereas for a residual porosity of 0.25, f_{CH} is zero. The lower and upper bounds provided in this plot correspond to indentation plane stress elastic modulus of the C-S-H globules of 65 GPa and 75 GPa respectively. The larger value is used to account for the potential enhancement in globule stiffness with decreasing C/S ratio of the gel [25] [28]. The indentation modulus of CH is taken as 46 GPa [69].

Considering the microsilica-based UHP paste after 30 days of hydration as an illustration (the nanoindentation stiffness is ~45 GPa, shown using the dotted line; See Table 3-3), if

the UHS phase is a composite of HD C-S-H and CH, the stiffness bounds indicate that residual porosity of this phase should be between 0.17 and 0.24 (or f_{CH} between 0.08 and 0.01). The f_{CH} determined from TGA for this mixture is 0.047, which lies in this range. Similar results for the other mixtures are noted and they all fall within or reasonably close to the bounds. Note that $\phi_{HD\text{-}CSH}$ is obtained from the packing density of C-S-H, which is an averaged value over many indents within a representative volume. Similarly, f_{CH} obtained from TGA is also averaged over a tested volume. As a confirmation, UHS stiffness data and f_{CH} values from [69] are also plotted, which is also observed to fall within the bounds. This analysis confirms that the UHS phase can be mechanically modelled as a composite of HD C-S-H and nanoscale CH.
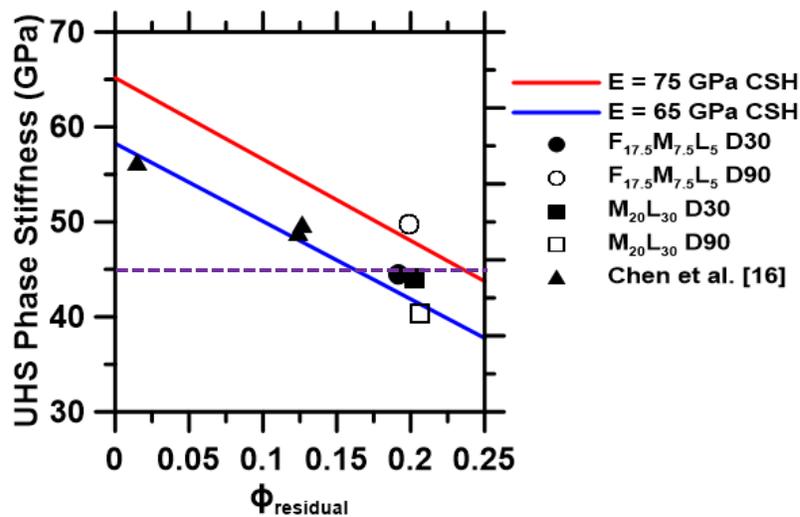
Figure 3-10: Homogenized UHS phase modulus as a function of residual porosity of the UHS phase.

3.4 SUMMARY

This chapter has discussed the nanoscale mechanical characterization of UHP cementitious matrices containing high volumes of commonly available cement replacement materials. Nanoindentation experiments coupled with a Bayesian information criterion were carried

out to determine the average modulus and hardness of the statistical phases. The microstructure was found to be highly heterogeneous, attributable mainly to the low w/b used and multitude of cement replacement materials of differing size ranges. LD and HD C-S-H along with an UHS phase was identified for the fly ash-based UHP paste (w/b ~ 0.20) after 30 days of hydration, while only the UHS phase was observed for the fly ash-based paste after 90 days of hydration. For the microsilica-based UHP paste, the volume fractions of LD and HD C-S-H phases were so low that the points were clustered together with the UHS phase at all ages. Both the UHP pastes contained several mixed/composite phases as determined from the M-H response. Nanomechanical analysis reveals the enhanced influence of physical packing of particles in low w/b systems such as UHPC. Analytical homogenization models based on Eshelby's solution for inclusions embedded in a matrix were used to upscale the elastic response of the individual phases in UHP paste to the elastic response of UHPC. For the fly ash-based UHPC, the homogenized E was in good agreement with that determined experimentally using strain controlled compression tests on UHPC specimens. However, for the microsilica-based UHPC, the upscaled E was found to be higher than that from experiments.

CHAPTER 4[2]

RELATING THE NANO-MECHANICAL RESPONSE AND QUALITATIVE

CHEMICAL MAPS OF MULTI-COMPONENT ULTRA-HIGH PERFORMANCE

CEMENTITIOUS BINDERS

4.1 INTRODUCTION

Ultra-high performance concretes (UHPC) are designed for critical infrastructural

systems to feature a wide-range of benefits including high compressive and flexural

strengths, improved ductility, and enhanced durability [18]. Of late, UHPCs are being

used in bridges as deck surfaces, deck connections replacing steel, or in the repair and

strengthening of bridge elements [140, 5]. UHPC mixtures are characterized by low

water-to-binder ratios (w/b), incorporation of cement replacement materials such as fly

ash, microsilica/nanosilica, metakaolin, mine tailings [22], fine fillers such as limestone

powder or ground quartz [36], and steel fibers to meet the stringent requirements of

strength, workability, ductility, and durability [17, 23]. The dense microstructures of

UHPC also signify excellent freeze-thaw resistance and chloride diffusion coefficients

30-600 times smaller than that of normal concrete [141]. While the requirement of high

compressive strengths (in excess of 130 MPa) generally demands high volumes of

cement, mixture proportions that maximize the use of more sustainable cement

replacement materials have been developed and implemented [140, 21, 6]. The reduced

degree of reaction of the cementitious materials (due to the low w/b) and the high

---

[2] This chapter is derived from the publication: E. L. Ford, C. G. Hoover, B. Mobasher and N. Neithalath, "Relating the nano-mechanical response and qualitative chemical maps of multi-component ultra-high performance cementitious binders," *Construction and Building Materials,* vol. 260, p. 119959, 2020.

volumes of fine materials that enable efficient particle packing results in a high degree of physical and chemical heterogeneity in the UHPC paste microstructure [18, 23].

It is well-documented that heterogeneous microstructures can be probed using statistical nanoindentation to determine the mechanical properties (i.e., stiffness and hardness) of constituent phases and their corresponding volume fractions [104, 51]. In the context of cementitious materials, this technique has been extensively used to discern the stiffness and hardness of different C-S-H phases (e.g., low density and high density C-S-H), calcium hydroxide (CH), unhydrated clinker grains, unreacted fly ash particles etc. [110, 64, 111] [59]. The properties of the constituent phases can be used in conjunction with micromechanical modeling schemes to provide upscaled properties of the composite (e.g., elastic modulus) [50, 122, 126]. However, when the complexity and heterogeneity of the system increases, such as in the case of UHPC mixtures containing components with differing levels of reactivity, the mechanical information in the indentation volume represents a composite response and needs to be supplemented by chemical data to identify the reaction products more succinctly. For instance, in pastes containing limestone filler and an alumina source, (mono- or hemi-) carboaluminates are formed, which are reported to possess similar stiffness as limestone or microsilica particles [38, 39]. Likewise is the case with the ultra-high density (UHD) reaction product [57, 69] and CH [113]. Reaction product identification in heterogeneous systems is accomplished through the use of qualitative or quantitative energy dispersive X-ray spectra (EDS) analysis of the probed volumes [69, 109, 70].

This paper demonstrates the combined use of statistical nanoindentation and chemical analysis on the probed volumes based on qualitative EDS mapping for two different,

71

highly heterogeneous UHPC binders. Both these binders were used in proportioning UHPCs that demonstrated similar 90-day compressive strengths (150-160 MPa [41]), albeit with different starting materials and overall cement replacement levels (30% and 50% by mass). This study advances our understanding of the relationships between the micromechanical properties of different phases and the microstructural chemistry in UHPC binders with varying source material chemistry and properties. The intensities of Ca, Si, and Al, their appropriate ratios in the reaction product phases, and their relationships with the nanoindentation stiffnesses of the phases are uncovered for the C-A-S-H gel which forms the major constituent of multicomponent UHPC binders.

4.2 EXPERIMENTAL PROGRAM

4.2.1 Materials and Mixtures

UHP cementitious pastes were prepared using Type I/II ordinary Portland cement (OPC) conforming to ASTM C 150, Class F fly ash conforming to ASTM C 618, limestone powder of two different median sizes (1.5 µm and 3.0 µm) conforming to ASTM C 568, and microsilica conforming to ASTM C 1240. Particle size distributions, extracted from laser particle size analysis, are shown in Figure 4-1 and the chemical composition of the source materials is shown in Table 4-1. A polycarboxylate ether (PCE)-based superplasticizer with a solids content of 43% was used to ensure workability of the very low w/b pastes. Two UHP pastes (w/b ~ 0.20) were prepared for the experiments in this study, the proportions of which are shown in Table 4-2. The binder contents in the UHP pastes were based on previous work of the authors [18, 41] that proportioned a family of UHPCs based on optimal particle packing, rheology, and substantial clinker factor

72

reduction. Paste samples were prepared by extensive mixing using high shear mixers and

moist cured for 30 or 90 days.
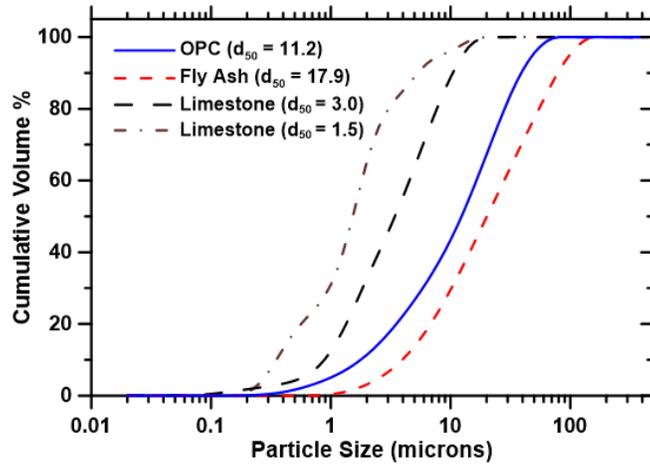
Figure 4-1: Particle size distribution (PSD) of the starting materials. The median size in microns is shown in parentheses. The PSD of microsilica is not shown, but the median size is $< 0.5\ \mu m$ in a well-dispersed state.

Table 4-1: Chemical composition and specific gravity of the starting materials used in this study.

Components of the binder	Chemical composition (% by mass)							Specific gravity
	SiO_2	Al_2O_3	Fe_2O_3	CaO	MgO	SO_3	LOI	
OPC	19.60	4.09	3.39	63.21	3.37	3.17	2.54	3.15
Fly Ash (FA)	58.40	23.80	4.19	7.32	1.11	3.04	2.13	2.24
Microsilica (MS)	> 90.0	-	-	< 1.0	-	-	-	2.18
Limestone (L), 1.5 μm			> 97% $CaCO_3$					2.7
Limestone (L), 3 μm								

Table 4-2: The components of the paste mixture. FA, MS, and L indicate fly ash, microsilica, and limestone respectively. The subscript numbers refer to the percentage of the corresponding material used as a replacement of OPC by mass. These pastes are termed FML and ML in the remainder of the paper.

Mixture Label	Binder component (mass %)				% solids content by mass of binder
	OPC	Class F Fly Ash (FA)	Microsilica (MS)	Limestone (L)	Superplasticizer
$FA_{17.5}MS_{7.5}L_5$ (FML)	70	17.5	7.5	5	1.25
$MS_{20}L_{30}$ (ML)	50	0	20	30	1.45

4.2.2 Sample Preparation

After the moist curing duration, the samples were cut in the form of discs of 12.5 mm thickness and 25 mm diameter using a Bruker IsoMet 1000 saw with a diamond wafer blade. Isopropyl alcohol (IPA) was used to cool the saw and to store the samples after cutting to prevent further hydration [74]. Sample surface preparation to minimize imperfections that could interfere with the results was carried out by sanding and polishing the samples successively using silicon carbide abrasive paper with sizes ranging from 240 grit to 1200 grit, and then polishing using alumina pads with particle diameters of 9, 3, and 1 µm [51, 76]. Figure 4-2 shows an example 10 µm x 10 µm atomic force microscopy (AFM) scan of a polished surface used to quantify the surface characteristics after preparation using the root mean square roughness (S_q), given as [50, 76] [60]:

$$S_q = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} z_{ij}^2} \tag{22}$$

where M is the number of pixels in the x direction, N is the number of pixels in the y direction, and z_{ij} is the height of the pixel at the position (i, j) from the mean plane. AFM was performed using a Nanosurf FlexAFM with EasyScan2 software in dynamic

(tapping) mode with a pyramidal silicon tip scanning at a rate of 1 sec/line with 512

pixels/line. The tip was calibrated using the Sader method [79]. A vibration frequency of

184 kHz, amplitude of 0.6V, and a spring constant of 26 N/m were obtained. Gwyddion,

an open source software for data visualization and analysis of scanning probe microscopy

techniques, was used to process the scan by mean plane subtraction and application of

row alignment with a first degree polynomial, which produced an RMS roughness value

of $S_q = 39.36$ nm. This roughness represents about 16% of the maximum indentation

depth of 250 nm. This value conforms to the recommendation that a sufficiently flat and

smooth surface for nanoindentation should have an S_q less than 20% of the maximum

indentation depth or 50 nm [76].
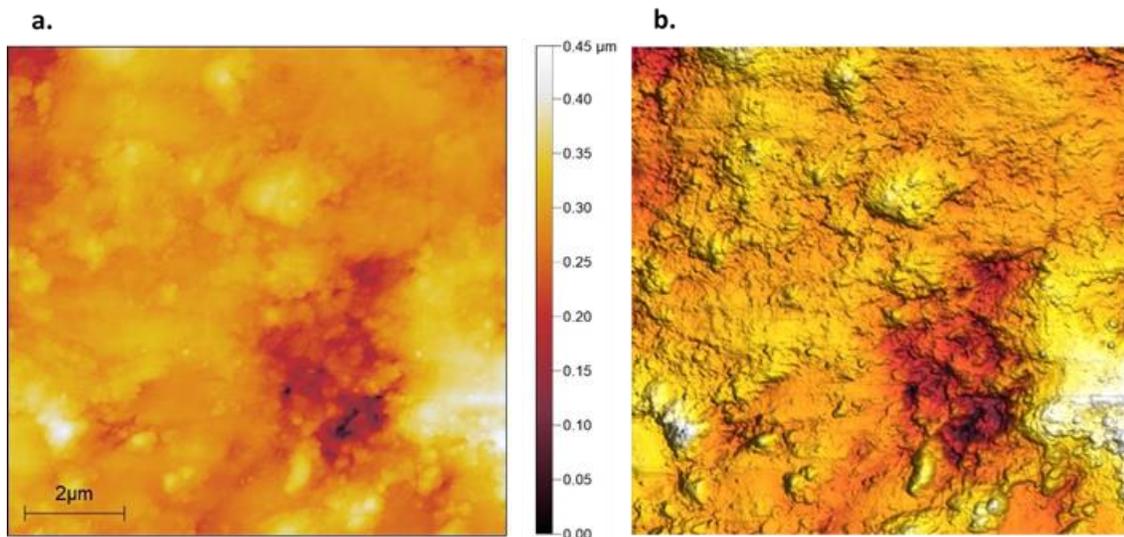
Figure 4-2: Representative 10 µm x 10 µm AFM scans over the polished surface of a 90-day cured FML specimen with $S_q = 39.36\ nm.$ (a) 2D View, and (b) 3D overlay projection.

4.2.3 Nanoindentation

Nanoindentation experiments were carried out using an Ultra Nanoindentation Tester (UNHT[3]; Anton Paar). Each sample had at least 1250 indents, split among multiple grids in different locations, to adequately capture the heterogeneous nature of the microstructure and minimize any effects of spatial dependence. Indentations were performed in force control mode with a maximum displacement cutoff of 250 nm, which is in line with other studies on cement-based materials [69, 52, 142]. This depth corresponded to an interaction volume idealized as a hemisphere with a radius 3-to-5 times the maximum cutoff, or 0.75 μm to 1.25 μm [69, 107] [143]. The distance between points in the grid was maintained at 5 μm to ensure that the influence of the indents on each other was minimum [50]. The loading and unloading rate used was 12 mN/min with a pause for 5 seconds at the peak load when the maximum displacement cutoff was reached. This rate and holding period are sufficiently short to avoid creep effects as discussed in [144]. The linear unloading portion of the force-displacement curve and the contact area of the tip is used to calculate the hardness (H) and the effective Young's Modulus (M) of the indented phases using the Oliver and Pharr method [65, 66]. The effective Young's Modulus (M) is related to the elastic moduli of the tip and the Poisson's ratios of the tip and the indented material as shown below:

$$\frac{1}{M} = \frac{1 - \upsilon^2}{E} + \frac{1 - v_i^2}{E_i} \tag{23}$$

Here, υ is the Poisson's ratio of the sample, and υ_i and E_i are the Poisson's ratio and Young's Modulus of the diamond tip (0.07 and 1141 GPa respectively) [65, 66]. υ for the different phases was assigned after identifying them through a clustering analysis [39]. As such, clinker was assigned a υ of 0.31, HD and LD C-S-H a υ of 0.25, UHS phase a υ

of 0.29 (approximated as the average υ of C-S-H and CH phases), and mixed phase a υ of 0.27 (between reactants and products). As part of post-processing, abnormal load-depth curves representing surface pores or partial material collapse were removed from the data sets as described in [59] [72].

4.2.4 Scanning Electron Microscopy (SEM) and Energy Dispersive X-Ray Spectroscopy (EDS) Chemical Mapping

A SEC Nanoimage Tabletop SNE-4500M Plus scanning electron microscope (SEM) equipped with a Bruker energy dispersive X-ray spectrometer (EDS) and ESPRIT software was used for chemical mapping of the indented regions of interest. A gold-palladium layer, about 20 nm thick, was sputter-coated on the sample surface to improve electron conduction [73]. Back-scattered electron (BSE) mode imaging was performed in the low vacuum mode (~10^{-3} torr), with a beam current of 110 µA, a working distance of 10 mm, sample tilt of 10°, and an accelerating voltage of 15 kV. After locating the grid on the sample surface, EDS was performed at about 50 kcps over a 256 x 192 pixel image, magnified for a pixel size of about 0.6 µm, with a dwell time of 128 µs (~6 seconds per frame) for 5-15 minutes. The total scan time was maximized on each sample to achieve the best EDS image prior to the loss of image quality due to surface charging. The interaction volume and escape depth of the signal X-rays are dependent on the microscope parameters as well as the sample density and composition [69, 72]. It has been shown through Monte-Carlo simulations that, for both clinker and C-S-H phases, most of the characteristic X-rays escaping the material are generated within a depth of 2 µm [69, 72], which is in line with the interaction depth for nanoindentation as reported earlier. It needs to be noted that continuous spectrum background (signal noise) is

inherently present in EDS mapping, leading to low elemental intensities even in spots where the element is not present [68]. This necessitates an understanding of the relative elemental intensities that can occur in cement-based materials. Moreover, electron microscopy and EDS are destructive processes where the electrons impinging on surface of the sample will damage the nanoindentation grid once scanned.

4.2.5   Optical Alignment

Figure 4-3 (a) shows an optical microscope image of the nanoindentation grid on a 30-day cured FML paste, and Figure 4-3(b) shows the grayscale BSE image of the same microstructure. In order to relate the elemental EDS analysis corresponding to an indent to the nanomechanical data of the same indent, a MATLAB localization algorithm was implemented to align the optical image of the nanoindentation grid to the EDS chemical maps. This algorithm employs image enhancement, indent identification, and Hough transformation-based grid detection [109] to determine the coordinates of the indents from the optical image. With a pixel length of 0.6 µm, an averaging filter of size 2x2 pixels was taken over the indent coordinates on the EDS maps to account for the full 1.25 µm expected diameter of influence about each indent. A 4x4 pixel filter was also applied as part of a sensitivity analysis, and it was observed that the chemical intensity results were almost invariant. This averaging of elemental pixel intensities over the entire region of the indent's influence ensures probing similar interaction volumes as those in nanoindentation experiments. Brightness of the qualitative EDS chemical maps is auto-scaled by the data-collection software. The RGB color brightness ranges from 0 to 255 with a higher intensity signifying a higher concentration of the element. Figure 4-3(c) is an example of the Ca EDS map. Al, Si, and Fe maps were similarly obtained. Across

different maps, the number of X-ray counts associated with the same brightness value

varies, and hence EDS maps are qualitative measures of the concentration of elements in

each indentation grid. For statistical analysis the averaged RGB intensities from the Al,

Ca, Fe, and Si EDS maps are matched with the corresponding nanomechanical data.

Figure 4-3(d) illustrates the translation of EDS map color intensity of Ca to the 0-255

scale.
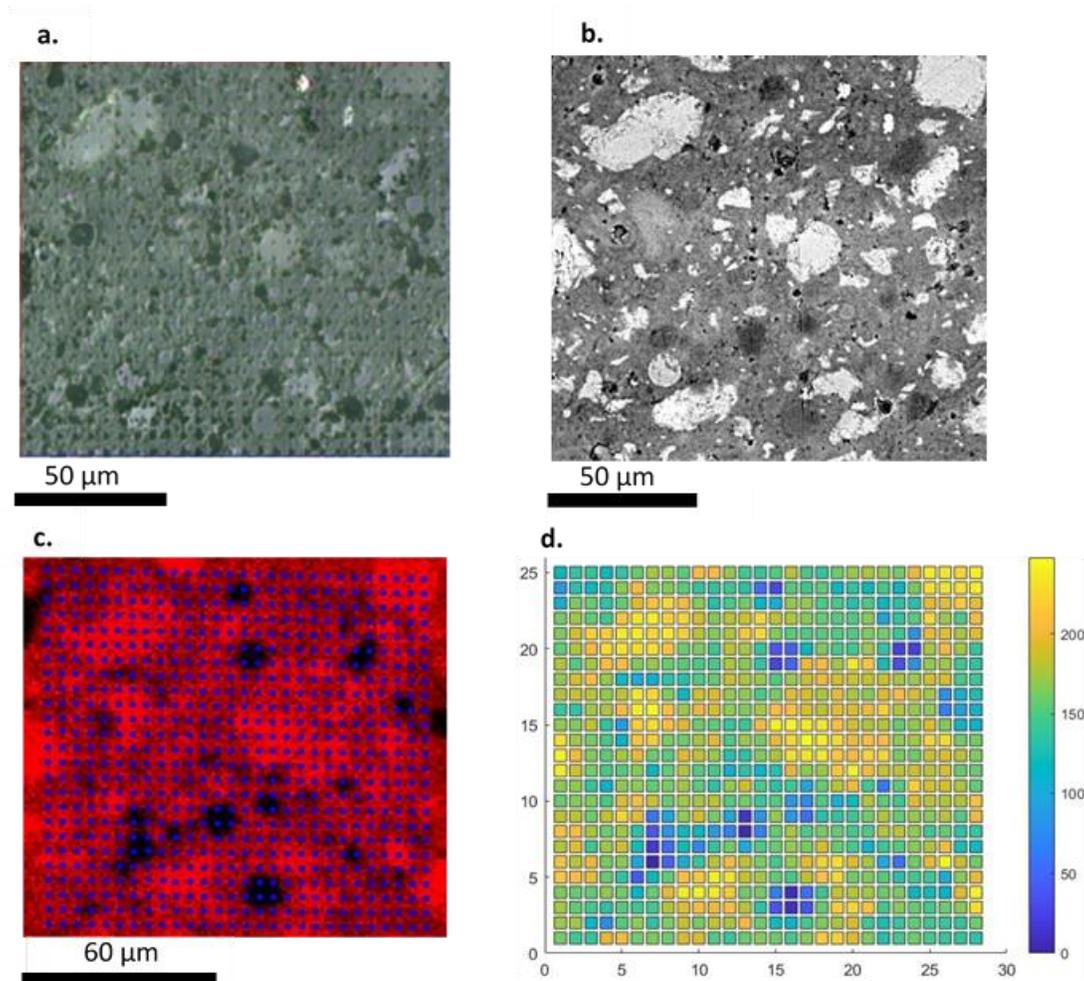
Figure 4-3: (a) Optical microscope image over a nanoindentation grid for the 30-day
cured FML paste, (b) BSE image over the same area, (c) Ca EDS map with blue dots
added to show the location of the indentation grid after the alignment procedure, and (d)
MATLAB graphic translating EDS map color intensity into values from 0 to 255 for Ca.

4.2.6 Nanomechanical and Chemical Statistical Clustering

If there exists n phases in the microstructure with each phase occupying a volume fraction of f_i (i = 1…n) such that $\sum_{i=1}^{n} f_i = 1$, the properties of each phase can be approximated by a Gaussian distribution with a probability density function (PDF) given as:

$$PDF = \sum_{i=1}^{n} f_i \, \psi_i \tag{24}$$

Here, ψ_i is the vector of classification variables of the phase. In this study, the six classification variables utilized were indentation modulus M, indentation hardness H, and the intensities of aluminum I_{Al}, calcium I_{Ca}, iron I_{Fe}, and silicon I_{Si}. A Bayesian Information Criterion (BIC) with negative log likelihood (NlogL) method was implemented for statistical deconvolution [109]. The maximum negative log likelihood estimation was used to find the PDFs that best represented the experimental data:

$$NlogL = -\max(\log(\textstyle\prod_n PDF(n_i))) \tag{25}$$

Here, n_i represents the distribution parameters, in the case of a Gaussian distribution its mean and standard deviation, that are iterated to maximize the likelihood function. Then, the BIC was minimized:

$$BIC = 2 * NlogL + p * \log(num) \tag{26}$$

In Equation 7, num is the number of indentation points and p is the number of identifying parameters available at each indentation point (in this case six; chemical intensities corresponding to Ca, Si, Al, Fe and the mechanical properties M and H) [109]. In cases where similar BIC values were obtained, the lower number of phases was chosen to

simplify the model. Occasionally the clustering algorithm yielded clusters with small populations (about 10% of the available points) that were embedded inside of one another, or unreasonably split the high hardness-stiffness values. In such cases, these small clusters were manually grouped with the nearest cluster, and the new average stiffness and hardness values were determined.

4.2.7   X-Ray Diffraction (XRD)

X-Ray Diffraction (XRD) spectra of the UHPC paste samples cured for 30 and 90 days were obtained using a Siemens D-5000 Powder X-ray Diffractometer utilizing a Cobalt (Co) Kα radiation with 2 mm and 0.2 mm slits to identify the crystalline phases in the samples. Following previous studies of cement pastes, the scanning angle (2θ) range was from 10$^\text{o}$ to 80$^\text{o}$ with a step size of 0.02$^\text{o}$ and 1 second per step [47, 48, 49]. MDI Jade 9 software was used for peak identification.

4.3 RESULTS AND DISCUSSION

4.3.1 Crystalline Phases in the UHPC Pastes
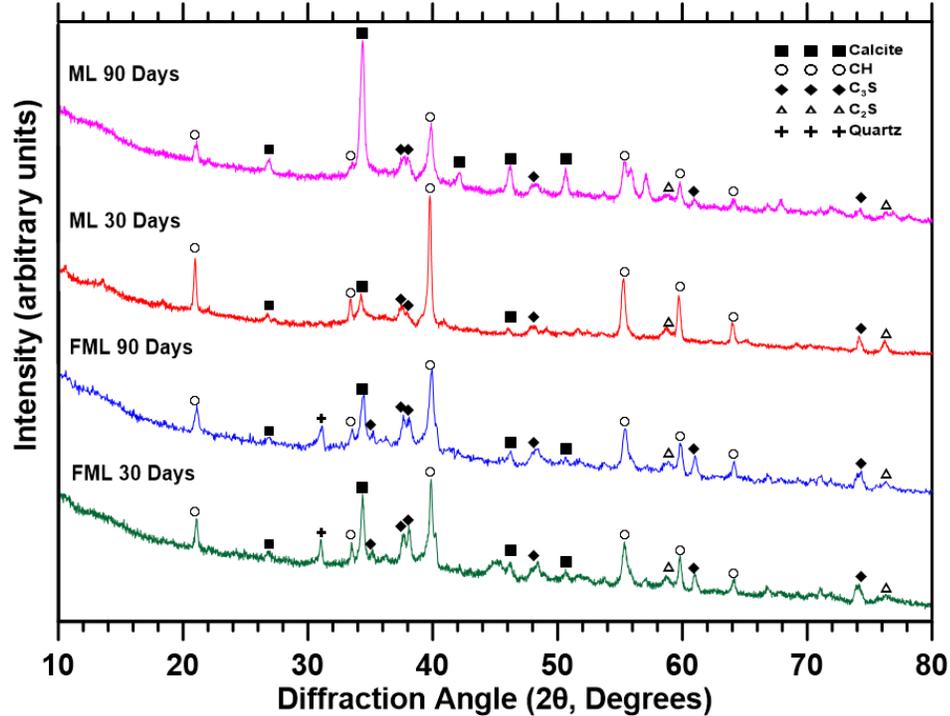
Figure 4-4: XRD spectra of 30-day and 90-day cured FML paste and ML paste.

Figure 4-4 shows the XRD spectra for the 30- and 90-day cured FML and ML UHP

pastes. The crystalline compounds identified in the FML paste are C_3S and C_2S from

clinker, quartz from fly ash, portlandite (CH) from cement hydration, and calcite

(limestone), while quartz is absent in the ML paste as would be expected. The intensity of

the calcite peaks are found to reduce between 30 and 90 days, indicating its consumption

due to the presence of reactive aluminates from fly ash [117, 119, 35]. The XRD spectra

does not clearly show the presence of carboaluminates, but they were evident in the

thermogravimetric and differential thermogravimetric (TG/DTG) analyses [4]. Beyond

this, the crystalline species present in the FML mixture do not vary greatly. The presence

of unreacted clinker and cement replacement materials even after 90 days of curing in

82

both the mixtures can be attributed to the reduced w/b and consequently, reduced reactivity. Notable in the ML paste spectra is a reduction in the CH content between 30 and 90 days, attributable to the consumption of CH by its pozzolanic reaction with microsilica [121] [145]. The low w/b in the UHP systems and the high degree of initial packing hinder the diffusion-controlled reactions, thereby delaying the pozzolanic reaction of microsilica, which is generally rapid in conventional paste mixtures (w/b ~ 0.4). The continued presence of CH even after 90 days in both the UHP pastes demonstrates the influence of low w/b on the reactivity and phase formation in such systems. The ML paste contains a high amount of limestone, and the lack of an aluminous cement replacement material in this system renders noticeable limestone consumption an unlikely scenario. This is evident in the 90-day XRD spectrum, but in the 30-day spectrum, calcite peaks are rather few. This is likely a result of the strong cleavage planes of calcite being susceptible to a preferred orientation spike in XRD [47]. TG/DTG data did indicate almost unchanged quantities of calcite [4].

4.3.2 Modulus-Hardness Relationships and Chemical Clusters from Nanoindentation and EDS

The clusters representing the indentation modulus and hardness as well as chemical intensities obtained from BIC and negative log likelihood statistical analysis for the different UHP pastes are presented here. The mean and standard deviation of the effective stiffness (M) and hardness (H) values of the different phases, along with their volume fractions are shown in Table 4-3 while the chemical clustering results for Al, Ca, and Si within the clusters are shown in Table 4-4. The chemical intensities are determined based on a qualitative map and not on quantitative spot chemical analysis [109, 72]. A holistic

qualitative map, where the indents were found and chemical intensities averaged over a grid of pixels (using the optical alignment procedure described earlier) avoids the issues with in-situ spot analysis where surface charging shifts the EDS maps and pixel placement over time [72]. However, the qualitative chemical maps only provide relative atomic ratios and not the exact ratios (e.g., Ca/Si). Figure 4-5 shows example composite SEM-EDS images with an indentation grid. The abundance of Ca and Si in the hydrated product phases can be noticed in these figures.

Table 4-3: Mechanical properties (M, H) and volume fractions (f) of the different phases in the UHPC pastes. FA, MS, L, and CA denotes fly ash, microsilica, limestone, and carboaluminates respectively

UHP Paste	Phase	M (GPa)		H (GPa)		f	
		30 d	90 d	30 d	90 d	30 d	90 d
FML paste $(FA_{17.5}MS_{7.5}L_5)$	LD CSH/Residual MS	28.86 ±5.94	-	1.62 ±0.51	-	0.17	-
	HD CSH	37.75 ±7.63	40.97 ±7.10	1.67 ±0.36	1.79 ±0.31	0.38	0.40
	UHS Phase	39.11 ±10.91	43.24 ±7.15	1.86 ±0.60	1.96 ±0.44	0.19	0.23
	Mixed (FA, L, MS, CA)	70.78 ±34.73	92.44 ±16.71	5.32 ±3.70	7.65 ±2.13	0.12	0.17
	Clinker	75.14 ±25.59	77.99 ±26.39	4.58 ±2.54	4.42 ±2.48	0.13	0.20
ML paste $(MS_{20}L_{30})$	UHS Phase/CSH	64.09 ±7.17	47.19 ±8.99	3.90 ±0.39	2.23 ±0.55	0.42	0.60
	Mixed (L, MS)	59.63 ±16.48	51.20 ±10.70	4.26 ±1.02	2.54 ±0.68	0.41	0.28
	Clinker/Unreacted	108.39 ±32.44	86.14 ±24.54	8.18 ±3.35	5.08 ±2.31	0.17	0.12

Table 4-4: Normalized chemical intensities of Al, Ca, and Si (I_{Al}, I_{Ca}, I_{Si}) of the different phases in the UHPC pastes. FA, MS, L, and CA denotes fly ash, microsilica, limestone, and carboaluminates respectively

UHP Paste	Phase	I_{Al}		I_{Ca}		I_{Si}	
		30 d	90 d	30 d	90 d	30 d	90 d
FML paste $(FA_{17.5}MS_{7.5}L_5)$	LD CSH/Residual MS	0.04 ±0.01	-	0.10 ±0.05	-	0.93 ±0.05	-
	HD CSH	0.10 ±0.04	0.08 ±0.03	0.68 ±0.09	0.74 ±0.07	0.24 ±0.12	0.28 ±0.10
	UHS Phase	0.18 ±0.10	0.20 ±0.09	0.56 ±0.15	0.68 ±0.08	0.34 ±0.20	0.25 ±0.11
	Mixed (FA, L, MS, CA)	0.39 ±0.21	0.58 ±0.20	0.47 ±0.23	0.43 ±0.16	0.42 ±0.25	0.44 ±0.20
	Clinker	0.08 ±0.03	0.09 ±0.04	0.77 ±0.13	0.81 ±0.11	0.39 ±0.15	0.28 ±0.09
ML paste $(MS_{20}L_{30})$	UHS Phase/CSH	0.28 ±0.11	0.28 ±0.10	0.66 ±0.10	0.67 ±0.08	0.13 ±0.04	0.13 ±0.04
	Mixed (L, M)	0.21 ±0.08	0.30 ±0.11	0.49 ±0.15	0.55 ±0.13	0.47 ±0.15	0.39 ±0.15
	Clinker/Unreacted	0.37 ±0.19	0.31 ±0.12	0.75 ±0.15	0.72 ±0.09	0.21 ±0.11	0.16 ±0.07
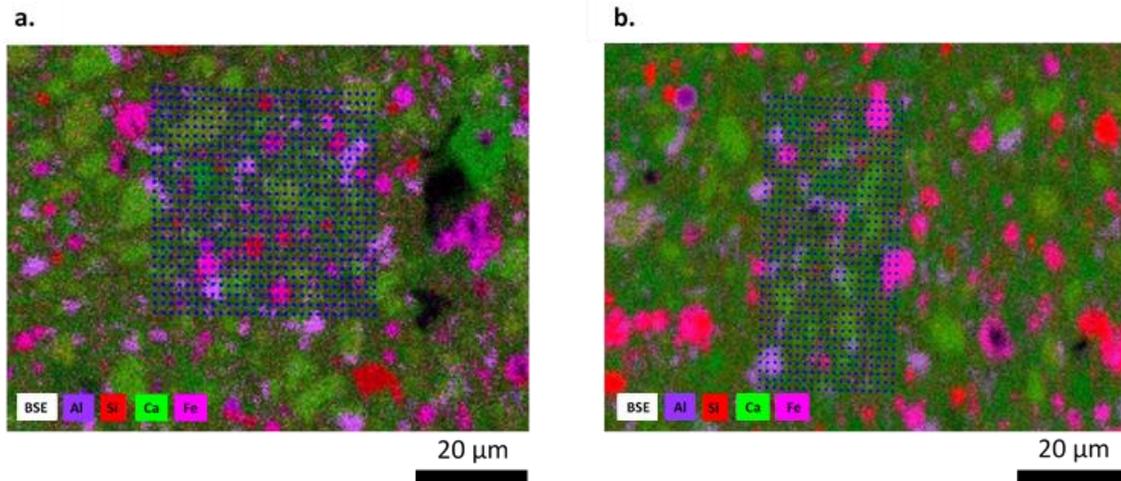
Figure 4-5: Example EDS composite maps with blue dots over grid locations (a) FML 30-day and (b) FML 90-day pastes.

4.3.3 Chemo-mechanical Analysis of the FML UHPC Paste

Figure 4-6(a) shows the modulus-hardness (M-H) relationship of the 30-day cured FML paste. Five different phases are identified by the BIC and negative log likelihood analysis. The hydration product phases, viz., C-S-H phases (high density or HD, and the ultra-high stiffness or UHS phases) are observed at $(M, H) \leq (65.0, 3.0)$ as reported elsewhere [53, 62]. The difference in the properties of C-S-H phases is attributed to the differences in the packing density of C-S-H particles [57, 116]. In low w/b mixtures, the LD C-S-H phase could be absent as reported in [50, 146]. In past works [50] [60], the low stiffness phase with a mean M of ~25 GPa has been identified as a LD C-S-H phase when only mechanical data was available [146]. However, the concomitant chemical intensities provide new insights into this phase, as will be explained later. The UHS phase has been shown to be a nanocomposite of HD C-S-H and CH. CH (indentation modulus in the 40 to 45 GPa range [39, 113]) has been identified in the XRD spectra (see Figure 4-4) and through thermal analysis [146]. Stiff and hard mixed phases are observed, which is

86

unsurprising in a blend of multiple fine particulates as described in [69]. They are

postulated to be constituted of carboaluminates (from fly ash-limestone reaction [117,

119, 35]) and other partly reacted phases. The HD C-S-H phase with an effective

modulus of 37.8 GPa is observed in the highest quantity in this mixture.
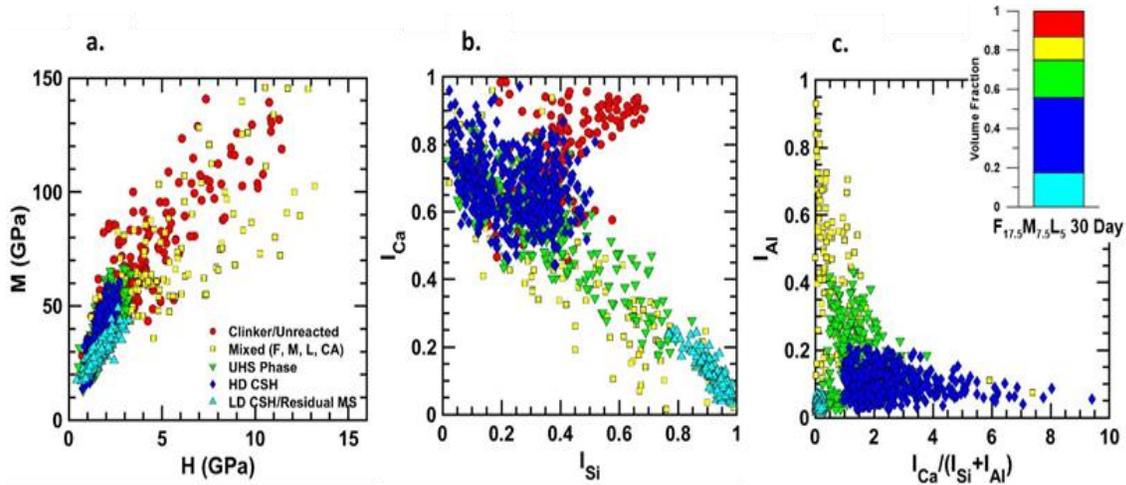
Figure 4-6: Clustering analysis output for the 30-day cured FML paste: (a) M vs. H, (b) I_{Ca} vs. I_{Si}, and (c) I_{Ca} vs. $I_{Ca/(Si+Al)}$, excluding the unreacted clinker. Inset in (c) is the volume fractions of the phases.

The normalized intensities of Ca and Si species, along with the mechanical property

description provided by nanoindentation allows further insights into the component

phases, especially for complex, heterogeneous systems such as UHPC pastes. The

normalized auto-scaled intensities of Ca and Si obtained from qualitative EDS analysis of

the FML paste microstructures after 30 days of curing are shown in Figure 4-6(b). Note

that the normalized intensities are achieved by dividing the EDS map RGB value by 255.

The same intensity value on different grids corresponds to a different number of X-ray

counts owing to differences in scan time and yield rate of X-rays from the surface in each

location. Thus, comparisons between systems with vastly different X-ray yields are

deemed to be not totally appropriate, even though they can allow for useful qualitative

inferences. The intensities are grouped into distinct classes which conform to their

nanomechanical information, as is reported for conventional cement pastes and mortars [109, 70], even though the scatter is larger for UHPC mixtures, in part due to the multiple materials used. The clinker phases, expectedly, have high Ca contents and moderate Si contents. The HD C-S-H and UHS phases also have similar Ca and Si intensities, even though the spread is larger for the UHS phase, likely due to the lower Ca/Si ratio of the gel. Both these phases demonstrate an elliptical spread that is common for the lower hardness and stiffness reactant phases [109, 70, 72]. Some CH identified in these mixtures using thermal analysis and XRD, shown to be included in the UHS phase [146], is also corroborated by the existence of compositions containing very high Ca and very low Si belonging to the UHS phase. The cluster with a mean stiffness of ~75 GPa is identified as silicate-rich clinker [70] [72], and the Ca/Si ratio of this phase as seen Table 4-4 agrees with that of C_2S and C_3S identified in the XRD spectra (Figure 4-4).

An anomaly with respect to the identifications based on mechanical properties is noticed at the higher Si end of the compositional map for this mixture. While nanomechanical signatures of M and H identified a LD C-S-H with mean (M, H) of ~ (23, 1.7) GPa, this phase is observed to have negligible amounts of Ca, based on the EDS analysis. Further evaluation (Figure 4-6(c)) also shows that this is an Al-deficient composition, indicating the phase to be pure silicates. As will be shown later, this phase is not present in 90-day cured mixtures, which likely suggests that this belongs to the microsilica added in the mixture, some of which could remain agglomerated in pastes (which is generally avoided in concretes through the shearing action provided by the aggregates) and thus demonstrate reduced reactivity early on.

The presence of fly ash containing higher amounts of aluminates results in the formation of C-A-S-H gel as the reaction product [71]. Figure 4-6(c), depicting the variation of Al with Ca/(Si+Al), which is a descriptor of C-A-S-H gel chemistry [71], shows that Al is present in the HD C-S-H and UHS phases, with a higher amount in the UHS phase. More discussions on the impact of Al are provided later in this paper. It is also possible that the interaction volume contains C-S-H and mono- or hemi carbonates (Mc or Hc) resulting from the reaction of limestone with aluminates from fly ash. It can be noted from this figure that Al incorporation is predominantly in C-S-H gels with low Ca/Si ratios (both HD and UHS phases). The mean Ca/(Si+Al) ratio of the UHS phase is lower than that of the HD C-S-H phase. In general, the hydrated phases exhibit a mean Ca/(Si+Al) ratio less than 2.0, which is in good agreement with previously reported results [31]. The mixed phase cluster with a M of ~70 GPa identified from the nanomechanical tests is denoted in this figure as belonging to the high Al, low Ca/Si region, which could be the carboaluminates or partly reacted fly ash phases. The regions of very high Al intensity could be assigned to the aluminoferrites; however, their identification in the mixed phase based on the nanomechanical signature, as opposed to the unreacted clinker phase, is another illustration of the challenges in accurately identifying chemical clusters in highly heterogeneous multicomponent binder systems.

The nanomechanical and chemical signatures of different phases in the 90-day cured FML mixture are shown in Figure 4-7. The major hydrated phases identified through nanomechanical clustering are the HD C-S-H and UHS phases, as observed from Figure 4-7(a). The auto-scaled Ca-Si intensity plots (Figure 4-7(b)) for both the 30- and 90-day cured mixtures have a similar shape, but it is important to note that the mean Ca/Si ratio

of the HD C-S-H phase dropped slightly from 30 to 90 days. This is in accordance with the understanding that increased pozzolanic reaction reduces the Ca/Si ratio of the C-S-H gel [25, 26]. This can also be related to the increased mean M and H values of this phase at 90 days (Table 4-3) since lowering the Ca/Si ratio of C-S-H gel, which is a result of increased silica polymerization, is noted to increase the elastic modulus and hardness of the constituent phases [25, 28, 29]. The absolute values of Ca/Si ratios are not given high importance here since only a qualitative estimation is carried out. The Ca and Si intensities of the mixed phase are quite spread out as noticed from Figure 4-7(b), as is the Al intensity from Figure 4-7(c). Al incorporation in the HD C-S-H and UHS phases is very similar to that observed for the 30-day cured paste (Table 4-4), and the mean Ca/(Al+Si) ratio of the hydrated phases is around 2.0. The mixed phase, a combination of carboaluminates and the unreacted non-OPC reactants, has a mean stiffness of ~73 GPa, and the volume fraction slightly increased from that at 30 days, an indication of increased carboaluminate formation. The carboaluminates are reported to have M values in the 42-88 GPa range [38, 39, 114], depending on whether it is the Mc or Hc phase.
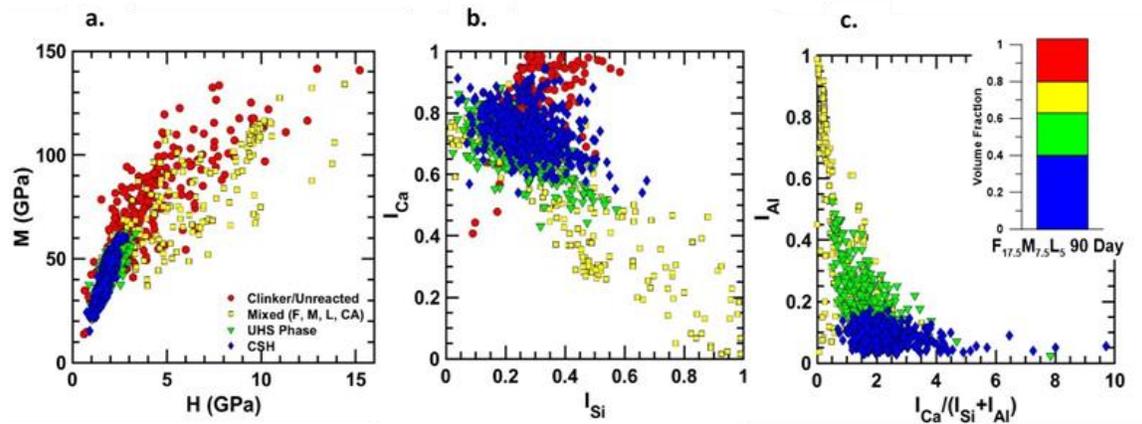
Figure 4-7: Clustering analysis output for the 90-day cured FML paste: (a) M vs. H, (b) I_{Ca} vs. I_{Si}, and (c) I_{Ca} vs. $I_{Ca/(Si+Al)}$, excluding the unreacted clinker. Inset in (c) is the volume fractions of the phases.

4.3.4   Chemo-mechanical Analysis of the ML UHPC Paste

The ML paste is characterized by a lower OPC content (50% instead of 70%), and the presence of higher amounts of limestone and microsilica (and therefore reduced Al content) as compared to the FML paste. Figure 4-8 shows the nanomechanical and chemical signatures for the different phases in the 30-day cured ML paste while Figure 4-9 shows the same for the 90-day cured mixtures.
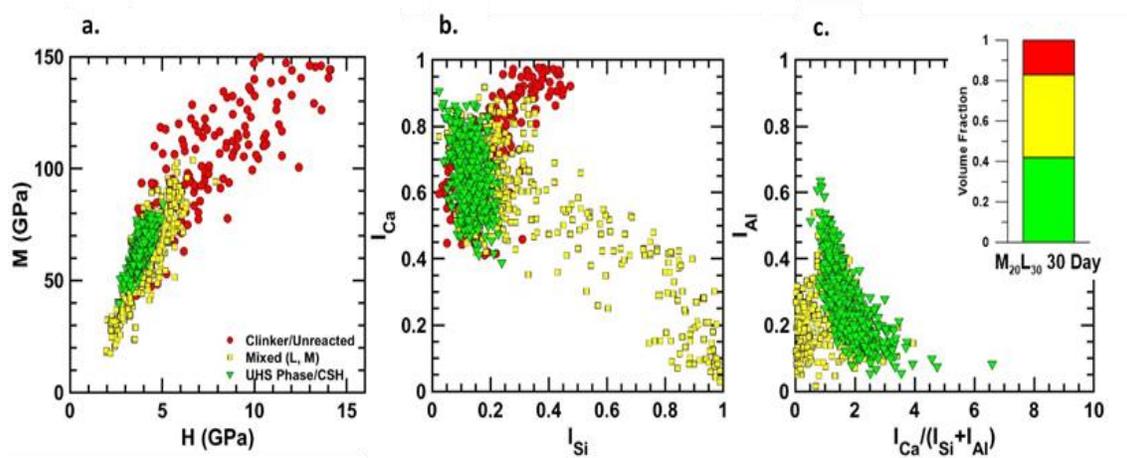
Figure 4-8: Clustering analysis output for the 30-day cured ML paste: (a) M vs. H, (b) I_{Ca} vs. I_{Si}, and (c) I_{Ca} vs. $I_{Ca/(Si+Al)}$, excluding the unreacted clinker. Inset in (c) is the volume fractions of the phases.
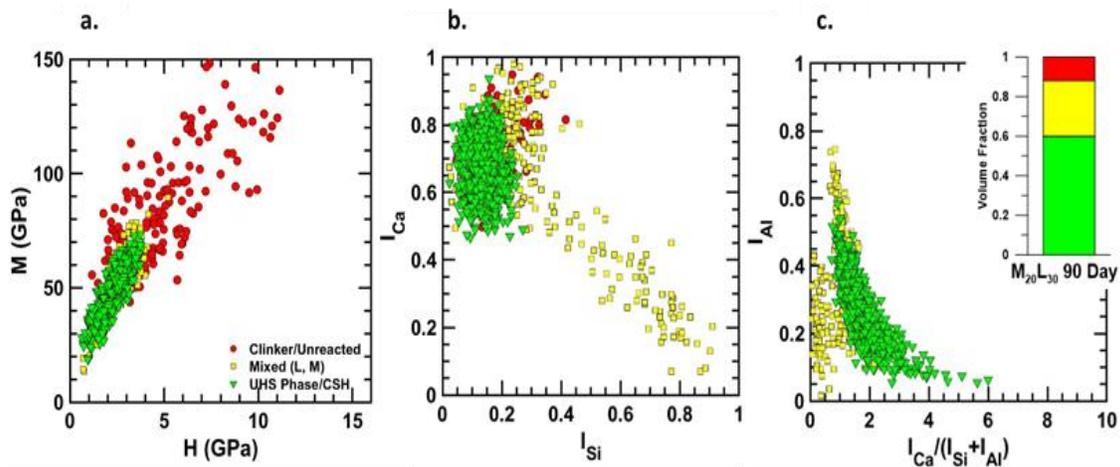
Figure 4-9: Clustering analysis output for the 90-day cured ML paste: (a) M vs. H, (b) I_{Ca} vs. I_{Si}, and (c) I_{Ca} vs. $I_{Ca/(Si+Al)}$, excluding the unreacted clinker. Inset in (c) is the volume fractions of the phases.

The nanomechanical clustering in Figure 4-8(a) and Figure 4-9(a) shows that the UHS phase is the only one belonging to the C-S-H family in this system, the reasons for which were elaborated in an earlier publication by the authors [146]. There is a mixed phase containing limestone, microsilica and some hydrates, along with the unhydrated clinker phase. The size of limestone powder (d_{50} of 1.5 to 3 µm) and microsilica ($d_{50} < 1$ µm) ensures that the indentation measurements attributed to them are of mixed phases containing these materials and the hydrated products. The volume of the UHS phase increases from 30 to 90 days as expected, at the expense of the unreacted clinker and the mixed phase. The UHS phases in both the 30- and 90-day cured ML paste show mean Ca/(Si+Al) ratios close to 2.0 (Figure 4-8(c) and Figure 4-9(c)).

The Ca/Si ratios demonstrate an elliptical spread similar to that of the FML paste, but with the major axis parallel to the Ca axis, while the major axis was inclined at ~45° for the FML paste. In other words, the Si content of the C-S-H gels lies in a narrow range for the ML paste, indicating that the reaction products are chemically more homogeneous than those in the FML paste. The higher amounts of more reactive cement replacement material (microsilica in this case) and the lower OPC content results in increased amounts of low Ca/Si ratio reaction products. The reduced heterogeneity in this mixture (fewer starting materials, and both microsilica and limestone being purer chemical species, as opposed to fly ash in the FML paste) also restricts the spread of the Ca/Si ratios. The mixed phase comprising of microsilica and limestone extends from the high Ca region to the high Si region, as would be expected. The increased concentration of points in the high Ca region denotes the unreacted limestone particles whereas the lower concentration in the high Si zone indicates that a large fraction of microsilica has reacted. A comparison

92

of the Ca-Si concentration plots between 30 and 90 days of curing shows that the number of points in the high Ca concentration region remains rather unchanged, due to the negligible reactivity of limestone in the absence of an alumina-bearing component [35, 33]. However, the number of points in the high Si concentration region reduced, indicating the consumption of microsilica to form more C-S-H gel.

The intensities of Al in the mixed phase in Figure 4-8(c) and Figure 4-9(c) show two distinct trends, one with almost no Ca content, denoting microsilica, and the other with some Ca content, overlapping with the UHS phase. The concentration of microsilica is also seen to reduce from 30 to 90 days when Figure 4-8(c) and Figure 4-9(c) are compared. The Al content in the C-S-H gel (UHS phase) can be noticed to reduce exponentially with increase in Ca content (or Ca/(Si+Al) ratio), especially for the mature pastes. This is once again confirmation that Al incorporation is higher in low Ca/Si ratio gels. Similar information can be gleaned from the data for the FML paste also; however, it is more evident in the ML paste due to reduced chemical variability of the C-S-H gel for reasons explained earlier.

4.3.5 A Closer Look at the C-A-S-H Phases in the FML and ML Mixtures and Their Mechanical Response

The M-H relationship for the C-S-H phases in the FML and ML pastes are shown in Figure 4-10(a). The FML paste containing the HD C-S-H and UHS phases have lower mean M and H values as compared to the ML paste containing the UHS phase alone. As described earlier, the HD C-S-H and UHS phases in both these mixtures are quite similar in their mean Ca/(Si+Al) ratio as observed in Figure 4-10(b), but with Ca/(Si+Al) ratio lying in a narrower range for the ML paste. In general, the mean Ca/(Si+Al) values are

located in the 1.5 to 2.0 range, as reported elsewhere [71] and mentioned earlier in the

paper. Broadly speaking, the modulus (and the hardness) does not appear to depend on

the Ca/(Si+Al) ratios based on the data shown in Figure 4-10(b) for both the pastes,

which is in line with published work [71]. However, it has been shown through

experiments and atomic simulations that a reduction in Ca/Si ratio indeed enhances the

stiffness of the solid C-S-H phases [25, 28, 29]. Combining all the data from different C-

S-H phases in this study does not capture this effect, likely because of the presence of CH

in the UHS phase which increases the Ca concentration along with the increase in M as

shown in Figure 4-10(c), especially for the FML paste. Moreover, the indentation

stiffnesses also account for the nanoporosity in the hydrated phases. Increase in Ca

concentration is shown to increase the stiffness of C-S-H gels in cement pastes [25],

while a decrease in Ca/Si ratio is also shown to increase the stiffness, especially when

cement replacement materials are used [25, 26].
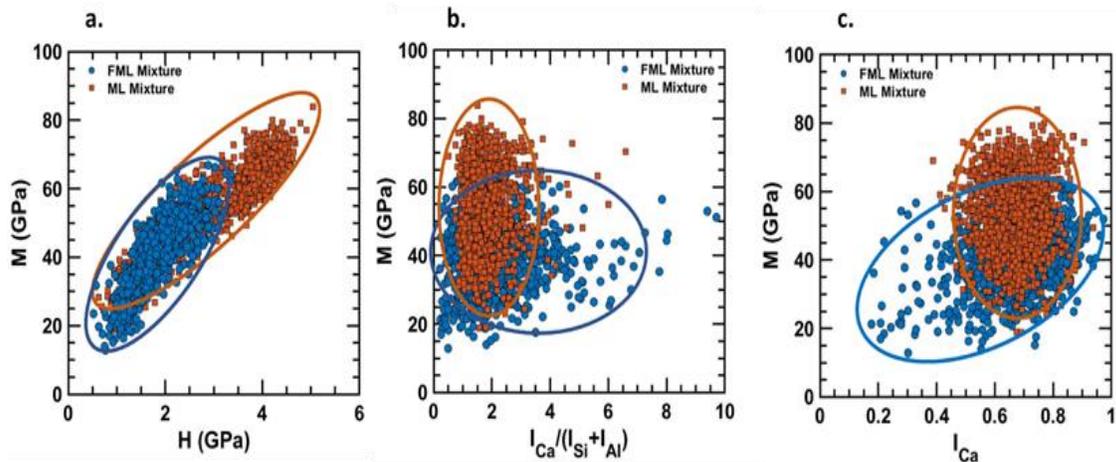
Figure 4-10: Properties and chemical intensities of the C-S-H phases in the FML and ML pastes at both ages: (a) M vs. H, (b) M vs. $I_{Ca/(Si+Al)}$, (c) M vs. I_{Ca}

It is instructive to examine the influence of Al on the properties of the gel, since the hydrates in multiple blend binder systems generally belong to the C-A-S-H family. The Ca/Si ratio is noted to be an important parameter that influences the incorporation of Al in the gel, with a reduction in Ca/Si ratio promoting the incorporation of more Al into the C-A-S-H gel [147, 148], which can be clearly noticed in Figure 4-11(a) from the rather exponential reduction in Al intensity with an increase in Ca/(Si+Al) for both the pastes. Also noteworthy is the fact that Al incorporation is not very different for the two pastes even though the ML paste has a lower Al content (see Table 4-1 and Table 4-2 for component chemistry and mixture proportions). For systems containing cement replacement materials, it has been reported that increase in Al incorporation may not necessarily be dependent on the amount of additional alumina present, but rather, due to the fact that active silica from the replacement material leads to a reduction in the Ca/Si ratio [147]. From Figure 4-10(b), Ca/Si depression is more prominent for the ML mixture, which justifies the enhanced incorporation of Al in this mixture as shown in Figure 4-11(b). It is also interesting to note that, for both the UHPC mixtures, the mean values of Ca/(Si+Al) in the hydration products remain relatively invariant with Al/(Si+Al), an observation reported in [71] for pastes containing aluminous replacement materials such as fly ash, slag, or metakaolin. When the FML and ML pastes are considered separately, there is no significant dependence of M on the Al/(Si+Al) values as shown in Figure 4-11(c) even though slightly negative correlations are sometimes reported [71]. However, atomistic simulations and synchrotron radiation-based high-pressure X-ray diffraction have shown that Al incorporation increases the bulk modulus of C-A-S-H [149].
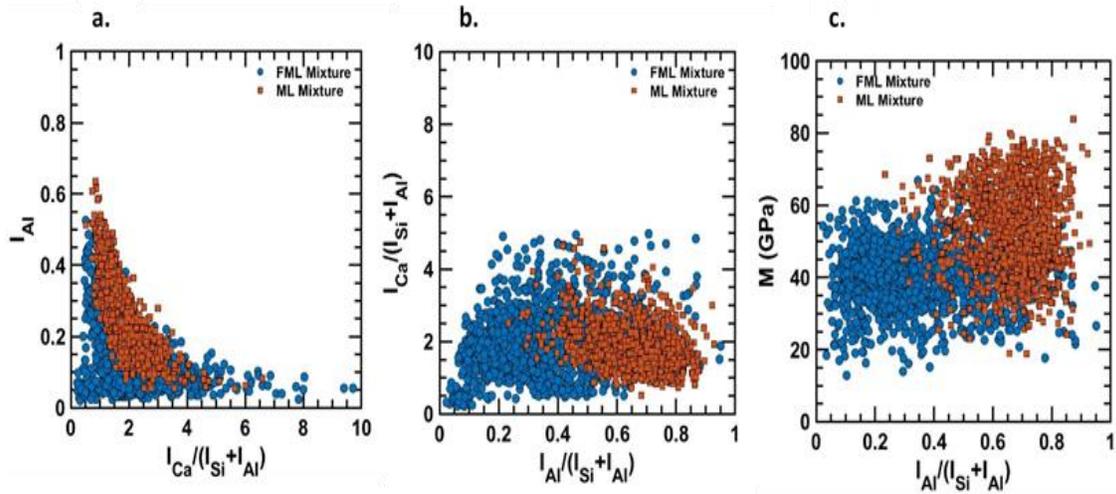
Figure 4-11: Chemical intensity ratios of the C-S-H phases of FML and ML pastes at both ages: (a) I_{Al} vs. $I_{Ca/(Si+Al)}$, (b) $I_{Ca/(Si+Al)}$ vs. $I_{Al/(Si+Al)}$, (c) M vs. $I_{Al/(Si+Al)}$

4.4 SUMMARY

Nanomechanical properties and analysis of the intensities of relevant chemical species at the indentation locations for two UHP pastes containing multiple starting materials were reported in this paper. Nanoindentation experiments were carried out on multiple grids at different ages, while the chemical analysis of the indentation points was carried out using qualitative EDS analysis, after performing a grid alignment procedure. Statistical clustering analysis of the mechanical and chemical data was performed assuming a Gaussian distribution by minimizing Bayes Information Criterion (BIC). The relationships between M, H, and the different chemical intensities were used to infer the fundamental nature of the reaction products in systems with high degree of heterogeneity, aided by the very low w/b and the presence of multiple cement replacement materials/fillers.

The UHP pastes consisted mostly of HD C-S-H and an UHS phase, along with mixed phases comprised of partly reacted starting materials and some reaction products. The

relationship between M and H of the reaction product phases in both the UHP pastes were found to be very similar, irrespective of significant changes in mixture composition. The scatter in the Ca/Si intensity plots for the UHPC pastes were found to be higher than those reported for conventional OPC pastes. The HD C-S-H and UHS phases were observed to have similar Ca and Si intensities, even though the spread was larger for the UHS phase. For both the UHP pastes, the Ca/Si ratio for the UHS phase demonstrated an elliptical spread, with reduced heterogeneity being visually identifiable for the ML paste. The normalized chemical intensities and ratios of Ca, Si, and Al species, along with the mechanical property description provided by nanoindentation, allowed for further insights into the microstructure of complex, heterogeneous systems such as UHPC pastes.

CHAPTER 5

MACHINE LEARNING ON MICROSTRUCTURAL CHEMICAL MAPS TO

CLASSIFY COMPONENT PHASES IN CEMENT PASTES

5.1 INTRODUCTION

It is well known that the microstructure of cementitious materials dictates the properties

and performance of the material. The microstructure in turn is a function of time,

processing techniques, as well as the constituent material properties and their proportions.

Cement paste microstructures are generally constituted of solid and pore phases; the

influence of porosity on the mechanical properties and durability of concrete has been

well-elucidated for many decades. The solid phase generally consists of cement hydration

products and unhydrated/unreacted materials, which depend on the water-to-binder ratio

(w/b) and the reactivity of the starting materials [150] [151] [40] [121]. While in well-

hydrated plain ordinary Portland cement (OPC) pastes, C-S-H gel, calcium hydroxide

(CH), and unhydrated clinker are invariably the only solid phases present, multi-

component blends like ultra-high performance (UHP) cementitious pastes contain

different types of C-S-H based on their density (e.g., low-density or LD, high-density or

HD), ultra-high stiffness phases, mixed reaction products, and unreacted particles of

cement, fly ash, and limestone [69] [33] [19] [36]. Thus, the microstructural complexity

increases with the use of multiple-blend binders, requiring more sophisticated and refined

methods for microstructural characterization and analysis.

Typically, scanning electron microscopy (SEM) coupled with energy-dispersive X-ray

spectroscopy (EDS) is used to extract the chemical information of the microstructure in

cement-based materials [67] [152] [153] [30]. Grid nanoindentation on these

98

microstructures provide the nanomechanical properties (or more accurately, micromechanical properties since the region of influence of the indents is of the order of 1-3 µm) [111, 146] [59]. Coupling nanoindentation data (i.e., modulus and/or hardness of the indented locations) with SEM-EDS-based microstructural chemical mapping (i.e., intensity of species such as Ca, Si, and Al) has been shown to provide much needed microscale chemistry-property relationships for cement-based materials [69] [154] [71]. Clustering algorithms such as k-means clustering or those based on Bayesian methods have been used in conjunction with nanoindentation and chemical maps of cement pastes [109] [51] [16]. The microscale properties thus obtained are upscaled using analytical or numerical tools to predict the bulk properties of the material such as elastic modulus, which are important in design [51] [124] [126].

Grid nanoindentation and chemical mapping produce large datasets, which when judiciously combined with machine learning (ML), enable the development of unbiased structure-property estimators. The use of ML to relate the properties of cement-based materials to the mixture proportions [9, 10, 11, 12, 13], or to a limited extent, to their constitutive phases [15, 16] has been reported. A recent work by the authors demonstrated the use of ML to predict the nanoindentation modulus of different phases in UHP cementitious pastes using the intensity of chemical species at indentation locations as inputs [155]. It was shown that the efficiency of predicting the modulus suffers when the microstructure becomes more complex. In addition, acquisition of nanoindentation data can be time-and-cost-prohibitive. Thus, a ML-based classification approach is adopted in this work. If ML models can be trained on elemental maps from SEM-EDS and corresponding nanoindentation data to classify locations in a SEM image as

99

belonging to the appropriate microstructural phase (e.g., LD C-S-H or unhydrated clinker, etc.), it facilitates real-time characterization and feedback loop to material processing. In this paper, the focus is on using SEM-EDS information (with or without nanoindentation data) to identify the constitutive phases as labeled from clustering analysis of nanoindentation data and chemical intensities. This allows for very quick first-order determinations of the effective material properties. Artificial Neural Networks (ANN) and hierarchical decision trees are the ML approaches adopted in this study. The classification models are implemented on two UHP cement pastes, whose properties have been extensively reported [18, 6], and validated on two other cement pastes whose characteristics are adopted from the literature [72] [70].

5.2 DATA AND ORGANIZATION

5.2.1   UHP Cement Pastes

Nanoindentation and SEM-EDS chemical data utilized in this study belong to two UHP cementitious pastes (referred to as UHP-1 and UHP-2 in Table 5-1) which have been studied in detail by the authors [146, 154, 18, 6]. As mentioned earlier, this dataset has been used in predicting the indentation modulus from chemical species intensities using ML [155]. Both UHPs contain multiple cement replacement materials (Class F fly ash, silica fume, fine limestone powder) of varying sizes and reactivity, and a low water-to-binder ratio (w/b), as shown in Table 5-1. Further details on chemical characteristics of the raw materials, mixture proportions, and mixing and curing conditions can be found in [146, 154]. The paste mixtures were cured in moist conditions until their testing duration.

Table 5-1: Proportions (mass-based) of the UHP cementitious pastes employed in this study.

| Mixture | Constituent mass fraction in the binder | | | | w/b | Curing regime |
	OPC	Fly ash	Silica fume	Limestone		
UHP-1	0.70	0.175	0.075	0.05	0.20	Moist curing, 30d, 90d
UHP-2	0.50	--	0.20	0.30	0.20	Moist curing, 30d

5.2.2 Nanoindentation and Chemical Mapping

A brief description of the procedure for nanoindentation and chemical mapping of UHP-1 and UHP-2 pastes is described here. The sample preparation included specimen cutting, ultrasonication in isopropyl alcohol (IPA) [74], and polishing [50, 156, 73]. Nanoindentation was carried out using an Ultra Nanoindentation Tester (UNHT[3]; Anton Paar). Each sample had at least 1250 indents split among several grids in different locations to capture the heterogeneity in the microstructure of multi-component UHP paste systems. Indentations were performed in force control mode with a maximum displacement cutoff of 250 nm (0.25 µm) with loading profile detailed in [146, 154]. This depth corresponded to an interaction volume idealized as a hemisphere with a radius 3 to 5 times the maximum cutoff [69, 107, 143]. The hardness (H) and the effective Young's Modulus (M) were determined following the Oliver and Pharr method [65, 66].

The specimen surfaces were imaged after the nanoindentation tests using a SEM (SNE-4500M Plus) coupled with EDS (Bruker EDS with ESPRIT software). The application of SEM-EDS for compositional identification of cement hydration phases is discussed in [68, 30]. Back-scattered electron (BSE) mode imaging Figure 5-1(a)) was performed with a beam current of 110 µA, a working distance of ~10 mm, and an accelerating voltage of

15 keV [154]. A BSE image was taken over the grid before EDS was performed at 50

kcps. It has been shown that, in cementitious materials, most of the characteristic X-rays

escaping the material are generated within a depth of 2 μm [72, 69], which is in line with

the interaction depth for nanoindentation. To relate the elemental EDS information to the

nanomechanical data, a MATLAB localization algorithm was implemented to align the

optical image of the nanoindentation grid to the EDS chemical maps, as detailed in [109]

[154] [155]. Brightness of the EDS chemical maps was auto-scaled by the data-collection

software. Figure 5-1(b) shows the Ca EDS map. Al, Si, and Fe maps were similarly

obtained. Across different maps, the number of X-ray counts associated with the same

brightness value varies, and hence EDS maps are qualitative measures of the

concentration of elements in each indentation grid. For statistical analysis, the RGB

intensities from the Al, Ca, Fe, and Si EDS maps (denoted as $I_{Al}, I_{Ca}, I_{Fe},$ and I_{Si}

respectively) were matched with the corresponding nanomechanical data. Figure 5-1(c)

illustrates the translation of EDS map color intensity of Ca to the 0-255 scale. In BSE

imaging, the cube of the brightness (γ^3) can be related to the density of the phase [67].

This local density information is also used as an input parameter in the ML models

described later.

102

(a)   (b)   (c)
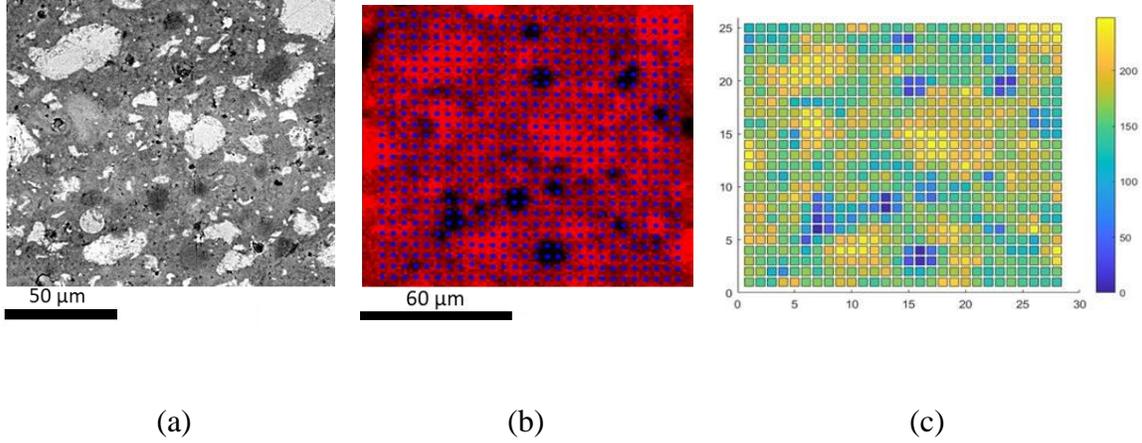
Figure 5-1: (a) BSE image of the 30-day UHP-1 paste, (b) Ca EDS map with blue dots added to show the location of one of the indentation grids after the alignment procedure, and (c) MATLAB graphic translating EDS map color intensity into 0-255 scale for Ca.

5.2.3   Statistical Cluster Analysis from SEM-EDS and Nanoindentation Data

To generate the labels of the constitutive microstructural phases to train the ML classification models, a Bayesian Information Criterion (BIC) with negative log likelihood method was implemented for statistical deconvolution (clustering) of the chemical intensities and the micromechanical properties [109]. If there exists n phases in the microstructure with each phase occupying a volume fraction of ϕ_i ($i = 1…n$) such that $\sum_{i=1}^{n} \phi_i = 1$, the properties of each phase can be approximated by a Gaussian distribution with a probability density function (PDF) given as:

$$PDF = \sum_{i=1}^{n} \phi_i \, \psi_i \tag{27}$$

Here, ψ_i is the vector of classification variables of the phase. The classification variables utilized in cluster analysis were indentation modulus M, indentation hardness H, and the intensities of aluminum I_{Al}, calcium I_{Ca}, iron I_{Fe}, and silicon I_{Si}. While the same statistical nanoindentation results can be fit using different number of phases and volume

103

fractions [108], a maximum negative log likelihood estimation was used to find the PDFs that best represented the experimental data:

$$\text{NlogL} = -\max(\log(\textstyle\prod_n \text{PDF}(n_i))) \tag{28}$$

Here, n_i represents the distribution parameters, in the case of a Gaussian distribution the mean and standard deviation, that are iterated to maximize the likelihood function. Then, the BIC was minimized such that:

$$\text{BIC} = 2\,\text{NlogL} + p\log(m) \tag{29}$$

In the above equation, m is the number of indentation points and p is the number of identifying parameters available at each indentation point (in this case six; four chemical intensities and two mechanical properties M and H) [109]. A summary of the constitutive phases identified from this clustering analysis is given in Table 5-2. They include low density (LD) C-S-H, high density (HD) C-S-H, an ultra-high stiffness (UHS) phase unique to the very low w/b cement pastes such as UHP mixtures, a mixed phase comprised of partially reacted starting materials such as fly ash or limestone and products such as carboaluminates, and residual clinker. The salient features of these phases have been elucidated in detail elsewhere [154] [69] [71] [39] [35]. As an example for the UHP-1 paste cured for 90 days, the clustering of M and H is shown in Figure 5-2(a), while Figure 5-2(b) depicts the normalized intensities of Ca at every indentation point and the corresponding M, and Figure 5-2(c) showcases the normalized intensities of Ca vs. Si. Detailed analysis of the UHP paste clusters identified, and justification for their corresponding constitutive phase labels are described in [154].

Table 5-2: Constitutive phases identified and their volume fractions (ϕ) in the UHP pastes. FA, MS, L, and CA denotes fly ash, microsilica, limestone, and carboaluminates, respectively. The phase labels from 0-4 are the inputs to the ML classification algorithm.

Mixture	Phase	Phase Label	Volume fraction (ϕ)	
			30 d	90 d
UHP-1	LD CSH/Residual MS	0	0.18	-
	HD CSH	1	0.38	0.40
	UHS Phase	2	0.19	0.23
	Mixed (FA, L, MS, CA)	3	0.12	0.17
	Clinker	4	0.13	0.20
UHP-2	UHS Phase/CSH	2	0.42	-
	Mixed (L, MS)	3	0.41	-
	Clinker/Unreacted	4	0.17	-

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
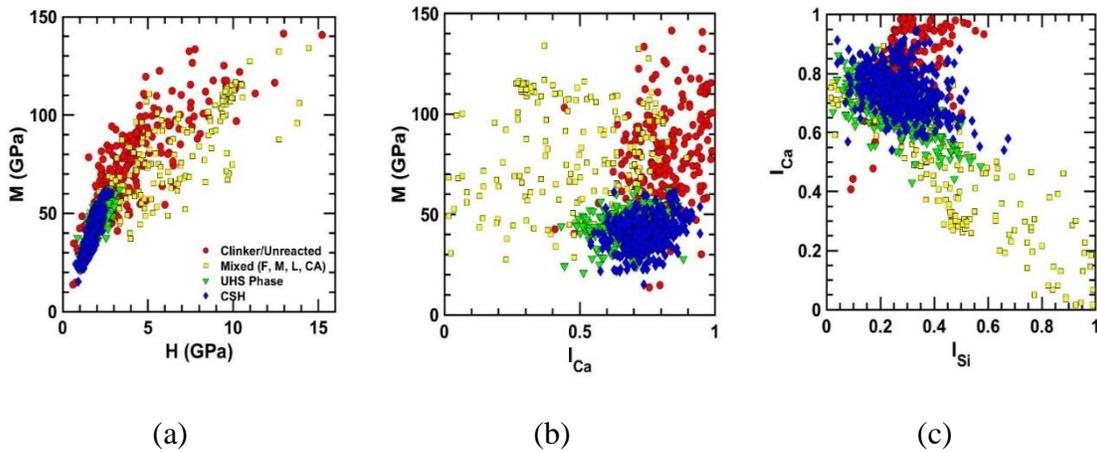
Figure 5-2: Clustering analysis of the 90-day cured UHP-1 paste: (a) M vs. H, (b) M vs. I_{Ca}, and (c) I_{Ca} vs. I_{Si}.

5.2.4 Inputs to the Machine Learning Classification Model and the Rationale

The ML classification models described in the forthcoming section uses the intensities at different indentation points to determine which of the phases (shown in Table 5-2), the point belongs to. The datasets for both mixtures and ages shown in Table 5-1 were combined to create the most generalizable ML classifier possible. The details of this large

dataset are shown in Table 5-3. The ability of ML algorithms to accurately classify the constitutive phases in complex microstructures belonging to multiple mixtures at different ages is explored. For the first set of ML models, 7 inputs were used (i.e., the 4 chemical intensities, γ^3, and M and H values from nanoindentation). In the second set, the ML models were trained only using 5 inputs (i.e., the 4 chemical intensities and γ^3). M and H are used as inputs in one set of ML models since the actual nanomechanical information is expected to facilitate better learning of the ML models to identify the phases during the training stage. This is shown to be true later in this paper, especially for more complex microstructures such as the UHP pastes. To test the correlation between the predicted phase labels and the 7 inputs, Pearson correlation coefficients (or linear correlation coefficients) [11] [16] were determined as shown in Figure 5-3. It can be noticed that all the inputs are reasonably correlated to the phase label output. M and H have the greatest correlation with the phase labels, and all the chemical intensities are quite similarly related to the phase label output. The high correlation between the phase label output and M and H means that the efficiency of ML classification models that uses only chemical intensities from SEM-EDS (which is the preferred approach, since this data is easier to obtain than M and H) could suffer, which is evaluated in this paper. Generating ML models with and without nanoindentation data provides quantification of the tradeoff of only including SEM-EDS data as inputs.

Table 5-3: Details of the input dataset for the ML models, including γ^3, H, M, and RGB intensities of Al, Ca, Fe, and Si.

Dataset (see Table 5-1 for mixture details)	No. of data points	Statistic	I_{Al}	I_{Ca}	I_{Fe}	I_{Si}	γ^3	H (GPa)	M (GPa)
Combined dataset belonging to UHP-1 @ 30d, 90d and UHP-2 @ 30d	3476	Max	252	252	252	252	1.47×10^7	23.20	235.54
		Mean	49	159	81	57	2.18×10^6	3.45	56.92
		Min	4	4	4	4	6.85×10^3	0.43	12.87
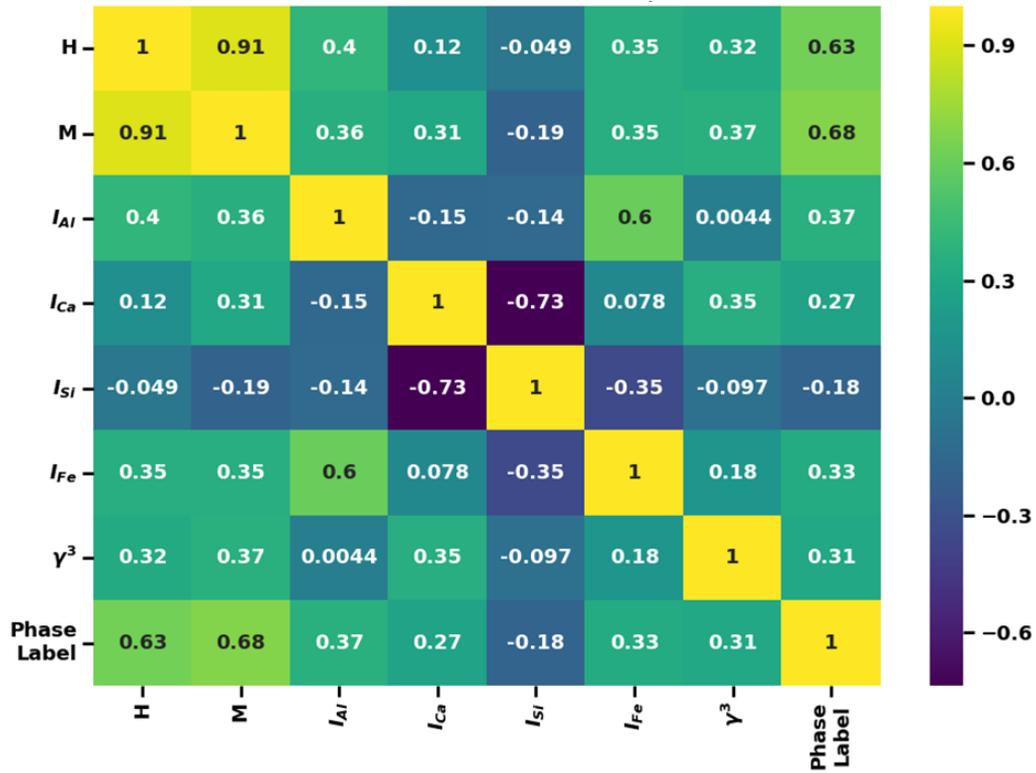
Figure 5-3: Pearson coefficient heat map for the correlation between the 7 inputs and the phase label output.

5.3 MACHINE LEARNING AND DATA PROCESSING

The different machine learning (ML) techniques used for classification, along with the data pre-processing and parameter optimization methods, are summarized here.

5.3.1 Machine Learning Techniques

Artificial neural network (ANN) and forest ensemble methods are the ML algorithms used for the multi-classification (i.e., more than 2 classes or phase labels) reported in this paper. ANNs can learn very complex patterns of data, and thus is a preferred ML algorithm for many materials-related problems [9] [12] [95] [14] [157]. The ANNs used in this study utilize 2 to 3 hidden layers, which are appropriate for the number of unique data records used. The chosen activation function to relate neurons [95] is the rectified linear unit (ReLu) with optimization performed using RMSprop, which features an adaptive learning rate formula [96]. Backpropagation, using the gradient of the previous iteration to train the weights of the ANN, was performed automatically by the Keras neural network framework written in Python to build and train the ANNs [158]. To minimize over-fitting, a dropout rate, i.e., the probability that any neuron and its connections will be temporarily excluded from the network, was incorporated into the ANN [97].

Machine learning forest ensemble methods are based on the structure of a decision tree that finds logical splits in the data leading from one branch to the next until ending at the leaf node [9, 87]. To reduce prediction inaccuracy and over-fitting, the predictions from a collection of decision trees are bagged [9] [159], termed ensembles. A basic form of forest ensemble is the Random Forest (RF) method in which the best split of the data is determined by considering all of the input features and checking a criterion, such as Gini

impurity, to select the most discriminative threshold [92, 87]. Each individual decision

tree in the RF ensemble does not use the entire set of training data, but a bootstrap sample

made from subsets of the training data with replacement [87] [159]. Another forest

ensemble is the Extra Trees (ET) regressor in which the splits are drawn at random for

each feature and the best split, as measured by the chosen criteria, is selected as the

splitting rule [92, 87]. In the ET regression model, the entire dataset is incorporated into

each individual tree [92]. The prediction results of the individual trees are averaged to

produce the output prediction in the RF and ET regressions. In a Gradient Boosted Tree

(GBT) ensemble, an initial tree is trained with the entire dataset. All subsequent trees in

the forest are trained to minimize the residual between the predicted and actual values of

the previous tree [9, 92] [93]. The final prediction is calculated as the weighted sum of

the predictions of each tree. For each tree beyond the first, the prediction is multiplied by

the learning rate, with typical values between 0.01 and 0.1 [9, 92]. A specialized form of

the GBT is Extreme Gradient Boosted (XGB) tree [93]. XGB performs shrinkage and

column subsampling techniques to prevent overfitting between boosted trees and

additionally offers scalability through parallel tree boosting (efficient computing

regardless of data size) [93].

5.3.2 Preprocessing and Evaluation

The input data points were pre-processed before separation into the testing and training

sets to ensure that all the inputs and outputs lie in the range [0, 1] such that:

$$z_{new} = \frac{z - z_{min}}{z_{max} - z_{min}} \tag{30}$$

Here, z_{new} is the value of the variable after transformation, z is the current value of the variable, and z_{min} and z_{max} are the minimum and maximum values respectively, of that variable.

The dataset mentioned in Table 5-3 was shuffled along the rows of indentation points (Figure 5-1(b)) such that adjacent points were separated, providing a greater chance of equal distribution of the various microstructural entities within the testing and training datasets. Training was performed by fitting the ML algorithm to the training dataset and allowing the algorithm to adjust its internal features to minimize the error. Model performance was evaluated using the testing dataset, which the ML algorithm has not yet seen, and measuring the resulting errors. To evaluate the accuracy of the ML predictions, a stratified n-fold cross-validation technique was employed [9, 87, 11]. Stratified splitting refers to preserving the percentage of samples in each class within each fold [92]. A 3-fold cross-validation, deemed sufficient for the size of the datasets, was performed using the following steps: (i) randomizing the dataset and performing a 3-fold stratified split, (ii) training the model using 2 of the folds, (iii) testing the model using the remaining fold, (iv) repeating steps (ii) and (iii) until each fold has been used for testing once, acquiring 3 independent performance measures, and (v) averaging the individual metrics measured to obtain the cross-validation value.

Among the several assessment methods for ML-based classification [160], the area under the Receiver Operator Characteristic curve (ROC-AUC) is chosen here since it is an important metric for checking any classification model's performance [160] [161] [162]. A ROC-AUC of 1.0 indicates most accurate classification. The ROC curve is created by plotting the true positive rate (also called sensitivity or recall) against the false positive

110

rate (also called false alarm rate or fallout) at various threshold settings [160] [161]. The

ROC-AUC is a measure of how well a model can discriminate between two classes (or

microstructural phase labels, in this case), and is insensitive to the changes in the class

distribution [160] [161]. In the case of multi-class labeling, however, ROC-AUC can be

calculated using two different methods. The One-versus-Rest (OvR) strategy calculates

the model's ability to discriminate between one class vs. the rest of the classes, while the

One-versus-One (OvO) strategy pairs each class against another such that, for n phase

labels, $\frac{n*(n-1)}{2}$ calculations are made [162]. The former is sensitive to class distribution

changes [161] while the latter is insensitive to class distribution, but computationally

more expensive when the class number increases. In this study with 5 phase labels and

data that is not significantly imbalanced, which requires special class distribution

considerations [163], the more general OvR method was employed. The multi-class

dataset was one-hot encoded (i.e., represented as binary vectors), and a ML classifier

trained to predict the probability that a data point belonged to each phase label. The phase

label with the highest probability is taken as the prediction for each point. In training, the

goal of the ML models was to maximize the objective function, which was the ROC-

AUC. Other metrics tracked, but not used to train the models, were the accuracy and the

F1 score, given as [160]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{31}$$

$$F1 = \frac{2TP}{2TP+FP+FN} \tag{32}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the

number of false positives, and FN is the number of false negatives, predicted by the

model for each class (or phase label). True positives indicate the success in identification of the correct phase label.

5.3.3 Hyperparameter Optimization

For all the models, the parameters which maximized the 3-fold cross-validation ROC-AUC were used as the basis for the final models, with some additional fine-tuning. The parameters to optimize in the ANN models were the number of hidden layers, the number of neurons in each hidden layer, and the dropout rate. ReLu activation function with a learning rate of 0.001 and an RMSprop optimization scheme was used. For the RF, ET, and GBT models, the number of trees in the forest, the maximum depth of the trees, the minimum number of samples before splitting, and the minimum number of samples per leaf were tuned. Coarse optimization of the hyperparameters for ANN and the forest ensembles followed a random search pattern, found to be the most efficient method to optimize parameters [164], by randomly generating 20 different combinations of hyperparameters. The hyperparameters for random testing were chosen from the uniform distributions shown in Table 5-4.

For the XGB models, there are many hyperparameters available to tune, nine of which were chosen for this study. The hyperparameters range from structure-based, such as the depth of the trees or the number of GBTs, to how splits are made via the subsample and colsample_bytree parameters, or even how big the leaf groups can be via min_child_weight. Additional parameters tuned included the learning rate, the minimum objective function loss required to split a leaf node called gamma, as well as the L1 and L2 regularization terms on the weights called alpha and lambda, respectively. Each hyperparameter was tested one at a time over a grid within the range of values indicated

in Table 5-4, where the best value was used when searching for the next parameter. The order of hyperparameter selection is given by the order of parameters in Table 5-4 for XGB. This process was continued until the end when several different learning rates and number of trees were tested as a final tuning effort. Detailed breakdown of the allowed ranges and significance of each of these hyperparameters are given in the XGB code documentation [93].

Table 5-4: Hyperparameters tuned based on a uniform distribution range of potential values.

Model	Hyperparameter	Uniform Distribution Range
	# hidden layers	[1, 4]
ANN	# starting neurons	[20, 75]
	Drop rate	[0, 0.3]
	# of trees	[50, 400]
Random Forest (RF), Extra Trees (ET) Forest, and Gradient Boosted Trees (GBT)	Maximum depth	[3, 21]
	Minimum# of samples before split	[2, 25]
	Minimum # of samples on leaf	[1, 10]
	# of trees	[0, 500]
	Maximum depth	[1, 9]
	min_child_weight	[1, 6]
	Gamma	[0, 0.8]
XGB	Subsample	[0.5, 1.0]
	Colsample_bytree	[0.2, 1.0]
	Alpha	[1E-5, 1]
	Lambda	[1E-5, 1.05]
	Learning Rate	[0.05, 0.3]

5.4 ML-BASED CLASSIFICATION OF CEMENTITIOUS PHASES

5.4.1   UHP Pastes

The predictive efficiency of the different ML models using SEM-EDS data with and without nanoindentation hardness (H) and stiffness (M) as inputs, to classify the UHP phase at each desired location is reported in this section. Each of the five ML algorithms (ANN, RF, ET, GBT, XGB) discussed above were implemented on the data to examine the applicability of the ML classification methodology to identify the phase labels in complex and heterogeneous UHP pastes. Table 5-5 lists the ROC-AUC, accuracy, and F1 values for the final ML classification models for the 7-input and 5-input cases. The bolded entries indicate the ML models where the OvR ROC-AUC results from 3-fold cross-validation were the highest. Note that the 3-fold cross-validation trials could not be plotted directly, instead, Figure 5-4 and Figure 5-5 were generated from a 75%/25% data split such that 75% of the data points were used for training and 25% were used for testing and displaying the plots, where the results were almost identical to the 3-fold cross-validation results reported in Table 5-5.

In the case of the datasets with 7 inputs (both SEM-EDS and nanomechanical data), all the ML models performed very well in terms of all three metrics (ROC-AUC, Accuracy, and F1), with the GBT model showing a slightly better performance. The ROC-AUC value was around 0.99 (1.0 being the absolute best) [160] [161], indicating the efficiency of the classification algorithms in being able to determine the phase labels based on the given input data. Even when the nanomechanical data was removed from the datasets and the input matrix reduced to 5 SEM-EDS input parameters, the ML classification

114

algorithms worked quite well with a ROC-AUC value of around 0.92. In this case, the

ANN model provided the best ROC-AUC value, while the forest models also showed

very similar performance. The high ROC-AUC values show that, in an OvR setting, the

classification ML algorithms used are successful in distinguishing one class compared to

all the others. However, it can be also seen from Table 5-5 that there is a sharp reduction

in the accuracy and F1 scores, which both depend on the number of correctly identified

data points as described using Equations 5 and 6 [160], when the nanomechanical

information is absent. This is to be expected, since M and H had the highest correlation

with the output phase label, as indicated in Figure 5-3. It is observed that high accuracy

and F1 values, along with high ROC-AUC, can be achieved when additional, relevant

input data such as M and H are available.

Table 5-5: Efficiency metrics of the ML classification algorithms for phases in UHP
pastes from SEM-EDS (5 inputs), and with two additional inputs, M and H, from
nanoindentation (7 inputs). Average and standard deviation from 3-fold cross-validation
is reported. The ML model with the greatest ROC-AUC for each number of inputs is
shown in **bold**.

# of inputs	Model Type	ROC-AUC	Accuracy	F1
7	ANN	0.988 ± 0.003	0.906 ± 0.009	0.912 ± 0.009
	Random Forest	0.988 ± 0.003	0.903 ± 0.010	0.911 ± 0.010
	Extra Trees Forest	0.986 ± 0.003	0.889 ± 0.015	0.898 ± 0.014
	Gradient Boosted Trees	**0.989 ± 0.002**	**0.908 ± 0.011**	**0.914 ± 0.012**
	XGB	0.988 ± 0.003	0.907 ± 0.017	0.914 ± 0.016
5	**ANN**	**0.926 ± 0.002**	**0.726 ± 0.013**	**0.745 ± 0.014**
	Random Forest	0.924 ± 0.002	0.728 ± 0.010	0.749 ± 0.012
	Extra Trees Forest	0.924 ± 0.003	0.715 ± 0.012	0.729 ± 0.013
	Gradient Boosted Trees	0.919 ± 0.004	0.719 ± 0.015	0.746 ± 0.016
	XGB	0.921 ± 0.003	0.721 ± 0.008	0.743 ± 0.010

Figure 5-4(a) and (b) show the ROC curves obtained from these best-performing models for the 7-input and 5-input cases, respectively. As expected, and shown in Table 5-5, the ROC curves shift downward when the nanomechanical inputs are excluded from the ML classification analysis. However, it is important to note that not including M and H, which correlated the most with the phase label output (see Figure 5-3), still produces reasonable identification of the microstructural phases just based on SEM-EDS information. This is significant in that, the use of SEM-EDS chemical maps along with a ML classification scheme allows for: (i) identification of potential phases present at those locations, which provides detailed insights into the influence of material composition on microstructure, and (ii) prediction of important paste properties (such as modulus) based on the known properties of the phases and their volume fractions.

(a)   (b)
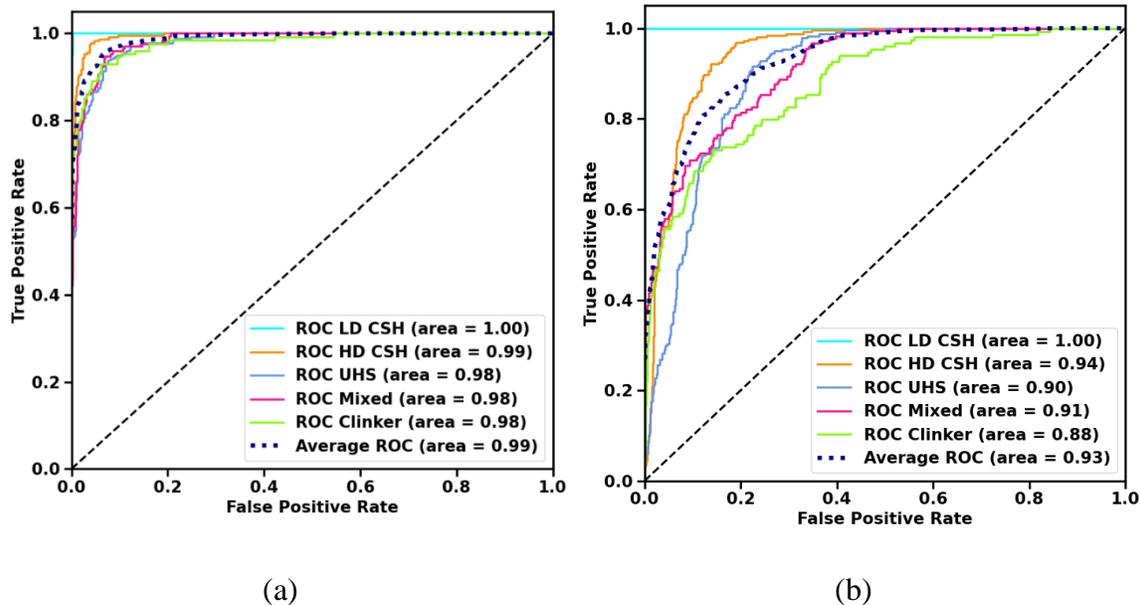
Figure 5-4: Receiver-Operator Curves (ROC) showing One-versus-Rest results for ML classification using 25% of data for testing: (a) GBT ML model with 7 inputs, (b) ANN ML model with 5 inputs. The dashed diagonal line represents the random guess of a class. The chosen models are the best performing ones based on Table 5-5.

116

Further information on the predictive performance of the classification models can be gleaned from confusion matrices presented in Figure 5-5(a) and (b) for the 7-input GBT ML and 5-input ANN ML classification models, respectively. For both the input types the LD C-S-H phase is accurately identified in 94-100% of the points by the ML models as shown in Figure 5-5. Similarly, the HD C-S-H phase is correctly classified in 83-94% of the points, depending on whether the 5-input or 7-input models are used. Since LD C-S-H and HD C-S-H have differences in their packing densities, which result in different mechanical properties [57, 116], it is only natural that a ML model that is trained using nanomechanical data also shows near-perfect capability in accurately identifying these phases. However, cluster analysis in several past work [71] [59] [165] have shown dissimilarities in chemical intensities between these phases, which enables the 5-input model also to perform satisfactorily in classifying these phases. As indicated in the authors' recent work [146] [154], the remaining three hard-stiff phases, viz., UHS, mixed phase containing limestone, carboaluminates and fly ash, and clinker, with indentation moduli of ~43 GPa [57], ~75 GPa [21] [39] [114] [38], and ~100 GPa [47] respectively, overlap in terms of chemical intensities and stiffnesses. This is clearly noticed in the scatter of points corresponding to these phases in Figure 5-2(c). Reducing the number of inputs from 7 to 5 clearly has a significant adverse effect on the classification of these phases as noted from Figure 5-5. In the 7-input model, the mixed phase is correctly identified in ~92% of the cases, while the classification accuracy drops down to ~63% in the 5-input model, where the mixed phase is confused with the UHS phase in many instances. In both the models, clinker classification has the lowest accuracy. In the 5-input model, the clinker classification accuracy is around 50%, with a significant number

of clinker locations mis-identified as HD C-S-H phases due to the absence of corroborating nanoindentation data. It is also notable that the EDS chemical maps were obtained based on qualitative measurements and not on quantitative spot chemical analysis [109, 72], and therefore only provide relative atomic ratios and not the exact ratios. As such, it is likely that cementitious phases with similar Ca/Si ratios, but different stiffnesses may be confused for one another in the 5-input ML model. Another explanation for the confusion between the clinker and HD C-S-H phases is that, in the UHP-2 mixture there was no HD C-S-H cluster identified, and the reaction product belonged to the UHS phase [154]; however, when the Ca and Si intensities were plotted for the clinker and UHS phases, they almost perfectly overlapped [154]. The high unreacted limestone content in this mixture could have resulted in excess Ca in the chemical map that contributed to a higher Ca/Si ratio, which is typical of clinker. This may have led to the confusion of the ML to differentiate between the clinker and UHS/HD C-S-H phases for the UHP-2 mixture. It is once again shown that, in complex microstructures where chemical intensities overlap between phases (as shown Figure 5-2(c)), the use of additional inputs in the form of nanomechanical properties help classification significantly.
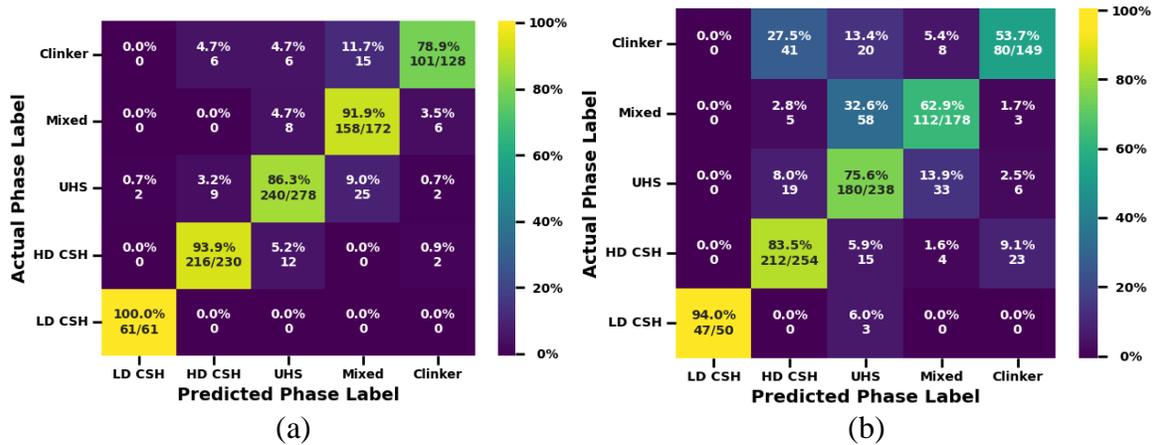
Figure 5-5: Confusion matrices showing results for ML classification using 25% of data for testing: (a) GBT ML model with 7 inputs, and (b) ANN ML model with 5 inputs. Percentage accuracy in each row is given based on the total number of data points in each phase label, as shown along the diagonal. In an ideal case, it is desirable to have a classification accuracy near 100% on the boxes along the diagonal, which would result in little to no misidentification, and thus, close to 0% on all the other boxes.

5.4.2 Validation of the Classification Approach Using Other Cement Paste Data

To validate the ML classification of microstructural phases through chemical intensities from SEM-EDS, two new datasets were curated from literature [72] [70] and similar ML models developed to classify their phases. These datasets are referred to as OPC (plain cement paste) [72] and NP (20% of cement by mass replaced with a natural pozzolan) [70]. Nanoindentation and chemical mapping data reported in [72] [70] identified several clusters of microstructural phases in these mixtures. The OPC data identified 5 clusters by BIC and negative log likelihood method in [72], however two clusters with the highest stiffness and hardness could be grouped together as part of the clinker phase to ensure that the same ML algorithms as described above can be used here. The remaining three clusters were labeled as LD C-S-H, HD C-S-H, and a mixed phase. For the NP data, 6

119

clusters were identified in [70], but clusters 5 and 6 were grouped as together as they were both identified as clinker phases [70], with LD C-S-H, HD C-S-H, UHS, and mixed phase labels given to the remaining clusters. The only available inputs were three elemental intensities, I_{Ca}, I_{Si}, and I_{Al}, along with M and H. Thus, ML models using all the 5 inputs, or just the 3 chemical signature inputs, were implemented. To keep the discussions succinct, only three forest ensemble models (RF, ET, and GBT) are used here for the validation tests. Table 5-6 lists the resulting ROC-AUC, accuracy, and F1 values for these datasets. Similar to the UHP pastes, there was a decrease across all metrics of classification going from 5 inputs (which included the micromechanical M and H) to 3 inputs. However, this decrease was to a much lesser extent owing to the greatly reduced complexity in these microstructures that were well hydrated. As compared to the UHP pastes evaluated in the previous section, these pastes demonstrate reduced heterogeneity with fewer starting ingredients, proportioned using a higher w/b, and having undergone higher degrees of reaction, as explained in detail in [155]. As shown by the results in Table 5-6, ML classification methods are quite successful in identifying the hydration phase given only the chemical intensities for less complex microstructures.

Table 5-6: Efficiency metrics of the ML classification algorithm for phases in OPC and NP pastes from SEM-EDS (3 inputs), and with two additional inputs, M and H, from nanoindentation (5 inputs). Average and standard deviation from 3-fold cross-validation is reported. The most accurate ML model for each number of inputs is shown in **bold**.

Dataset	# of inputs	Model Type	ROC-AUC	Accuracy	F1
OPC	5	Random Forest	0.975 ± 0.010	0.888 ± 0.018	0.893 ± 0.018
		Extra Trees Forest	**0.981 ± 0.005**	**0.897 ± 0.011**	**0.899 ± 0.011**
		Gradient Boosted Forest	0.968 ± 0.011	0.858 ± 0.043	0.867 ± 0.037
	3	Random Forest	0.951 ± 0.011	0.808 ± 0.062	0.801 ± 0.064
		Extra Trees Forest	**0.958 ± 0.012**	**0.829 ± 0.040**	**0.837 ± 0.037**
		Gradient Boosted Forest	0.945 ± 0.017	0.817 ± 0.040	0.827 ± 0.031
NP	5	Random Forest	0.988 ± 0.006	0.891 ± 0.023	0.891 ± 0.019
		Extra Trees Forest	0.989 ± 0.007	0.902 ± 0.038	0.897 ± 0.039
		Gradient Boosted Forest	**0.991 ± 0.006**	**0.925 ± 0.026**	**0.925 ± 0.022**
	3	**Random Forest**	**0.973 ± 0.008**	**0.860 ± 0.019**	**0.859 ± 0.022**
		Extra Trees Forest	0.973 ± 0.006	0.832 ± 0.041	0.826 ± 0.039
		Gradient Boosted Forest	0.965 ± 0.006	0.822 ± 0.028	0.819 ± 0.032

Figure 5-6 shows the confusion matrices and ROC curves for the OPC and NP mixtures, for the 3-input cases. The classification accuracy is very high as noted from the confusion matrices for both the pastes, attributable to the relative simplicity of their microstructures as compared to the UHP pastes. There are very few mis-labeled indentation points even when the nanomechanical data is not provided. The results show the application of ML-based classification algorithms in labeling the microstructural phases in cementitious systems.
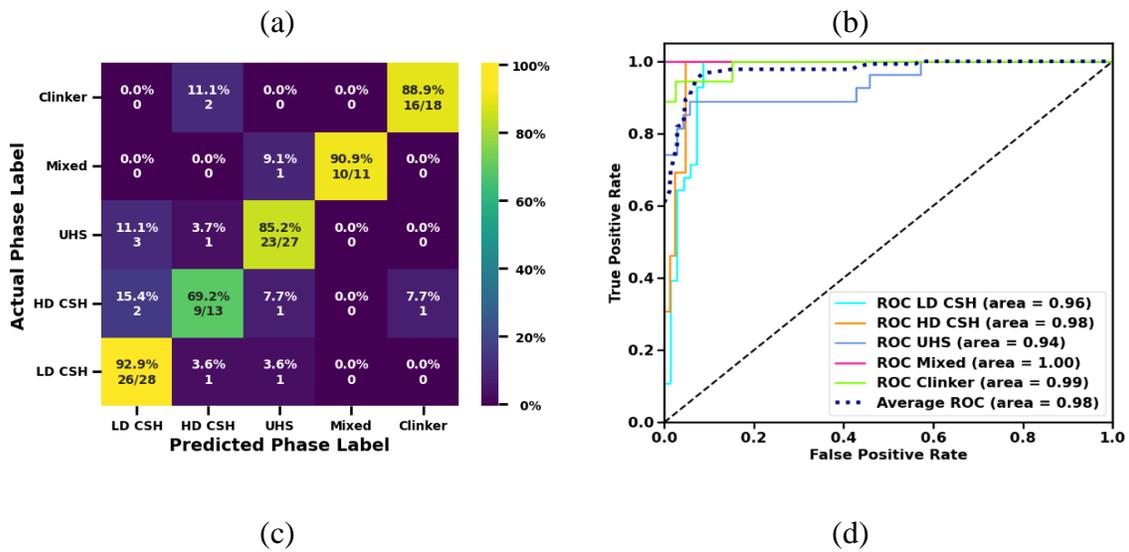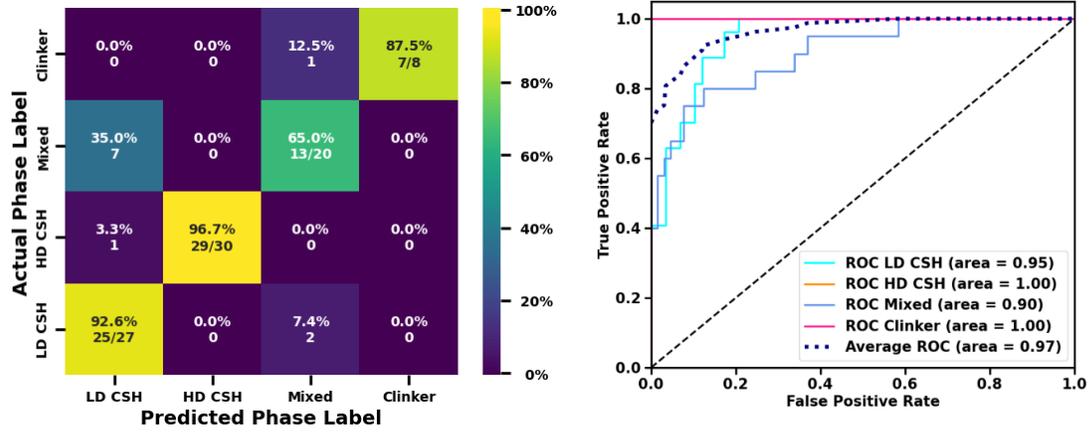
Figure 5-6: ML classification using 25% of data for testing: (a) and (b) confusion matrix and ROC curves for the 3-input ET model for the OPC paste; (c) and (d) confusion matrix and ROC curves for the 3-input RF model for the NP paste.

Final hyperparameter details for each data set are given in Table 5-7 for the ANN models, Table 5-8 for the ensemble models, and Table 5-9 for the XGB models. A reminder in ANN models that for each subsequent hidden layer the number of neurons is halved. For example, an ANN which began with 50 neurons in hidden layer 1 will have 13 neurons in hidden layer 3.

Table 5-7: Final hyperparameters of ANN models for each dataset to classify cementitious phases given qualitative chemical intensity and/or nanoindentation hardness and stiffness.

Dataset	# of inputs	# of Hidden Layers	Starting # Neurons	Dropout	Activation	Optimization	Learning Rate	Epochs	Run Time (sec)
UHP	7	3	75	0.05	ReLu	RMSprop	0.001	400	387
UHP	5	2	67	0.15	ReLu	RMSprop	0.001	400	375

Table 5-8: Final hyperparameters of ensemble models for each dataset to classify cementitious phases given qualitative chemical intensity and/or nanoindentation hardness and stiffness.

Dataset	# of inputs	Ensemble	n_estimators (# of Trees)	max_depth	min_sample_split	min_samples _leaf	Run Time (sec)
UHP	7	RandomForestClassifier()	335	19	3	2	6.36
		ExtraTreesClassifier()	221	19	5	2	2.61
		GradientBoostingClassifier()	255	20	3	10	102.9
	5	RandomForestClassifier()	345	10	7	3	5.78
		ExtraTreesClassifier()	60	10	7	1	0.74
		GradientBoostingClassifier()	86	3	16	6	5.50
OPC	5	RandomForestClassifier()	283	13	5	1	2.70
		ExtraTreesClassifier()	285	13	3	1	1.76
		GradientBoostingClassifier()	300	25	5	8	8.36
	3	RandomForestClassifier()	127	3	15	3	1.21
		ExtraTreesClassifier()	124	14	13	1	0.89
		GradientBoostingClassifier()	62	15	17	6	1.97
NP	5	RandomForestClassifier()	110	16	6	3	1.03
		ExtraTreesClassifier()	100	16	6	3	0.75
		GradientBoostingClassifier()	345	16	20	9	11.69
	3	RandomForestClassifier()	144	15	5	3	1.54
		ExtraTreesClassifier()	141	15	5	3	1.08
		GradientBoostingClassifier()	55	13	23	10	2.09

Table 5-9: Final hyperparameters of XGB models for each dataset to classify cementitious phases given qualitative chemical intensity and/or nanoindentation hardness and stiffness.

Hyperparameters	Dataset	
	UHP	
# Inputs	7	5
n_estimators	32	34
max_depth	10	5
min_child_weight	1	2
gamma	0	0.1
subsample	1	0.8
colsample_bytree	1	0.8
reg_alpha	0	0.0095
reg_lambda	1	1.05
learning_rate	0.3	0.1
Run Time (sec)	1.61	1.92

5.5 SUMMARY

This chapter has presented a novel approach to accurately predict cement hydration

phases from chemical intensity maps, using machine learning (ML) methods. Chemical

intensity data from SEM-EDS for different UHP cement paste datasets representing

multiple cementing materials and hydration ages were combined. Micromechanical

information from nanoindentation as well the elemental intensities from qualitative EDS

maps were then coupled with Bayesian statistical clustering. With the phase labels (e.g.,

LD or HD C-S-H, clinker etc.) thus identified, different ML classification techniques

based on Artificial Neural Networks (ANN) and forest ensemble methods were

implemented on the dataset. The classification algorithms were implemented on the 5-

input dataset (chemical intensities of Ca, Si, Al, and Fe, along with the cube of the

brightness of the BSE image (γ^3)), and 7-input dataset (the above 5 inputs, and M and H

from nanoindentation). The area under the Receiver Operator Characteristic curve (ROC-

AUC) was chosen as the indicator of model performance.

It was observed that, for the combined dataset of the UHP pastes, the ROC-AUC values were higher than 0.90 for both the 7-input and 5-input datasets. In such complex systems, the use of additional inputs in the form of nanomechanical properties help classification significantly. The same approach was also used on two less complex microstructures (i.e., fewer starting materials and more complete hydration), one of a plain OPC paste and the other a paste with 20% OPC replaced using a highly reactive natural pozzolan. Here, normalized intensities of just the three chemical species (Ca, Si, and Al) were deemed sufficient (without nanoindentation data) to generate a highly accurate classifier. It is shown that chemical intensity mapping of microstructures, coupled with machine learning, can be used to accurately (in the case of common cementitious microstructures) classify the microstructural phases, which can lead to *apriori* property (e.g., stiffness) predictions.

CHAPTER 6[3]

MACHINE LEARNING APPROACHES TO PREDICT THE MICROMECHANICAL

PROPERTIES OF CEMENTITIOUS HYDRATION PHASES FROM

MICROSTRUCTURAL CHEMICAL MAPS

6.1 INTRODUCTION

The macroscale mechanical properties of cementitious materials that are important in

structural design possess a complex relationship to constitutive materials, hydration, and

other microstructural effects that are not well understood, nor physically modeled [9].

Enhanced mixture design that not only achieves desired mechanical properties with fewer

trial batches [10], but also optimizes the concrete mixture for cost and environmental

factors, can only be produced through a fundamental understanding of the relationship

between the properties of the constituent microstructural phases and the macroscale

response. In cementitious pastes the macroscale properties differ widely depending on the

degree of hydration, the amount of unhydrated cement, the type of hydration products (C-

(A)-S-H gels with varying local chemistry, various other crystalline phases such as

calcium hydroxide and ettringite), and the remaining amounts of additional reactants such

as fly ash and slag. The fundamental properties of such phases are typically determined

using nanoindentation techniques, which are quite challenging for cement-based

materials [78, 51, 104]. In general, a grid is applied to the area of interest in the

microstructure, and nanoindentation tests are performed at each point in the grid. In

combination with statistical analysis techniques, the micromechanical properties (since

---

[3] This chapter is derived from the publication: E. L. Ford, S. Kailas, K. Maneparambil, and N. Neithalath, "Machine learning approaches to predict the micromechanical properties of cementitious hydration phases from microstructural chemcial maps," *Construction and Building Materials,* no. 265, pp. 120647, 2020.

the region of influence of the indentation is of the order of 1-3 µm, the terms

"micromechanical properties" or "nanomechanical properties" are used interchangeably)

of individual phases are determined [70, 72, 51]. The use of such an approach to

determine the stiffness and hardness of the constituent phases in plain ordinary Portland

cement (OPC) pastes as well as those containing multiple cement replacement materials,

or high- and ultra-high performance pastes have been well-documented [111] [59] [69,

146, 154]. The nanomechanical response of individual phases, in conjunction with

homogenization models, have been used to predict the macroscale response of normal,

high, and ultra-high-performance concretes [50, 122]. The limitations of statistical

nanoindentation has resulted in its coupling with electron microscopy and chemical

mapping (using energy dispersive X-ray spectra, for instance) to identify the different

phases in the microstructure and give a chemical context to the nanomechanical

properties [70, 109].

While the challenges with nanoindentation technique are well-accepted, it has also been

shown that there are some relationships between the chemistry of the hydration products

in the microstructure and their micromechanical properties [25, 28, 29]. For C-S-H gels,

recent papers have shown plots that represent the nanoindentation stiffness as a function

of Ca and Si contents in the gel [72, 71]. However, these relationships are very complex,

not unique, and not easily modeled using physics-based or statistical models. This

necessitates the use of machine learning (ML) methods that model complex, multi-

dimensional relationships that cannot easily be fitted using traditional statistical and

regression methods [8]. Large data sets can be used to relate the nanoindentation stiffness

of phases to several chemical features of the microstructure through supervised learning.

ML models, when trained on high-quality data sets: (a) utilize the semi-empirical rules that inform the relationship between phase chemistry and phase properties, and (b) perform predictions on previously untrained data sets. A significant number of previous publications have dealt with the prediction of compressive strength of conventional, high-performance, or geopolymer concrete from mixture proportions using a variety of ML models [9, 10, 11, 12, 13]. Recent work has also elucidated the use of ML methods to predict the elastic modulus of concretes [166]. ML has also been utilized in a multi-objective optimization of concrete mixture proportions to satisfy desired strength, cost, and slump [14], and paired with nanoindentation mapping of mortar surfaces to reconstruct the constituent phases [15, 16].

As discussed earlier, being able to predict the nanoindentation stiffness of phases from energy dispersive X-ray maps on electron micrographs that portray the distribution of chemical species, is efficient and expeditious. If successful, this method opens up new avenues to easily upscale the properties of cementitious materials through homogenization. Prediction of mechanical properties from ML will advance the design and implementation of complex, heterogeneous cementitious mixes through minimization of expensive and time-consuming testing as well as shifting the focus from the macroscopic to microscopic results of mixture design. This paper, the first of its kind, is a step in that direction. ML models are developed using relative concentrations of chosen chemical species at the indentation points in the hydrated microstructure as inputs. Different cementitious pastes, ranging from simple, well-hydrated OPC pastes, to complex, multi-component blends that are poorly hydrated, are utilized to demonstrate the range of applicability of the developed ML models. The relative performance of

different ML models (e.g., Support Vector Regression, Gaussian Process Regression, Artificial Neural Network, and Ensemble methods) are brought out, along with the complexities in the microstructure that challenge the predictive ability.

6.2 A BRIEF INTRODUCTION TO MACHINE LEARNING TECHNIQUES USED IN THIS STUDY

A concise overview of the different machine learning (ML) techniques used in this paper is provided in this section.

6.2.1   Support Vector Regression (SVR)

Support Vector Regression (SVR) is a nonparametric technique that uses a regularization parameter C and kernels to transform predictor variables into higher-dimensional feature space, as well as threshold-based error to guide training [9, 14, 10]. SVRs have fast learning speed and noise-tolerating ability [14]. When the objective function is minimized (i.e., ε-insensitive loss), errors smaller than a threshold, ε, do not add to the overall error measure, while the training data outside the ε-tolerance will be used as support vectors to build the regression function [9, 10]. The regularization term C quantifies the penalization of data outside of the tolerance [14] (in other words, C controls the trade-off between achieving a low error on the training data and minimizing the norm of the weights). The kernel is used to quantify the Gaussian distance between the predicted inputs and those of the training data [9]. Each type of kernel has its own set of parameters which must be tuned via cross-validation to achieve the best prediction (See Section 6.4.2 for further details). The squared exponential kernel (also called the radial basis function, RBF) is used in this study [12].

6.2.2 Gaussian Process Regression (GPR)

Gaussian Process Regression (GPR) is a ML technique in which the known data y is fitted to a predictive Gaussian distribution $\mathcal{N}$, with mean µ and variance σ^2, represented in Equation 1 [12, 91]. The model uses a stochastic process to determine random variables having Gaussian distributions without any prior knowledge.

$$y = \mathcal{N}(\mu, \sigma^2) = \mathcal{N}(\mathbf{0}, [\mathbf{K} + \sigma_n^2 \mathbf{I}]) \tag{33}$$

The distribution is assumed to have a mean of zero, with a covariance (kernel) matrix $\mathbf{K}$ generated from an assumed covariance (kernel) function k, and additive independent Gaussian noise with a mean of zero and a variance of σ_n^2 multiplied by the identify matrix $\mathbf{I}$ of size $\mathbf{K}$. The distribution of known y and predictions f_* from new inputs x_* is a joint distribution written as:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} = \mathcal{N}(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K} + \sigma_n^2 \mathbf{I} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{bmatrix}) \tag{34}$$

where $\mathbf{K}_* = [k(x_*, x_1), \dots, k(x_*, x_M)]^T$ for M dimensions (different types) of inputs, and $\mathbf{K}_{**} = k(x_*, x_*)$ where x_* is input into the kernel function k. The mean μ_* and variance σ_*^2 of the prediction distribution f_* are therefore given as:

$$\mu_* = \mathbf{K}_*^T(\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1}y \tag{35}$$

$$\sigma_*^2 = \mathbf{K}_{**} - \mathbf{K}_*^T(\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1}\mathbf{K}_* \tag{36}$$

The mean of the new input data represents the prediction, and the standard deviation gives an error range of estimation, which marks GPR as one of the few ML techniques that provide an error range (confidence interval) for the predictions [167, 168]. Cholesky decomposition is used to determine the inverse of the covariance matrix $\mathbf{K}$. The assumed kernel function describes the structure and similarity of the data and therefore has a large

impact on the prediction results [12, 169]. The four types of kernels used in this study are the squared exponential (i.e., radial basis function; See section 6.2.1), the rational quadratic, the Matern 1.5, and Matern 2.5, whose formulations are given as [12]:

Squared Exponential/Radial Basis Function (RBF):

$$k(x, x^*) = \sigma_f^2 * e^{-\frac{r^2}{2\ell^2}} \tag{37}$$

Rational Quadratic:

$$k(x, x^*) = \sigma_f^2 * \left(1 + \frac{r^2}{2\alpha\ell^2}\right)^{-\alpha} \tag{38}$$

Matern 1.5:

$$k(x, x^*) = \sigma_f^2 * \left(1 + \frac{\sqrt{3}*r}{\ell}\right) * e^{\frac{-\sqrt{3}*r}{\ell}} \tag{39}$$

Matern 2.5:

$$k(x, x^*) = \sigma_f^2 * \left(1 + \frac{\sqrt{5}*r}{\ell} + \frac{\sqrt{5}*r^2}{3\ell^2}\right) * e^{\frac{-\sqrt{5}*r}{\ell}} \tag{40}$$

Here, r is the Euclidean distance between variables x and x^*, σ_f^2 is the variance of the output data y, ℓ is the characteristic length-scale, and α is the scale parameter (when $\alpha \rightarrow \infty$ the rational quadratic is identical to the squared exponential). The Matern kernels are generalizations of RBF and the parameter 1.5 or 2.5 controls the smoothness of the resulting function [92]. Each kernel has unique parameters that are automatically optimized via maximum likelihood estimation [91] and applied using "fit" on the scikit learn "GaussianProcessRegressor" command [92]. To further improve the predictive capabilities of the GPR models, the kernels are treated as additive in which each dimension (input variable) is given a base kernel to train on, and the sum of these kernels plus white noise is treated as the kernel to generate predictions [169]. A limitation of this

131

additive kernel technique is that, as more kernels are introduced, more parameters must be optimized and potentially overfit to the training data [169]. GPR is known to work well on sparse data sets [167] and is more effective when variables are reduced to only the most essential ones [170].

6.2.3 Artificial Neural Networks (ANN)

An artificial neural network (ANN) mimics synaptic responses in the brain and is organized into an input layer, hidden layer(s), and an output layer. Each layer contains neurons with values of information that typically ranges between 0 and 1 [95]. Between layers, neurons are fully connected to every other neuron. Each connection has a unique weight, w, relating the two neurons. The number of input layer neurons is equal to the number of inputs, the number of output layer neurons is equal to the number of outputs, and there can be any number of neurons in the hidden layer(s). The ANN can be shallow with only a few hidden layers, or deep with many layers. The exact number of hidden layers to be considered "deep" is not well-agreed upon, but generally networks with > 10 hidden layers are considered deep networks [83]. The ANNs presented in this study utilize 1 to 3 hidden layers, which are appropriate for the number of unique data records used, and the number of inputs are 3 or 6.

The value within each neuron of the hidden layer(s) and output layer depends on the previous neurons, the weights, and the chosen activation function [95]. Although a sigmoidal function is commonly used as the activation function [9, 95, 171], more accurate results were found in this study using a rectified linear unit (ReLu). Optimization is performed using RMSprop, which features an adaptive learning rate formula [96]:

132

$$v_t = \beta * v_{t-1} + (1 - \beta) \left(\frac{\partial f}{\partial w_t} \right)^2 \tag{41}$$

Here, v_t is the average of the squared gradient at time t, v_{t-1} is the average of the squared gradient of the previous iteration, β is the moving average parameter with a default value of 0.90, and $\frac{\partial f}{\partial w_t}$ is the gradient of the objective function f with respect to the weight at time t (w_t). This is then used to automatically adjust the initial learning rate, η, for each parameter when generating the updated weight, w_{t+1}:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{v_t}} * \frac{\partial f}{\partial w_t} \tag{42}$$

Utilization of the gradient of the previous iteration to train the weights of the ANN is known as backpropagation.

Neural networks feature many fitting parameters that allow them to predict nonlinear interactions. A disadvantage of neural networks is the potential for over-fitting the data, or training the weights to precisely match the training data set and render the algorithm unable to accurately predict results of the test data set [9]. To minimize over-fitting, a dropout rate was incorporated into the ANN. A dropout rate is a hyperparameter probability that any neuron and its connections will be temporarily excluded from the network [97]. While testing, the entire neural network is used, but the connection weights are multiplied by the dropout rate to combine the effect of the thinned-out training networks.

In this study, the Keras neural network framework written in Python is utilized to build and train the ANNs [158]. TensorFlow is used as the back-end engine. Additionally, a wrapper was implemented to use the scikit learn GridSearchCV and

RandomizedSearchCV functions [92] in order to determine and compare hyperparameter settings, as detailed in Section 6.4.2.

6.2.4   Ensembles

Machine learning forest ensemble methods are based on the structure of a decision tree. The decision tree finds logical splits in the data leading from one branch to the next until ending at the final leaf node. A simple operation such as multiplication is then used to predict the output of the partition [9, 87]. To reduce prediction inaccuracy and over-fitting, the predictions from a collection of decision trees are accumulated [9]. A basic form of forest ensemble is the Random Forest (RF) method in which the best split of the data into branches and nodes is determined by considering all of the input features and checking a criterion, such as mean-squared error, to select the most discriminative threshold [92, 87]. Each individual decision tree in the RF ensemble does not use the entire set of training data, but a bootstrap sample made from subsets of the training data with replacement [87]. Another forest ensemble is the Extra Trees (ET) regressor in which the splits are drawn at random for each feature and the best split, as measured by the chosen criteria, is selected as the splitting rule [92, 87]. In the ET regression model, the entire dataset is incorporated into each individual tree [92]. The prediction results of the individual trees are averaged to produce the output prediction in the RF and ET regressions. In a Gradient Boosted Tree (GBT) ensemble, an initial tree is trained with the entire data set. All subsequent trees in the forest are trained to minimize the residual (least squares) between the predicted and actual values of the previous tree via steepest descent [9, 92]. Steepest descent follows the negative gradient of the residual to

134

determine the learning rate of the next tree; values between 0.01 and 0.1 are typically

used as the learning rates [9, 92]. The final prediction is calculated as the weighted sum

of the predictions of each tree, where for each tree beyond the first, the prediction is

multiplied by the learning rate [9].

6.3 EXPERIMENTS AND DATA COLLECTION

6.3.1 Materials and Mixtures

In this study, experiments were performed on two ultra-high performance (UHP)

cementitious pastes, which have been studied in detail by the authors [146, 154, 18, 6].

These pastes are referred to as UHP-1 with a mortar 28 day compressive strength of ~145

MPa and UHP-2 with a mortar compressive strength of ~128 MPa [6] [41]. Both these

UHPs contain multiple cement replacement materials (Class F fly ash, silica fume, fine

limestone powder) of varying sizes and reactivity, and a low water-to-binder ratio (w/b),

thereby introducing high degrees of complexities in the microstructure. Hence, for

comparison, nanoindentation and chemical maps corresponding to relatively less complex

microstructures were extracted from recent literature [70, 72, 109]. These mixtures are

referred to as OPC (plain cement paste), NP (20% of cement by mass replaced by a

natural pozzolan), and SF (35% of oil well cement by mass replaced by silica flour).

Table 6-1 gives the proportions and details of all the mixtures evaluated in this study.

Further details on chemical characteristics of the raw materials, mixture proportions, and

mixing and curing conditions can be found in [70, 72, 146, 154, 109].

Table 6-1: Details of cement pastes with proportions given by mass which ML models are used to predict nanomechanical properties.

Mixture	Constituent mass fraction in the binder						w/b	Curing regime	Ref.
	OPC	Fly ash	Silica fume	Limestone	Natural pozzolan	Silica flour			
UHP-1	0.70	0.175	0.075	0.05	--	--	0.20	Moist curing, 30d, 90d	[146, 154]
UHP-2	0.50	--	0.20	0.30	--	--	0.20	Moist curing, 30d, 90d	[146, 154]
OPC	1.0	--	--	--	--	--	0.40	Moist curing, 1 year	[72]
NP	0.80	--	--	--	0.20	--	0.35	Moist curing, 90d	[70]
SF	0.65[*]	--	--	--	--	0.35	0.44	Hydrothermal curing for 6d at 60°C and 20 MPa	[109]

*The cement used in this case alone is a Class G oil-well cement.

6.3.2   Nanoindentation and Chemical Mapping

The procedure for nanoindentation and chemical mapping experiments carried out for UHP-1 and UHP-2 mixtures are described here. For the OPC, NP, and SF mixtures, similar procedures have been carried out as reported in [109, 70, 72]. These experimental techniques have been extensively reported by the authors and many other researchers, and hence only a brief description is provided here. After curing for the respective durations, sample preparation including specimen cutting, ultrasonication in isopropyl alcohol (IPA) [74], and polishing with silicon carbide abrasive paper from 240 grit to 1μm, were performed [50, 156, 73]. Nanoindentation grids were placed on regions in the microstructure with minimal pores or scratches and indented using an Ultra Nanoindentation Tester (UNHT[3]; Anton Paar). Each sample had at least 1250 indents

split among several grids in different locations to capture the heterogeneous microstructure of multicomponent blends. The linear loading profile had a loading and unloading rate of 12 mN/min with a pause for 5 seconds at the peak load when the maximum displacement cutoff was reached. Indentations were performed in force control mode with a maximum displacement cutoff of 250 nm (0.25 µm). This depth corresponded to an interaction volume idealized as a hemisphere with a radius 3 to 5 times the maximum cutoff [69, 107, 143]. The linear unloading portion of the force-displacement curve and the contact area of the tip to the sample surface was used to calculate the hardness (H) and the effective Young's Modulus (M), following the Oliver and Pharr method [65, 66].

The specimen surfaces were imaged using Scanning Electron Microscopy (SEM; SEC Nanoimage TableTop SNE-4500M Plus) coupled with Energy Dispersive X-ray Spectroscopy (EDS; Bruker EDS and ESPRIT software) after the nanoindentation experiments. The application of SEM and EDS for compositional identification of cement hydration phases is discussed in [68, 30]. Back-scattered electron (BSE) mode imaging (Figure 6-1(a)) was performed in low vacuum (~10^{-3} torr), with a beam current of 110 µA, a working distance of approximately 10 mm, sample tilt of 10°, and an accelerating voltage of 15 keV. A 1280 x 960 pixel BSE image was taken over the grid at a magnification of 500x before EDS was performed at 50 kcps over a 256 x 192 pixel image using a dwell time of 128 µs (~6 seconds per frame) for 5-15 minutes. The interaction volume and escape depth of the signal X-rays are dependent on the microscope parameters as well as the sample density and composition [72, 69]. It has been shown that in cementitious materials, most of the characteristic X-rays escaping the

material are generated within a depth of 2 µm [72, 69], which is in line with the

interaction depth for nanoindentation.

To relate the elemental EDS analysis to the nanomechanical data, a MATLAB

localization algorithm was implemented to align the optical image of the nanoindentation

grid to the EDS chemical maps. This algorithm employs image enhancement, indent

identification, and Hough transformation-based grid detection [109] to determine the

coordinates of the indents from the optical image. With a pixel length of 0.6 µm (256 x

192 pixels image) and 0.27 µm (1280 x 960 pixel image), an averaging filter of size 4x4

pixels and 9x9 pixels was taken over the indent coordinates on the EDS and BSE maps,

respectively, to account for the full 1.25-2.5 µm expected diameter of influence about

each indent. Brightness of the qualitative EDS chemical maps was auto-scaled by the

data-collection software. The RGB color brightness ranges from 0 to 255 with a higher

intensity signifying a higher concentration of the element.

Figure 6-1(b) is an example of the Ca EDS map. Al, Si, and Fe maps were similarly

obtained. Across different maps, the number of X-ray counts associated with the same

brightness value varies, and hence EDS maps are qualitative measures of the

concentration of elements in each indentation grid. For statistical analysis the averaged

auto-scaled RGB intensities from the Al, Ca, Fe, and Si EDS maps (denoted as

I_{Al}, I_{Ca}, I_{Fe}, and I_{Si} respectively) are matched with the corresponding nanomechanical

data.

Figure 6-1(c) illustrates the translation of EDS map color intensity of Ca to the 0-255

scale. In BSE imaging, the cube of the brightness coefficient (γ^3) can be related to the

density of the phase as shown in [67]. This local density information is also used as an

additional parameter in the ML models described later.

(a)   (b)   (c)
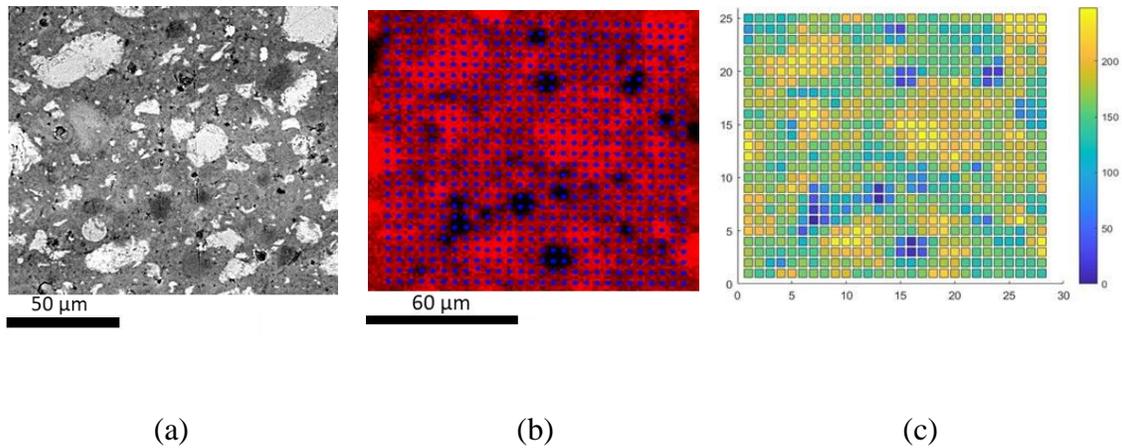
Figure 6-1: (a) BSE image of the 30-day UHP - 1 paste, (b) Ca EDS map with blue dots added to show the location of the indentation grid after the alignment procedure, and (c) MATLAB graphic translating EDS map color intensity into 0-255 scale for Ca.

A Bayesian Information Criterion (BIC) with negative log likelihood method was

implemented for statistical deconvolution (clustering) of the input chemistry and the

micromechanical properties [109]. The properties of each cementitious phase were

approximated by a Gaussian distribution and the maximum negative log likelihood

estimation was used to find the distribution that best represented the experimental data.

For the data obtained from other works (i.e., OPC, NP, and SF pastes in Table 6-1), the

mechanical and chemical clustering data were not fully aligned since we do not have

details on one-to-one mapping of the mechanical and chemical data. However, well-

known patterns in the relationship between identified clusters and their corresponding M,

H, and chemical intensity values were followed to accomplish this objective. For

example, in the C-S-H dominated clusters, an increasing M and H was taken to

corresponding to a decreasing Ca and increasing Si. The values from the Al vs. Si

intensity graphs were aligned with those from the Ca vs. Si intensity graphs by ordering

139

the extracted Si values from least to the highest. The Si values from the Ca vs. Si intensity graphs were ultimately used as inputs in order to preserve the relationship where lowering the Ca/Si ratio of C-S-H gel increases the elastic modulus and hardness of the constituent phases [25, 28, 29]. It is important to note the extracted data sets from [109, 70, 72] only had three chemical intensities, I_{Ca}, I_{Si}, and I_{Al} available along with M, compared to the UHP mixtures where I_{Fe}, γ^3 were also available, along with the age of the mixtures, which influences the phase stiffness.

6.3.3 Details of the Input Data for ML Models, Based on Experimental Datasets

The normalized chemical intensities of different indent points are used to predict the nanomechanical data (effective modulus, M) using the ML models explained in Section 6.2. Models for all the mixtures shown in Table 6-1 were developed using these techniques. The details of the datasets are shown in Table 6-2. The OPC, NP, and SF mixtures have less complex microstructures – fewer number of ingredients, higher w/b ratio, longer curing durations, and/or exposure to hydrothermal conditions, all contributing to more homogeneous microstructure than those for the UHP pastes that have multiple starting materials and a very low w/b ratio. Individual predictions for the NP mixture, and predictions where the datasets of NP, OPC and SF mixtures are combined, are discussed. By combining the NP, OPC, and SF datasets, the effectiveness of ML algorithms to predict the performance of multi-mixture, multi-age data sets are explored. For the UHP pastes, the data for UHP-1 and 2 are combined for a certain age of hydration (30 days), and the data for UHP-1 at both 30 and 90 days of hydration are combined. Unlike common ML models that use mixture proportions as inputs to predict concrete strength, the inputs here are the relative intensities of the chemical species at the

140

indentation points. Since statistical nanoindentation accounts for the spatial distribution of components and their chemistry, ML models to predict nanoscale modulus from chemical information is deemed to be an appropriate approach.

Table 6-2: Details of data sets used as inputs to the ML models.

Data set (Table 6-1 for mixture details)	No. of data points	Input information	Source
NP	387	I_{Ca}, I_{Si}, I_{Al}	[70]
NP+OPC + SF[4]	1177	I_{Ca}, I_{Si}, I_{Al}	[70, 72, 109]
UHP-1 and 2 @ 30 days	2309	$I_{Ca}, I_{Si}, I_{Al}, I_{Fe}, \gamma^3$, Mixture type	[146, 154]
UHP-1 @30d and 90d	2416	$I_{Ca}, I_{Si}, I_{Al}, I_{Fe}, \gamma^3$, Age	[146, 154]

6.4 DATA PROCESSING

Several data preprocessing and parameter optimization methods are needed as part of implementing ML techniques for property prediction. They are discussed here.

6.4.1   Preprocessing and Evaluation

First, each data set shown in Table 6-2 was shuffled along the rows of indentation points (Figure 6-1(b)) such that adjacent points were separated, thus providing a greater chance of an equal distribution of the various microstructural entities within the testing and training data sets. The entire set of data was split such that 25% of the points were assigned to the testing set and 75% of the points were assigned to the training set. An artifact of the weight assignments in ML is that larger values will inherently be given a larger weight, which can skew the prediction significantly [87]. To address this, the input and output data points were pre-processed before separation into the testing and training

---

[4] The predictions for OPC and SF mixtures by themselves were also carried out. Though not as accurate as the NP mixture model (see Section 5.0), these mixtures were also able to bring out the influence of microstructural complexity on predictive efficiency, which is explained in the Discussions section.

sets. Equation 9, based on MinMaxScaler in scikit-learn [92], was used on the data to ensure that all of the inputs and outputs lie in the range [0, 1].

$$z_{new} = \frac{z - z_{min}}{z_{max} - z_{min}} \tag{43}$$

where z_{new} is the value of the variable after transformation, z is the current value of the variable, z_{min} is the minimum value of that variable, and z_{max} is the maximum value of that variable. After training and predicting, the normalized test data is converted back to the original scale using the "inversetransform" function.

Training was performed by fitting the ML algorithm to the training data set. This allows the algorithm to adjust its internal features to minimize the error. Model performance was evaluated using the testing data set, which the ML algorithm has not seen yet, and measuring the resulting errors. Each instance that the entire set of data (testing and training) is processed through the ML algorithm is called an epoch. Selection of different hyperparameters (see next section for more details) is accomplished by utilizing the same number of epochs to train MLs and comparing their accuracy. In training, the goal of the ML models is to minimize the objective function, which was the mean squared error (MSE), given as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (P_i - A_i)^2 \tag{44}$$

where n is the total number of data points, A_i is the actual value, and P_i is the predicted value. Other metrics tracked, but not used to train the models, were the mean absolute error (MAE) and the coefficient of determination (R^2), given as:

$$MAE = \frac{1}{n} \left(\sum_{i=1}^{n} \left| \frac{A_i - P_i}{A_i} \right| \right) \tag{45}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - A_i)^2}{\sum_{i=1}^{n}(A_i - \bar{A})^2} = \left(\frac{n(\sum_{i=1}^{n} A_i * P_i) - (\sum_{i=1}^{n} A_i) * (\sum_{i=1}^{n} P_i)}{\sqrt{n(\sum_{i=1}^{n} A_i^2) - (\sum_{i=1}^{n} A_i)^2} * \sqrt{n(\sum_{i=1}^{n} P_i^2) - (\sum_{i=1}^{n} P_i)^2}} \right)^2 \tag{46}$$

Where $\bar{A}$ is the mean of the actual values.

6.4.2 Hyperparameter Optimization

In SVR, the hyperparameters to be tuned were C and the length-scale for the squared exponential kernel. The ε error threshold was held constant at 0.90 for every data set to equally penalize training data that fell out of the range. The three parameters to optimize in the ANN models were the number of hidden layers, the number of neurons in each hidden layer, and the dropout rate. ReLu activation function with a learning rate of 0.001 and an RMSprop optimization scheme was used. For the RF, ET, and GBT models, the number of trees in the forest were tuned. Coarse optimization of the hyperparameters for SVR, ANN, and the forest ensembles followed a random search pattern, found to be the most efficient method to optimize parameters [164], by randomly generating 20 different combinations of hyperparameters. The hyperparameters for random testing were chosen from the uniform distributions shown in Table 6-3. GPR automatically searches through the allowed range of kernel parameter values and selects the parameters that maximize the likelihood estimation [91], and hence no iterative search over the hyperparameters was implemented. Instead, four types of kernels, i.e., the squared exponential, the rational quadratic, the Matern 1.5, and Matern 2.5, a range of the variance parameter σ_f^2 from 10^{-3} to 10^3, length-scales ℓ from 10^{-7} to 10^3, and an α bound of 10^{-5} to 10^2, were utilized to ensure the selection of optimized parameters.

Table 6-3: Hyperparameters tuned based on a uniform distribution range of potential values.

Model	Hyperparameter	Uniform Distribution Range
SVR	Regularization Parameter, C	[1, 500]
	RBF kernel length-scale, ℓ	[0, 1]
ANN	No. hidden layers	[1, 4]
	No. starting neurons	[10, 75]
	Drop rate	[0, 0.15]
Random Forest, Extra Trees Forest, and Gradient Boosted Forest	No. of trees	[50, 400]

To test the accuracy of the predictions under each set of test parameters, an n-fold cross-validation technique was employed [9, 87, 11]. A 3-fold cross-validation, deemed sufficient for the size of the data sets, was performed using the following steps: (i) randomizing the data set and splitting into 3 folds, (ii) training the model with selected parameters using 2 of the folds, (iii) testing the model using the remaining fold, (iv) repeating steps (ii) and (iii) until each fold has been used for testing once, acquiring 3 independent performance measures, and (v) averaging the individual accuracy measures to obtain the cross-validation errors. The parameters which minimized the cross-validation error was used as a basis for the final models with some additional fine-tuning searches. A concise summary of the parameter selection, training, and testing procedures are shown in Figure 6-2.
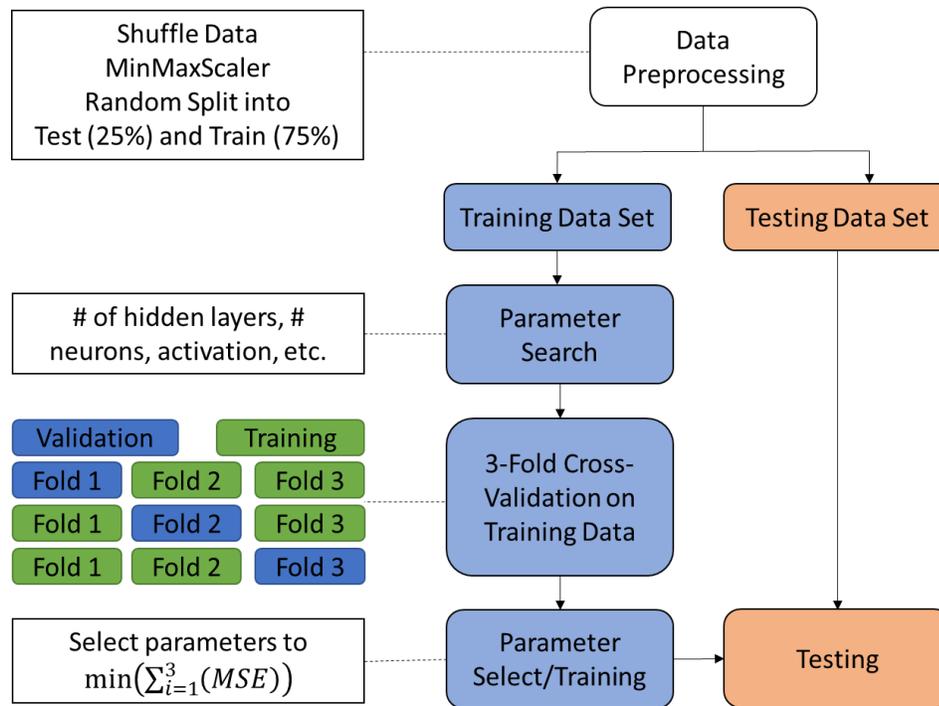
Figure 6-2: A summary of steps employed in ML model generation and selection. During the coarse hyperparameter search, the parameters are randomly chosen from a uniform distribution of potential values. During the fine hyperparameter search, the parameters are tested in a grid around the best coarse search model.

Final hyperparameter details for each data set are given in Table 6-4 for the SVR models, in Table 6-5 for the GPR models, in Table 6-6 for the ANN models, and Table 6-7 for the ensemble models. A reminder in ANN models that for each subsequent hidden layer the number of neurons is halved. For example, an ANN which began with 50 neurons in hidden layer 1 will have 13 neurons in hidden layer 3.

Table 6-4: Final hyperparameters of SVR models for each dataset to predict the nanoindentation stiffness given qualitative chemical intensity.

Dataset	# of inputs	Kernel	Run Time (sec)
NP	3	rbf, C= 275, Epsilon = 0.1, Gamma = 0.2036	0.01
NP + OPC + SF	3	rbf, C= 378, Epsilon = 0.1, Gamma =0.95	0.15
UHP-1 and 2, 30d	6	rbf, C=38, Epsilon=0.1, Gamma = 0.25	0.10
UHP-1 30d and 90d	6	rbf, C= 241, Epsilon = 0.1, Gamma = 0.15	0.25

Table 6-5: Final hyperparameters of GPR models for each dataset to predict the nanoindentation stiffness given qualitative chemical intensity.

Dataset	# of inputs	Kernel	Run Time (sec)
NP	3	31.6**2 * Matern(length_scale=0.155, nu=1.5) + 3.2**2 * Matern(length_scale=0.000125, nu=1.5) + 31.6**2 * Matern(length_scale=710, nu=1.5) + WhiteKernel(noise_level=0.00219)	16.96
NP + OPC + SF	3	21.7**2 * RationalQuadratic(alpha=0.377, length_scale=0.0615) + 31.6**2 * RationalQuadratic(alpha=1e-05, length_scale=1.01e-07) + 31.6**2 * RationalQuadratic(alpha=100, length_scale=720) + WhiteKernel(noise_level=30.9)	366.36
UHP-1 and 2, 30d	6	0.0355**2 * RBF(length_scale=0.031) + 14.8**2 * RBF(length_scale=2.11e-06) + 31.6**2 * RBF(length_scale=464) + 0.112**2 * RBF(length_scale=0.0109) + 11.7**2 * RBF(length_scale=24) + 31.6**2 * RBF(length_scale=83.1) + WhiteKernel(noise_level=53.9)	938.02
UHP-1 30d and 90d	6	18.2**2 * Matern(length_scale=0.00489, nu=2.5) + 31.6**2 * Matern(length_scale=300, nu=2.5) + 31.6**2 * Matern(length_scale=300, nu=2.5) + 0.128**2 * Matern(length_scale=0.704, nu=2.5) + 19**2 * Matern(length_scale=50.5, nu=2.5) + WhiteKernel(noise_level=15.8)	1789.14

Table 6-6: Final hyperparameters of ANN models for each dataset to predict the nanoindentation stiffness given qualitative chemical intensity.

Dataset	# of inputs	# of Hidden Layers	Starting # Neurons	Dropout	Activation	Optimization	Learning Rate	Epochs	Run Time (sec)
NP	3	2	30	0.067	ReLu	RMSprop	0.001	800	13.52
NP + OPC + SF	3	3	53	0.09	ReLu	RMSprop	0.001	800	32.08
UHP-1 and 2, 30d	6	1	26	0.135	ReLu	RMSprop	0.001	300	17.87
UHP-1 30d and 90d	6	1	40	0.055	ReLu	RMSprop	0.001	500	27.74

Table 6-7: Final hyperparameters of ensemble models for each dataset to predict the nanoindentation stiffness given qualitative chemical intensity.

Dataset	# of inputs	Ensemble	n_estimators (# of Trees)
NP	3	RandomForestClassifier()	141
		ExtraTreesClassifier()	65
		GradientBoostingClassifier()	68
NP + OPC + SF	3	RandomForestClassifier()	223
		ExtraTreesClassifier()	315
		GradientBoostingClassifier()	76
UHP-1 and 2, 30d	6	RandomForestClassifier()	222
		ExtraTreesClassifier()	263
		GradientBoostingClassifier()	56
UHP-1 30d and 90d	6	RandomForestClassifier()	345
		ExtraTreesClassifier()	346
		GradientBoostingClassifier()	53

6.5 RESULTS AND DISCUSSIONS

The afore-described ML models were implemented on the four datasets culled from the five different mixtures shown in Table 6-1. The predictive efficiencies of the models are discussed in this section, followed by the physical reasoning behind the variable performances of the models among different mixtures.

Previous works [51, 69, 146, 154, 106] have revealed the complex relationships between the nanoindentation stiffness of the phases and the associated chemical intensities. As illustrated in Figure 6-3(a) and (b) for the NP+OPC+SF and UHP-1 pastes, there is a non-linear, complex relationship between M and the intensities of Ca and Si, which are the primary constituents of the product phases. It has been shown through experiments and atomic simulations that a reduction in Ca/Si ratio enhances the stiffness of the solid C-S-H phases [25, 28, 29]; however this is not the lone influential factor for many hydration product phases, for example: (i) the ultra-high stiffness (UHS) phase in UHP pastes with embedded nanocrystals of CH [69], and (ii) C-A-S-H gel in systems with significant cement replacement levels where alumina incorporation markedly changes the response [154]. The following sections describe how ML models perform in predicting the hydration product phase stiffnesses.

(a)   (b)
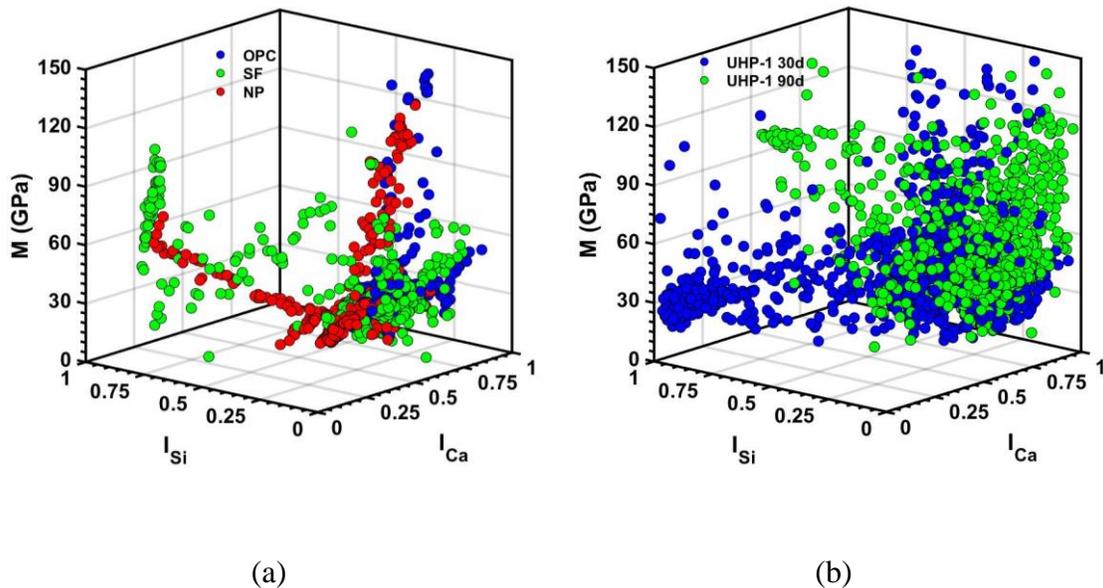
Figure 6-3: Relationship between the nanoindentation stiffness and the normalized intensities of Ca and Si for: (a) NP+OPC+SF combined dataset, and (b) UHP-1 30d and 90d pastes. Note that the chemical intensities are normalized, and thus vary from 0-1.

148

6.5.1 Predictive Efficiency of ML Models using Microstructural Data as Inputs

In this section, the predictive efficiency of the different ML models as applied to the different data sets are explained. The graphs shown in this section are representative, and do not include all the models explored. Table 6-8 lists all the models implemented for all the data sets and summarizes the prediction metrics. Note that for the first two rows in Table 6-2, comprising of mixtures that are less complex, only three input variables are available, while the UHP data sets are augmented with three more input variables.

6.5.2 SVR

Unlike the other ML models, SVR does not have an element of randomness in the hyperparameter training and testing. Thus, the same data set with a similar distribution of testing and training data sets will result in little to no change in consecutive runs. Additionally, as reported elsewhere [14], SVR has one of the fastest computing times, averaging at around 0.15 seconds per run on the largest two data sets (see SI). Figure 6-4(a-c) show the predictive efficiencies of SVR model for the NP, NP+OPC+SF, and UHP-1 and 2 30-day datasets, respectively. As can be noticed here, the prediction of modulus values for the NP data set are more accurate as compared to the NP+OPC+SF or the UHP data sets. As a general observation, when compared to the other models described later, SVR is found to be less accurate, regardless of the data set used. A few studies have shown the use of SVR to predict the compressive strength of concrete mixtures from mixture proportions [10, 11, 172]. However due to the natural spread in the data and the ε being fixed for all the data sets, the SVR model had moderate success in predicting modulus given chemical intensity and γ^3, especially when multiple mixtures are combined or more complex mixtures such as the UHP ones are used. Figure 6-4 (b)

149

and (c) show that a large amount of high modulus data is underpredicted by this ML model (and will be shown subsequently, that this is true for other models too), the reasons for which are explained in detail in the section discussing ANN models.

(a)   (b)   (c)
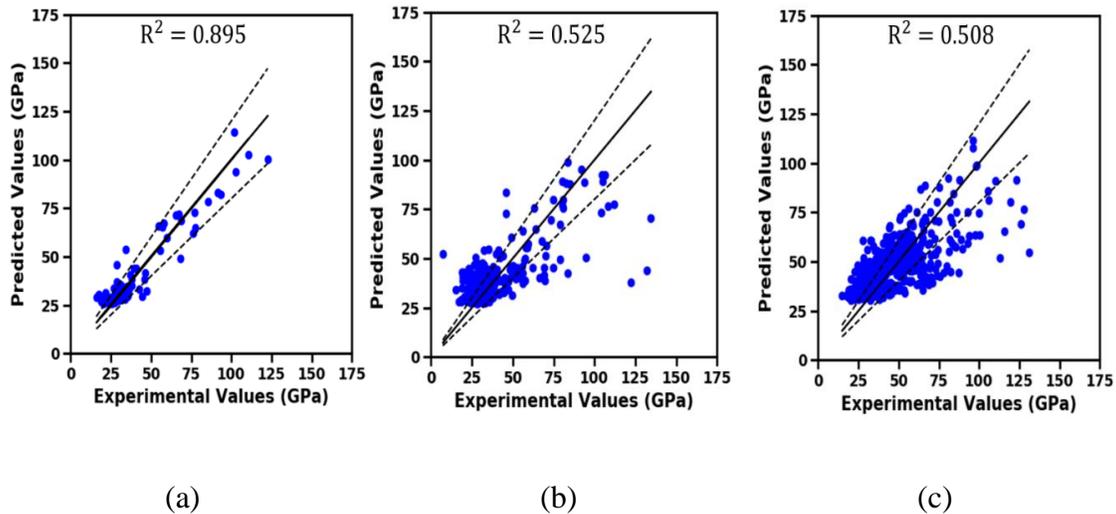
Figure 6-4: SVR estimation of M from testing data sets corresponding to: (a) NP, (b) NP+OPC+SF combined dataset, and (c) UHP-1 and 2 30-days. The solid line represents the line of ideality, and the dashed lines represent a $\pm$ 20% bound.

6.5.3 GPR

GPR is a non-parametric, probability-based ML method allowing for the representation of error bars as the standard deviation of the estimates at each point, as shown in Figure 6-5. GPR was most successfully able to predict the modulus of the phases from the smallest data set (NP), with a RMSE of 5.45 GPa and an R^2 of 0.947 as shown in Figure 6-5(a). The ability of GPR to work on sparse data sets is well documented [167], attributed to its non-parametric nature. With few inputs, there are few parameters that vary depending on the training data set and GPR is much less likely to experience overfitting compared to a method such as ANN. Conversely, as the data size and complexity in the mixtures increased, the GPR estimations became the least accurate of

150

all techniques (see Table 6-8), as evidenced from Figure 6-5(b) for the UHP paste. For the NP+OPC+SF data set, GPR achieved the second-highest R^2, but with an elevated MSE and MAE compared to the best model. This large MSE and MAE error indicates that the residual difference between the experimental and predicted results had nearly tripled with the inclusion of two additional data sets. As an additive GPR kernel method was performed, the addition of every input indicated an increased tendency to overfit the data [173]. This overfitting is evident by comparing the R^2 for the testing data (0.395) with that of the training data (0.996) for the UHP data set. The GPR model clearly is appropriate for smaller, more uniform data sets, and is less capable of handling large diversity in the data set.

(a)   (b)
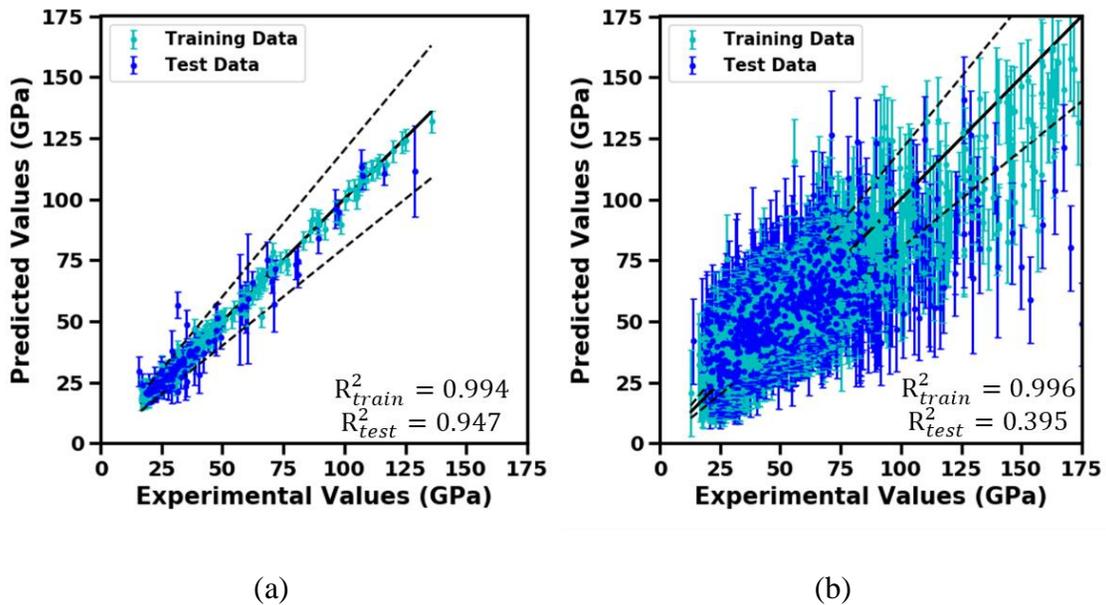
Figure 6-5: GPR estimation of M from testing and training data sets corresponding to: (a) NP, and (b) UHP-1 and 2 30-days. The solid line represents the line of ideality, and the dashed lines represent a $\pm$ 20% bound. Both training and testing data sets are shown for the GPR case alone.

6.5.4 ANN

The ANN predictions of phase moduli based on the input parameters are shown in Figure 6-6. Models for all four data sets from Table 6-2 are shown here. Once again, the best performance is noted for the NP mixture as seen from Figure 6-6(a). With increasing complexity, either through the incorporation of multiple mixtures (Figure 6-6(b)) or multiple starting materials (Figure 6-6(c) and (d)), the predictive ability is significantly lower. In Figure 6-6(c) and (d), the different microstructural phases are shown separately as obtained from the clustering analysis for the UHP pastes [146, 154]. In Figure 6-6(c), which shows the data for UHP-1 and 2 pastes at 30 days, one combined C-S-H reaction product cluster is shown corresponding to HD C-S-H for UHP-1 and the ultra-high stiffness (UHS) phase for UHP-2 [146, 154], which explains the two groupings for that cluster. While the indentation moduli of the major reaction product phases (C-S-H/UHS) are predicted quite accurately (within the 20% error band) using the chemical intensity information as shown in these figures, the overall prediction quality is lowered by the less-than-sufficient predictive quality for the mixed reaction product phases and unreacted phases. More discussions on the mixed product phase is provided in Section 6.5.2.

Further examination of Figure 6-6(b-d) also reveals that the predictions are underestimated for a large proportion of data with higher M (for e.g., in these figures, the scatter in the indentations with M > 65 GPa can be easily recognized), which belongs to the clinker/unreacted phases. One possible explanation for the ML models' tendency to under-estimate the stiffness of phases such as clinker or the unreacted cement replacement materials stems from the contrasting influence of Ca. As stated earlier, a

152

reduction in Ca/Si ratio generally enhances the stiffness of the solid C-S-H phases [25, 28, 29]; however for the UHS phase [146, 154] with nanocrystals of CH, an increase in Ca increases the M [69]. Also, the high stiffness clinker has a higher concentration of Ca. While the use of γ^3 as an input did contribute to some improvement in predictive capability, it still leaves a lot to be desired. There also exists an inherent imbalance in the distribution of the nanoindentation data in complex binders, where most of the indentation points in hydrated systems belong to the matrix/C-S-H phases, while only a small fraction (~20% in the UHP pastes) belongs to the unreacted/partly reacted cement replacement materials, and remaining (~ 10% or less) to the clinker [146]. Abundantly available data for the matrix/C-S-H phases allows the ANN to adjust the weights according to the most prominent phases in the UHP pastes (i.e., ANN models, like many ML methods, are skew-sensitive, meaning they are biased towards the trends favored by the majority of the data [163]), such as the HD C-S-H phase with a known stiffness around 35 GPa [62]. To predict with accuracy the chemical intensity-mechanical property relationship in complex microstructures requires MLs such as ANN to adapt to a large data set where only a few datapoints represent phases such as clinker.
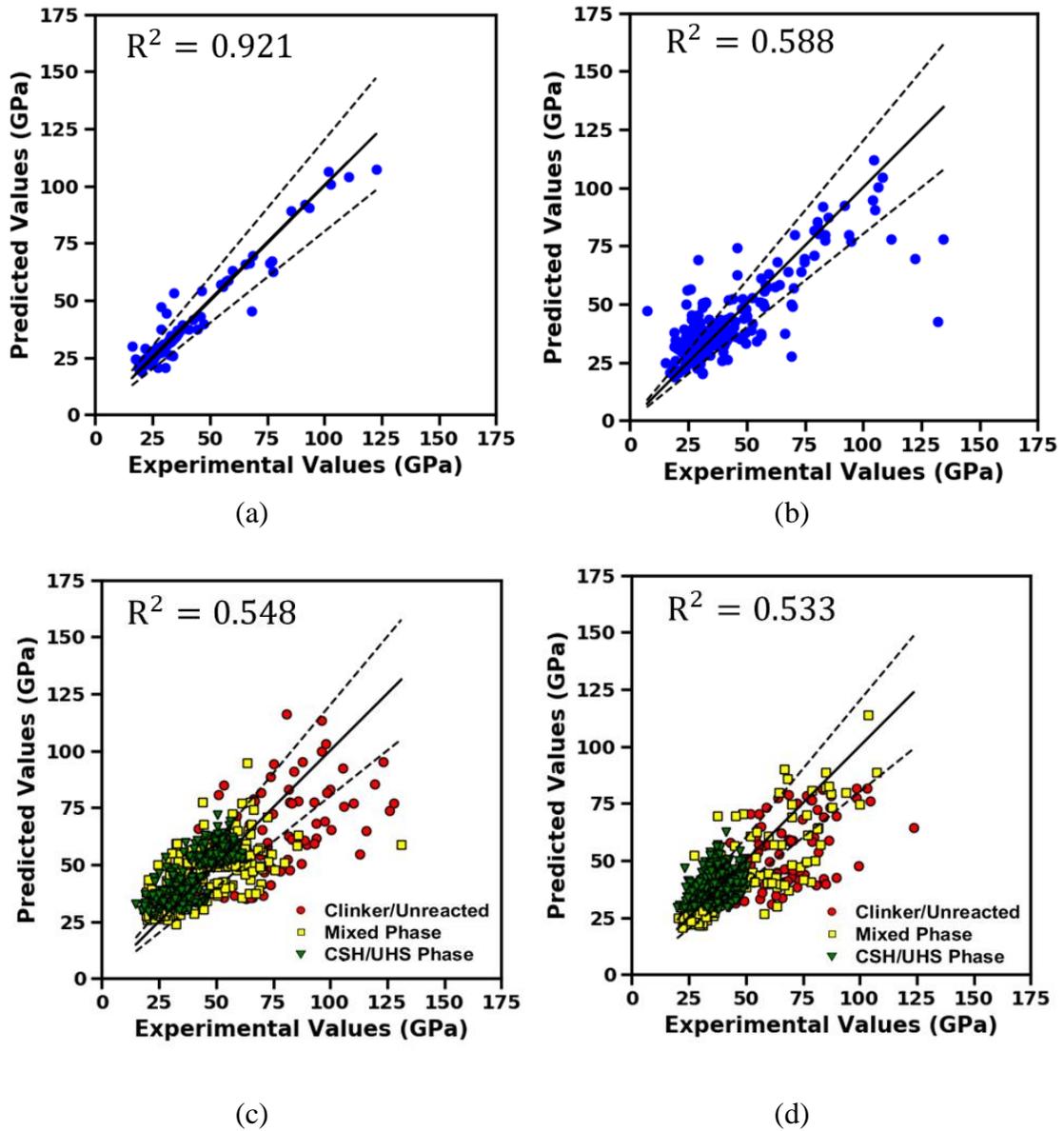
Figure 6-6: ANN estimation of M from testing data sets corresponding to: (a) NP, (b) NP+OPC+SF combined dataset, (c) UHP-1 and 2 30-days and (d) UHP-1 30 and 90-days. The solid line represents the line of ideality, and the dashed lines represent a $\pm 20\%$ bound. The different microstructural phases, according to the clustering analysis are shown in (c) and (d). For ease of explanation, they are grouped into just three distinct phases.

6.5.5 Ensembles

For all the datasets considered, the forest ensembles performed similar or better than the other ML models. Among the three forest ensembles tested, ET and RF models performed better than the GBT model. Beyond changing the parameters such as the interaction depth (maximum number of nodes per tree) or the minimum number of observations per node, it is unclear how the GBT model could be further improved to enable better predictions [92]. Here, ET regression estimates for the NP+OPC+SF and UHP-1 data sets are shown in Figure 6-7(a) and (b). The metrics for all mixtures and all ensembles can be found in Table 6-8. The ET model was able to achieve a significant improvement in all accuracy measures compared to other models for the NP+OPC+SF and UHP-1 data sets, with the next closest in accuracy being the RF model. A previous work on concrete mixture optimization found that forest models outperformed ANN in the case of unbalanced and discrete data [14]. This improvement was credited to the improved generalization in the forest methods, which reduces the instability of individual trees through random sampling of data and random selection of input features [10, 163]. Like ANN, RF and ET models are also skew-sensitive.
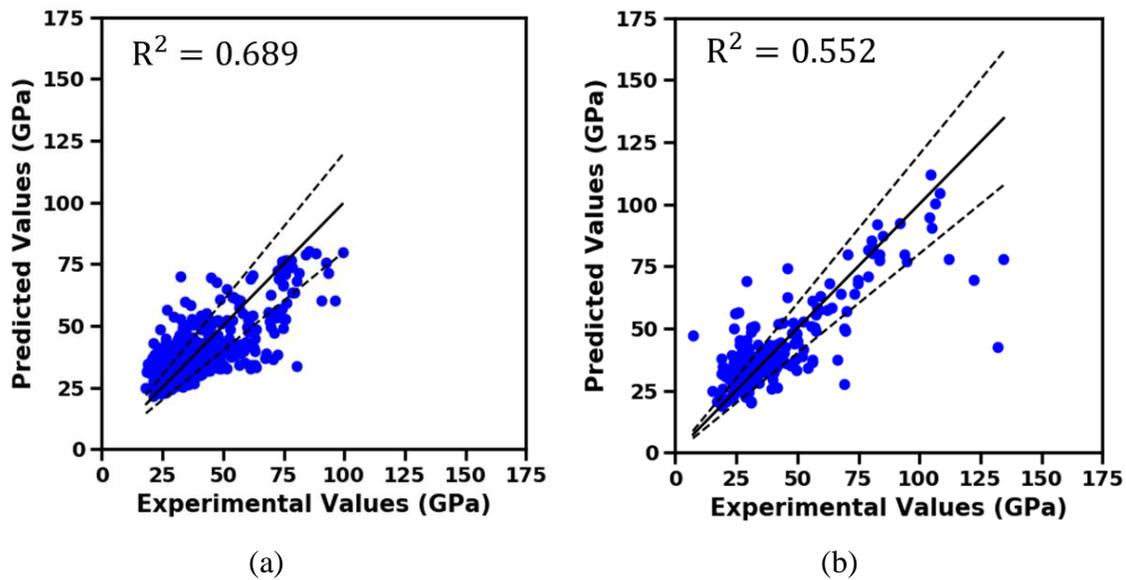
Figure 6-7: ET Forest estimation of M from testing data sets corresponding to: (a) NP+OPC+SF combined dataset, and (b) UHP-1 30 and 90-days.

6.5.6 Comparison of the ML Models

Table 6-8 shows the metrics used to evaluate the performance of the ML models for all

the data sets shown in Table 6-2. Figure 6-8 shows the RMSE values for all the ML

models and data sets. All the ML methods fitted the NP data set very well, with the

lowest R^2 being 0.895 and the highest RMSE being 7.25 GPa, using SVR. The chemical-

nanomechanical relationship of the NP paste was easily handled by all algorithms and

serves as a "best case" application of ML to predict nanoindentation moduli. When

handling multiple-blend mixtures, or those with lower degrees of hydration and at

different ages, the accuracy of ML predictions using the other three data sets were

significantly lower in comparison with the NP mixture, as can be seen in Figure 6-8. For

these data sets, the low predictive ability was consistent among the SVR, GPR, ANN, and

ensemble models, the reasons for which are further explored below. In addition to these

methods, voter methods [166] that use a combination of two or more ML models were

found to provide some improvement in predictive capabilities depending on the ML
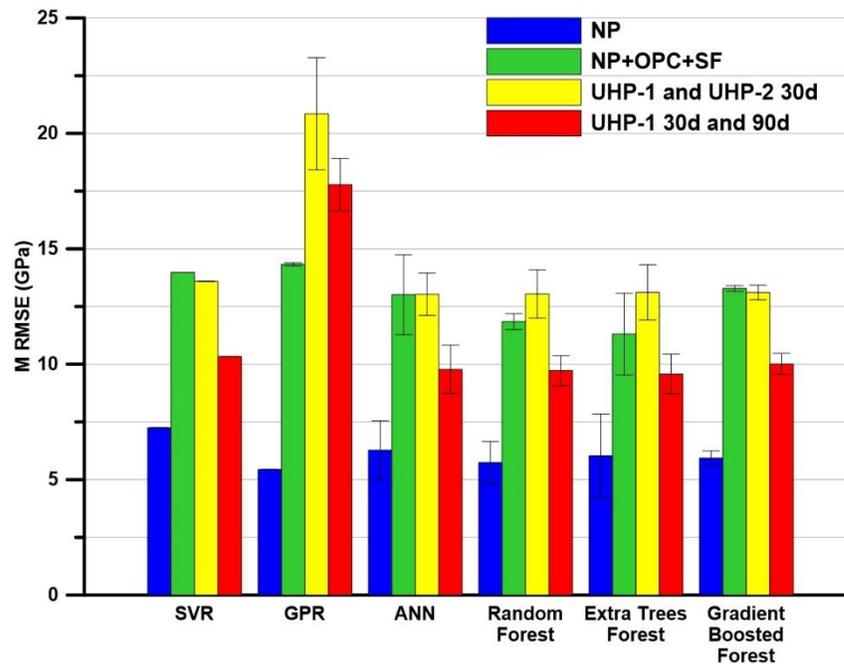
model combinations used; however, they are not discussed further in this paper.

Figure 6-8: Comparison of RMSE of M predictions from different ML techniques for all data sets. Error bars shown for standard deviation are from three consecutive runs of the models.

Table 6-8: ML results for all the data sets, with the average and standard deviation values over three consecutive runs. The most accurate ML model for each data set is shown in **bold**.

Data Set	Model Type	RMSE M (GPa)	MAE M (GPa)	R^2
NP	SVR	7.25 ± 0.00	5.53 ± 0.00	0.895 ± 0.000
	GPR	**5.45 ± 0.01**	**3.29 ± 0.00**	**0.947 ± 0.000**
	ANN	6.28 ± 1.27	4.08 ± 0.12	0.921 ± 0.003
	Random Forest	5.74 ± 0.92	3.26 ± 0.07	0.934 ± 0.002
	Extra Trees Forest	6.03 ± 1.82	3.34 ± 0.10	0.927 ± 0.007
	Gradient Boosted Forest	5.93 ± 0.31	3.68 ± 0.01	0.930 ± 0.000
NP + OPC + SF	SVR	13.97 ± 0.00	9.26 ± 0.00	0.525 ± 0.000
	GPR	14.33 ± 0.07	9.10 ± 0.01	0.663 ± 0.000
	ANN	13.01 ± 1.73	8.16 ± 0.22	0.588 ± 0.007
	Random Forest	11.85 ± 0.34	6.85 ± 0.04	0.659 ± 0.000
	Extra Trees Forest	**11.30 ± 1.77**	**6.42 ± 0.06**	**0.689 ± 0.008**
	Gradient Boosted Forest	13.28 ± 0.13	8.30 ± 0.00	0.571 ± 0.000
UHP-1 and 2, 30d	SVR	13.59 ± 0.00	0.00 ± 0.51	0.508 ± 0.000
	GPR	20.85 ± 2.43	14.62 ± 0.09	0.395 ± 0.008
	ANN	**13.03 ± 0.92**	**9.20 ± 0.30**	**0.548 ± 0.002**
	Random Forest	13.04 ± 1.05	9.08 ± 0.02	0.547 ± 0.003
	Extra Trees Forest	13.11 ± 1.20	9.12 ± 0.03	0.542 ± 0.004
	Gradient Boosted Forest	13.11 ± 0.32	8.93 ± 0.01	0.542 ± 0.000
UHP-1 30d and 90d	SVR	10.34 ± 0.00	8.38 ± 0.00	0.478 ± 0.000
	GPR	17.78 ± 1.13	12.55 ± 0.03	0.438 ± 0.002
	ANN	9.78 ± 1.05	6.97 ± 0.12	0.533 ± 0.005
	Random Forest	9.72 ± 0.65	6.88 ± 0.01	0.538 ± 0.002
	Extra Trees Forest	**9.58 ± 0.86**	**6.79 ± 0.03**	**0.552 ± 0.004**
	Gradient Boosted Forest	10.01 ± 0.46	7.11 ± 0.01	0.510 ± 0.001

6.5.7 Implications on the Use of ML for Predictions Related to Complex Microstructures

For the NP mixture reported in [70], the predictive capacity of all the ML models are found to be excellent. Because of the high reactivity of the natural pozzolan, a higher w/b used, and the longer curing duration, the microstructure is more uniform, and there are fewer unhydrated clinker remnants. The reaction product in this microstructure is also more homogeneous because of the improved reaction of natural pozzolan with calcium hydroxide (CH) forming more C-A-S-H gel. Thus, with three relative chemical intensities as inputs, all of which appear in the chemical description of C-A-S-H, ML models are

found to accurately predict the micromechanical response. Moreover, the dataset is from one mixture at one age, giving it more uniformity, and therefore certainty in predictions. This indeed is an ideal and desired scenario. Similar is the case when the plain OPC mixture was considered. However, when the data sets corresponding to NP, OPC, and SF are combined, the predictive capability is reduced. This could be because of differences in "relative intensities" of the chemical species between the different data sets acquired from different experimental set up, ages, and mixtures, as well as the relative contributions of phases such as unreacted clinker in the data sets (see Figure 6-6), the nanomechanical properties of which may not follow the same relationships with chemical intensities as the hydration phases.

As explained in the previous sections, the UHP pastes are not well-modeled by any of the ML methods. The presence of multiple starting materials of varying reactivity and nanomechanical properties and the low w/b ratio (to achieve high strength and durability) result in highly heterogeneous microstructures as has been reported elsewhere [146, 154]. The relative paucity of data from the high stiffness phases as compared to those from reaction products result in its underprediction for the case of UHP pastes, as shown in Figure 6-6 (c) and (d). Moreover, in the < 65 GPa region in Figure 6-6 (c) and (d), the predictive trends are better for the C-S-H/UHS phases, as explained earlier. The mixed phase identified in these microstructures, that shows decent amount of scatter, is likely a composite of different chemical species, e.g., C-S-H phases of varying stiffness, unreacted/partly reacted limestone powder and fly ash, microsilica, and carboaluminates [169]. This is because the region of influence under the indent is larger than the characteristic sizes of some of the constituents in the system. It is also possible that some

159

phase interfaces are also captured in the mixed phase. This has been shown by the authors through a closer look at the M-H cluster of the UHP-1 paste in Figure 6-9 [146] which also suggests that for M < 65 GPa, about 20%-25% of the reaction products show H > 3 GPa, indicating that a multi-phase response is indeed acquired. Also note the significant scatter in the clinker/unreacted phases. The chemistry-nanomechanical property relationships for the large number of mixed/unreacted phases present in heterogeneous microstructures (see Figure 6-9) could also be different from those for the majority hydrate phases, resulting in less than ideal predictive ability.

These point to the need for more robust input variables for more complex microstructures; though the inclusion of a pseudo local density term through γ^3 only slightly improved the predictions. The addition of physically informed variables such as Ca/Si ratios did not alter the predictive capabilities. In the absence of any conceivable means to identify and label the individual chemical species in a mixed phase, this will likely remain an impediment to accurate prediction of phase stiffness from chemical maps for highly heterogeneous systems. Accurately identifying the multiple constituents currently represented as clinker in Figure 6-9 and thus reducing the scatter in the > 65 GPa region (e.g., in Figure 6-6(c) and (d)) could help improve predictive capability. Another potential improvement could be through data augmentation, i.e., the use of more data points especially from under-represented phases such as clinker. The inclusion of more inputs that would offer further insights into the relationship between chemistry and micromechanical properties, such as fineness of reactants, or the location of the indentation points relative to one another, is also a possible option. The location of hydration phases within a cementitious matrix is dependent on the proximity to clinker or

other starting materials [112]. For example, high density (HD) C-S-H with an effective

nanoindentation modulus of 35 GPa [62] is found next to clinker while low density (LD)

C-S-H of 23 GPa stiffness [53] is typically located farther from the clinker in the matrix.

As such, an image-based ML method such as convolutional neural networks (CNNs) may

be able to provide the needed additional inputs to expand the ability of ML in predicting

the micromechanical properties of microstructural phases from the local chemical

composition. Such approaches are being evaluated. Meanwhile, for plain OPC pastes,

pastes containing smaller amounts of reactive cement replacement materials, and well-

hydrated systems, chemical intensity maps coupled with ML provide an easy and reliable

methodology to predict the phase stiffnesses.
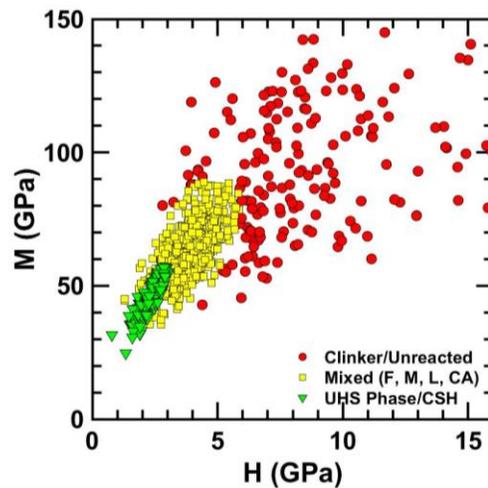
Figure 6-9: M-H response for the UHP-1 paste cured for 90 days [146]. F, M, L, and CA denotes fly ash, microsilica, limestone, and carboaluminates, respectively. Note the grouping of the mixed phases with similar mean stiffness and hardness, but widely differing chemical compositions.

6.6 SUMMARY

This chapter has presented a novel approach to predict the nanoindentation stiffness of

cement hydration phases from chemical intensity maps using ML methods. To

demonstrate the methodology and application, nanoindentation modulus and normalized

chemical intensity data from EDS for four different data sets representing various cementitious mixtures with multiple cement replacement materials, w/b, hydration ages, and microstructural complexities were chosen. Four different classes of ML methods including Support Vector Regression (SVR), Gaussian Process Regression (GPR), Artificial Neural Networks with backpropagation (ANN), and ensemble methods (e.g., RF, ET, and GBT) were implemented on data sets comprising of ~400 to 2400 data records, each corresponding to a distinct nanoindentation point in the microstructure. The models were trained using 75% of the data and tested on the remaining 25% of the data. EDS maps were obtained for the microstructures so that the local chemistry could be used to predict the hydration phase modulus using ML models. The relative intensity of the major chemical species (Ca, Si, Al, Fe), the mixture ages, as well as the local density of the microstructure represented by the cube of the brightness coefficient from backscattered SEM were used as the inputs.

For the NP mixture, all ML models were able to predict the phase stiffness with an R^2 value ~ 0.90 using just three chemical intensities (Ca, Si, Al) as inputs, all of which appear in the chemical description of C-A-S-H gel. GPR was shown to be the best model for this mixture in terms of the metrics (RMSE, MAE, R^2), attributed to its ability to work on sparse data sets with fewer inputs. When data sets belonging to NP, OPC, and SF mixtures were combined, the relative predictive efficiency dropped irrespective of the ML model used. For the UHP pastes demonstrating complex microstructures from multiple binding materials of varying chemical constitution and reactivities, and incomplete reaction due to low w/b, all the ML models were found to have lower predictive abilities.

CHAPTER 7

IMAGE-BASED MACHINE LEARNING TO PREDICT UNIDIRECTIONAL FIBER

MECHANICAL PROPERTIES

7.1 INTRODUCTION

Composite materials are ubiquitous in nature, and artificial composites are employed for

a large number of engineering applications. A matrix phase and one or multiple fiber or

inclusion phases are judiciously combined to obtain desirable structural (e.g., strength,

stiffness, toughness) and/or functional (e.g., electrical or thermal conductivity) properties.

Most generic fiber reinforced composites are two-phase composites, consisting of a

strong and stiff fiber phase distributed uniformly, or aligned in a particular direction, in a

compliant matrix. The fact that fiber size is uniform and cross sections are generally

circular, enables the microstructure of such composites to be generated easily, for

analysis. In more complex multi-phase composites (e.g., concrete, rocks), different

inclusion phases of random shapes and sizes are randomly distributed, rendering the

analysis, and consequently the prediction of properties, more complex.

The mechanical properties of composite materials are highly dependent on the type and

properties of the constituents, and their spatial arrangement in the microstructure [174]

[39] [175]. In addition to experimental methods to determine the relevant composite

properties, analytical homogenization methods are in use to obtain macroscopically

averaged properties. These micromechanical homogenization methods include Mori-

Tanaka mean field method, double inclusion method, and self-consistent approaches

[129] [131][63]. These approaches are generally effective for simple microstructures, but

when microstructural complexity increases, or when constitutive relations are demanded,

numerical methods are resorted to [176] [177]. Microstructure-based finite element method (FEM) is normally employed. High fidelity FE simulations are known to provide accurate predictions of a range of composite properties, both in the linear elastic and non-linear regimes. They also allow for accurate relationships between the microstructural parameters (inclusion size, volume fraction etc.) and the measured response. The major drawback with FEM-based predictions is that the representative volume element (RVE) size needs to be large enough, especially for complex microstructures containing a diversity of inclusion sizes and shapes [178], for extraction of non-linear response [179], or for crack propagation simulations [180], which renders the entire process tedious and time-consuming.

The microstructural dependence of properties of composite materials enables the use of properly trained machine learning (ML) methods to rapidly evaluate the properties of composite materials, given the material microstructure or its adequate representation [181] [95] [85]. ML methods are finding tremendous applications in wide and disparate fields such as image classification and identification, language and text processing and analysis, and material design and discovery [8] [182]. These methods are finding increased acceptance in materials science and engineering. Using large data sets of constituent materials and their properties, ML methods based on neural networks or forest ensembles are used to predict material properties. Using image-based ML methods (including artificial neural networks or ANN, conventional neural networks or CNN, or long-term recurrent convolutional networks or LRCN) coupled with FEM, several recent studies investigate the response of different types of composites [86] [183]. In this study, we use virtual microstructures of three common two-phase composite materials: (i) an

164

aerospace unidirectional fiber reinforced composite - T800 carbon fiber in F3900 matrix, with uniform size fibers and different fiber volume fractions, (ii) an ultra-high performance (UHP) mortar composed of different median size sand particles in a strong cementitious matrix, and (iii) a metallic particulate reinforced cement mortar. Several virtual microstructures belonging to each of these composites are created using a microstructure generation algorithm, meshed, and subjected to linear elastic FE analysis to determine the elastic modulus, Poisson's ratio, and the stresses in both the phases, under the application of a given strain. Limited experimental data is used to validate the FE analyses. Two-point correlations (TPC) corresponding to the virtual microstructures are extracted, and the dimensionality reduced using principal component analysis (PCA). The principal components thus extracted are used as inputs to the ML algorithms to predict the composite properties from microstructural images. This high throughput, negligible cost, highly generalizable computing method can be used to rapidly determine the properties of two-phase composites from microstructure, enabling efficient design of such materials.

7.2 METHODOLOGY

7.2.1 Generation of Stochastic Microstructures

A synthetic microstructure generation algorithm developed in MATLAB capable of generating non-overlapping ellipses of varying sizes, aspect ratios, and orientations was used to construct the two-phase microstructures [184] [185]. For the unidirectional fiber reinforced composite microstructure (hereinafter labeled as UD-FRC), circles with a set radius were consecutively added to the window space until the specified area fraction was obtained. For the UHP mortar microstructures, circles of differing dimensions were

165

generated based on a particle size distribution, while for the metallic particulate

reinforced cement paste (hereinafter termed as MPR paste), elliptical particles of different

dimensions based on a particle size distribution were incorporated in the paste matrix.

The program created the square representative element area (REA) and inclusion radii in

units of pixels, so that the simulations are generalized and not constrained by physical

dimensions. To eliminate the possibility of overlapping, particles were assigned a major

and minor radius 1 unit higher than the real values, which ensured that a gap of at least 1

unit existed between the inclusion edges. Periodic boundary conditions were maintained

such that if any part of the inclusion were generated at the edge of the REA, it would

wrap around to the other side, including particles located at the corners of the REA.

Figure 6-7 demonstrates representative images corresponding to a UD-FRC, and mortars

with sand or iron particle as inclusions, and Table 7-1 details the dataset volume fractions,

inclusion shapes, and size distributions. In addition to the images, the microstructural

generation algorithm was also used to output a .txt file to reproduce the microstructures

into Coreform Trelis™ for meshing prior to the finite element analysis (FEA).

Given that the microstructures were studied in the linear-elastic plane strain regime with

a small applied strain, changing the inclusion volume fraction would significantly change

the FEA mechanical results, while altering the particle size or shape would not. As in any

ML training regime, it was important that the output mechanical properties be sufficiently

different that the ML would learn meaningful connections between the inputs and outputs

- instead of trying to predict noise in the FEA simulated results. Therefore, the three size

distributions of sand mortars at the same volume fraction (ϕ) had to be combined with

the MPR paste inclusions to form a dataset that would successfully train the ML

algorithms.

(a)   (b)   (c)
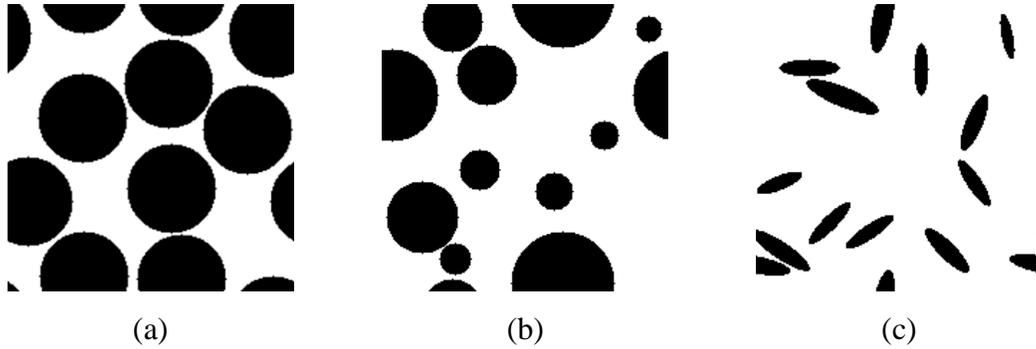
Figure 7-1: Examples of randomly generated 2D microstructures with a square window size of 200 pixels: (a) UD-FRC with a volume fraction of 0.60, (b) UHP mortar distribution 2, and (c) MPR paste.

Table 7-1. Details of the two-phase microstructural images generated and analyzed. a_0 is the major ellipse axis, b_0 is the minor ellipse axis, μ is the mean value of the distribution, and σ is the standard deviation of the distribution.

Dataset	Two-Phase Microstructure	Inclusion volume fraction (ϕ)	Ellipse size (pixels)	Particle shape distribution
UD-FRC	UD-FRC	0.35, 0.40, 0.45, 0.50, 0.55, and 0.60	$a_0 = 31$ $b_0 = 31$	Constant
UHP mortar and MPR paste	UHP mortar distribution 1	0.35	$a_0 = 10$ $b_0 = 10$	Lognormal: $\mu = 10, \sigma = 0.7$
	UHP mortar distribution 2	0.35	$a_0 = 20$ $b_0 = 20$	Lognormal: $\mu = 10, \sigma = 0.7$
	UHP mortar distribution 3	0.35	$a_0 = 10$ $b_0 = 10$	Constant
	MPR paste	0.12	$a_0 = 20$ $b_0 = 5$	Gaussian Normal for a_0: $\mu = 20, \sigma = 0.25$ $\frac{a_0}{b_0} = 4$ Constant

As all the microstructures are considered truly random, for each unique microstructure

generated an additional one was created by rotating the parent microstructure

counterclockwise by 90°. This allowed for efficient creation of a multitude of

microstructures that can be expected to have unique mechanical properties under the chosen loading conditions because of the differences in organization of the phases. As machine learning algorithms require large data sets for successful training and testing [181], such an approach can be utilized to create large datasets. However, it also needs to be noted that, for such simple microstructures, generating more of them from scratch is not very time consuming, but the abovementioned approach is adopted here to serve as a trial case for more complex microstructures (e.g., concrete) that will be investigated in the future. Here, 125 samples belonging to each matrix-inclusion combination were developed for every volume fraction or particle size distribution investigated. This sample size proved to be sufficient to train all the ML models, as will be demonstrated later.

7.2.2   Extracting the Relevant Microstructural Descriptors

In this work, the analysis and predictions are based on 2D microstructures, which have been shown to be sufficient to describe the linear elastic properties of two-phase (matrix-inclusion) composites considered here [178] [186]. In general, for a 2D microstructure, the volume fractions of the phases and sizes of the inclusions are considered to be the most important microstructural descriptors. However, for random composites, finer aspects of the geometry, including non-trivial information on the arrangement of the phases and the relative positions of the inclusions, are important in accurate property prediction [187] [188]. A large family of statistical descriptors are available for such characterization [189], among which n-point correlation functions [190] [191] [192] [193] are deemed to be most suitable for these types of random heterogeneous materials.

7.2.3 Two-point Correlation (TPC) Function

n-point correlation functions have been implemented in microstructure-based analysis of heterogeneous materials, including higher-order reconstructions and property prediction [181] [174] [190] [188]. The one-point correlation function indicates the volume fraction of a certain phase in the microstructure [190]. The two-point correlation function (TPC), $S_2(r)$, quantifies the probability of finding phases α and β at the start and end of a vector of length r, randomly placed in the microstructure. This quantity is invariant with respect to translations of the periodic microstructure [188]. For periodic microstructures where the characteristic particle density function is positive and bounded everywhere, the TPC is sufficient to generate a unique solution up to translation and inversion [194] [192]. In this study, periodic boundary conditions and a constant particle density are maintained such that the TPC is sufficient to quantify and differentiate between the different microstructures randomly generated, for the same specimen. The $S_2(r)$ of an image of pixel size $\sqrt{S} * \sqrt{S}$ is given as [181] [195]:

$$f(\alpha, \beta | r) = S_2(r) = \frac{1}{S} \sum_{s=1}^{S} m_s^\alpha * m_{s+r}^\beta \tag{47}$$

Here, S is the total number of pixels in the microstructure, s is the spatial position of the pixel, and r is a vector of a given length and direction. TPC is known as an "autocorrelation" when α and β are the same phase, and cross-correlation when α and β are different phases [181]. For a two-phase composite where the two phase labels are 1 or 0, the four TPCs, i.e., $f(0,0), f(1,0), f(0,1)$, and $f(1,1)$, are related, so that only one needs to be calculated to find the rest [196] [181]. Figure 7-2 shows an example microstructure of a UD-FRC specimen with a fiber volume fraction of 0.35, and the resulting contours of the $f(0,0)$ and $f(1,0)$ TPC functions. Note that the center of the

169

autocorrelation image in Figure 7-2(b) corresponds to a TPC value of 0.35, equal to the

volume fraction of the fibers, while the center of the cross-correlation image in Figure

7-2(c) is equal to 0, indicating that there is no region where a combination of both the

phases are present in any single pixel [192]. Further details on the computation of the

TPC statistic, where the convolutional theorem is applied to use the Fast Fourier

Transform (FFT), can be found in [195] [192] [195] [188]. The TPC calculations for this

study were performed using the Python-based PyMKS library [197].

(a)   (b)   (c)
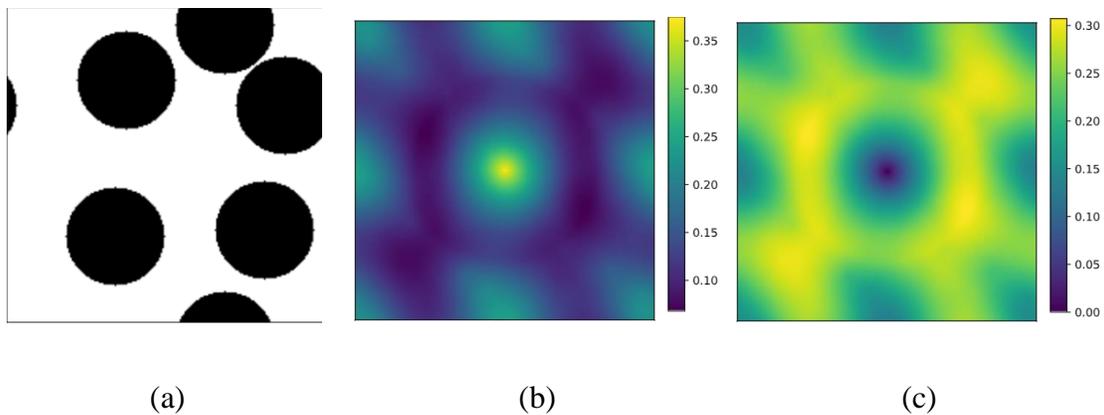
Figure 7-2: (a) Microstructure of a UD-FRC with a fiber volume fraction of 0.35, and the
contours of corresponding (b) two-point autocorrelation function $f(0,0)$, and (c) two-
point cross-correlation function $f(0,1)$. The centers of (b) and (c) correspond to the
origin, and the total image sizes are 199 x 199 pixels.

7.2.4   Principal Component Analysis for Dimensionality Reduction

The dimensionality reduction technique of principal component analysis (PCA) has been

utilized in other image-based machine learning efforts [181] [198] [199]. Although the

principal components lack a precise physical meaning, research suggests that a detailed

study of the PCs can provide new insights into comparative microstructural image

analysis [182].

For the m pixel square microstructure images generated in this study, the number of TPCs would be equal to $(m-1)^2$. Given the size of microstructures, e.g., 200 pixels x 200 pixels, this would entail a large number of inputs into the machine learning algorithm, making the predictive effort extremely tedious and inefficient in terms of computations. To reduce the dimensionality of the inputs and maximize the variability, a principal component analysis (PCA) was performed. PCA projects the data to a set of orthogonal coordinates corresponding to the directions with the most variance in the input data [200] [201] [181]. Principal components (PCs) are found from an eigenvalue decomposition of the covariance matrix [201], meaning that they are a linear combination of the original feature variables [199]. The covariance matrix is a square symmetric matrix of size p (the number of TPCs) that is typically denoted as:

$$C = \begin{bmatrix} E[(x_1 - E[x_1]) * (x_1 - E[x_1])] & \dots & E[(x_1 - E[x_1]) * (x_p - E[x_p])] \\ \vdots & \ddots & \vdots \\ E[(x_p - E[x_p]) * (x_1 - E[x_1])] & \dots & E[(x_p - E[x_p]) * (x_p - E[x_p])] \end{bmatrix} \tag{48}$$

where $E(x)$ is the expected value (mean) of the TPCs. The eigenvectors of the covariance matrix C are solved as:

$$V^{-1}CV = D \tag{49}$$

Where D is the diagonal matrix of size p × p containing the eigenvalues of the covariance matrix C, and V is a matrix also of size p × p containing the eigenvectors, which are orthogonal vectors in this case because the covariance matrix is symmetric. The eigenvalues and their corresponding eigenvectors are sorted in order of greatest to least to rank the PCs by the fraction of the variance they capture. Consider a matrix A of n data points (microstructures) by p number of TPCs which is centered by subtracting the mean from each row, represented as the column vector $\boldsymbol{\mu}$ of size n x 1:

171

$$A_{centered} = A - \mu = \begin{bmatrix} f_{n=1,p=1} & \cdots & f_{n=1,p=40,000} \\ \vdots & \ddots & \vdots \\ f_{n=750,p=1} & \cdots & f_{n=750,p=40,000} \end{bmatrix} - \begin{bmatrix} E[A] \\ \vdots \\ E[A] \end{bmatrix} \qquad (50)$$

To project this data into the new PC basis, take only the first r columns (r < p) of the

sorted V matrix and multiply by matrix $A_{centered}$:

$$T = A_{centered} * V_r \qquad (51)$$

One representation of the principal component analysis is through singular value

decomposition (SVD). Consider a matrix A of N data points by P number of TPCs with a

singular value decomposition [181]:

$$A = \begin{bmatrix} f_{n=1,p=1} & \cdots & f_{n=1,p=P} \\ \vdots & \ddots & \vdots \\ f_{n=N,p=1} & \cdots & f_{n=N,p=P} \end{bmatrix} = U\Sigma W^T \qquad (52)$$

Where U is a N × N matrix whose columns are orthogonal unit vectors called the left

singular vectors, Σ is a N × P diagonal matrix of the positive singular values σ_i, which

are the square root of the principal component variances, and W^T is a P × P matrix whose

columns are orthogonal unit vectors called the right singular vectors. The magnitude of

the singular values represents the fraction of the input variance they capture. The

principal component axis is ranked by sorting the singular values and their corresponding

singular vectors in decreasing order (highest magnitude to lowest). Transferring the

matrix A into the PC basis using only r terms (r < P) is completed by taking only the first

r singular value columns:

$$T = U * \Sigma_r = A_r * W_r \qquad (53)$$

In this study, the PCA was also performed using the Python-based PyMKS library [197].

The principal components are used to quantify the geometry of the microstructure. The maximum number of PCs is limited to the number of features (TPCs) or the number of data points (microstructure images) in the dataset [200]. In the case of the UD-FRC, there were 750 data points, corresponding to 125 microstructures each for 6 volume fractions used in the analysis, while in the case of UHP mortar and MPR paste there were 500 data points with 125 microstructures each for the 4 distributions and particle shapes. Figure 7-3(a) and (b) show the first and second PC feature space maps (or contours) for the $f(0,0)$ TPC of the UD-FRC dataset. The first PC provides information on the most dominant geometric feature of the microstructure, which for the UD-FRC, is the volume fraction of the inclusion phase [201] [181]. As shown in Figure 4(a), the highest weights are located about the center of the TPC image. Figure 4(b) shows the second PC that likely indicates the fiber size and its constant distribution [181]. It is well known that, while PCs of higher order are also important and represent other spatial characteristics of the microstructure, their interpretation is not straightforward [181]. Moreover, the efficiency of dimensionality reduction also drops when a larger number of PCs need to be included in the analysis. A plot of the percentage variance in the UD-FRC images as a function of the number of PCs used in analysis is shown in Figure 7-3(c) to demonstrate that increasing the number of PCs beyond a certain value does not add additional value to the predictive capacity of the model. This plot shows that 5 PCs account for ~95% of the variance in the UD-FRC dataset, while 7 PCs accounted for ~95% of the variance in the combined UHPC mortar and MPR paste dataset.

173
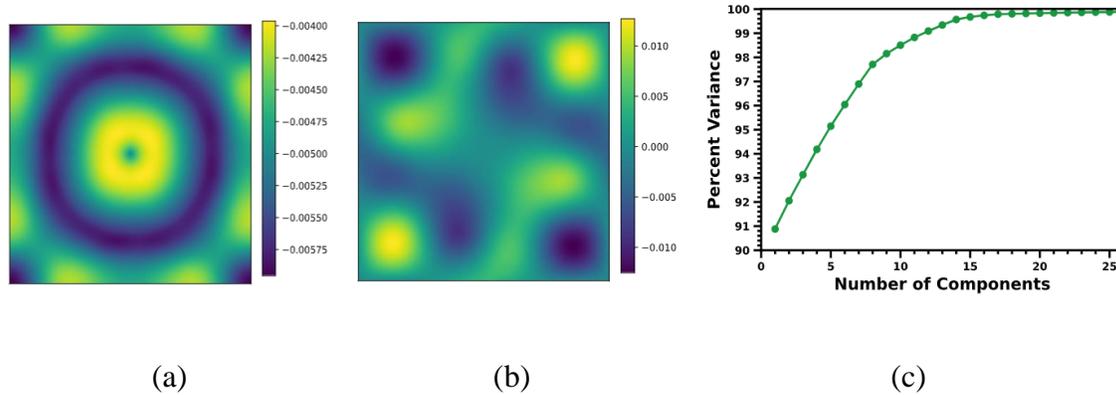
Figure 7-3: Feature space maps for: (a) 1st PC, and (b) 2nd PC for the UD-FRC dataset. The centers of (a) and (b) correspond to correspond to the origin, and the total image sizes are 199 x 199 pixels (same as the TPC). Graph (c) is the UD-FRC dataset percentage variance as a function of number of PCs.

7.3 FINITE ELEMENT ANALYSIS

Following the generation of the microstructure images, the .txt files were input into the Coreform Trelis™ software for microstructural meshing. A two-dimensional plane strain finite element model was developed in ABAQUS format using CPE4 elements. These meshed microstructures were input into a publicly available finite element solver to determine the elastic modulus in the 2-2 direction (E_{22}), the Poisson's ratio (v_{23}), and the average stress in the matrix (σ_{mat}). This finite element solver was developed and validated from experimental results studying the elastic behavior of T800 carbon fiber strands in a F3900 matrix at a strand volume fraction of 0.60. As such, the microstructural meshing for the UD-FRC analysis assigned strand elements the T800 elastic modulus of 2250000 psi (~15.51 GPa) and a Poisson's ratio of 0.25 and matrix elements the F3900 elastic modulus of 409000 psi (~2.82 GPa) and a Poisson's ratio of 0.387. Table 7-2 details the properties of the matrix and inclusions assigned for finite element analysis of each microstructure studied. The boundary conditions and direction

174

that the unit strain was applied (tensile for UD-FRC and compressive for the UHP mortar

and MPR paste) is illustrated in the meshed microstructure shown in Figure 7-4(a).

Table 7-2. Details of the two-phase microstructural images generated and analyzed.
UHP stands for ultra-high performance, E is the Young's modulus, and ν is the
Poisson's ratio.

Two-Phase Microstructure	Matrix	E_{matrix} (GPa)	ν_{matrix}	Inclusion	$E_{inclusion}$ (GPa)	$\nu_{inclusion}$
UD-FRC	F3900 plastic	2.82	0.387	T800 carbon fiber	15.51	0.25
UHP mortar	UHP cement paste	36	0.24	Quartz Sand	72	0.24
MPR paste	Cement paste	27	0.24	Steel	200	0.29

(a)   (b)
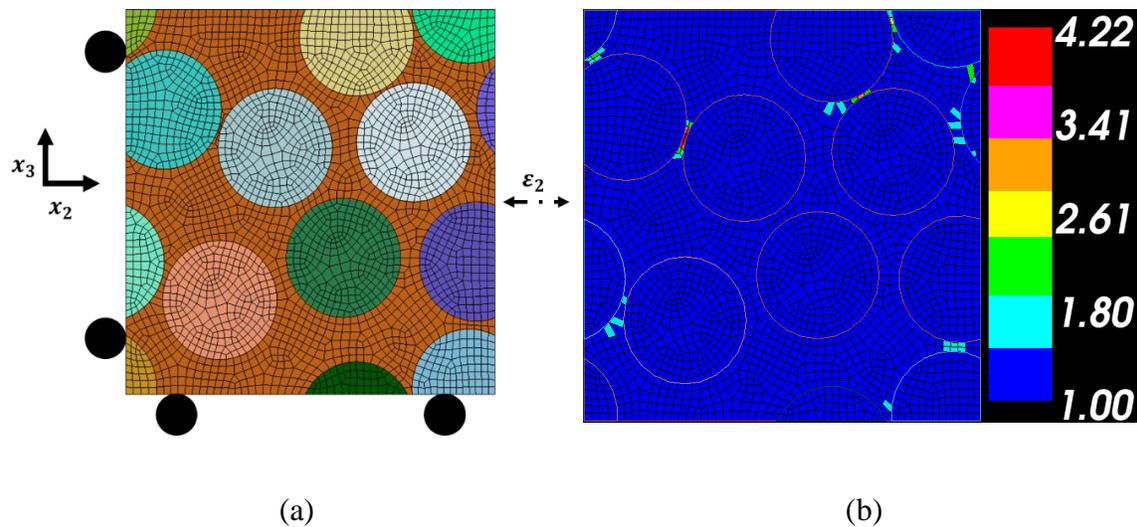
Figure 7-4: Microstructure with a volume fraction of 0.60: (a) After meshing with
boundary conditions, coordinate axis, and direction the unit strain shown, and (b) Color
map of the aspect ratio of each finite element.

A convergence analysis testing the image window size and fineness of the mesh was

performed using a volume fraction of 0.60 to compare to the experimental results. The

square image size was varied from 100, 200, and 300 pixels long and the auto factor

mesh size in Trellis varied from 1 (finest) to 7 (coarsest). This convergence analysis

found that auto factor 1 with the 200 pixel window size had the smallest difference, about

-4% and -17% respectively, between the experimentally determined E_{22} and v_{23} and those predicted by the finite element analysis. As shown in Figure 7-4(b), auto factor size 1 produced a mesh where almost all elements had an aspect ratio of 1. Only the tight regions between strands produce elements with aspect ratios greater than 1.

7.4 MACHINE LEARNING ALGORITHMS AND APPROACH

A concise overview of the different machine learning (ML) techniques used in this paper is provided here.

7.4.1   Regression Machine Learning Algorithms

Artificial neural network (ANN) and forest ensemble methods are the ML algorithms used for the regression-based predictions reported in this paper. ANN is a preferred ML algorithm for many materials-related problems [9] [12] [95] [14] [157] due to its ability to model nonlinear functions. Inside each neuron of the ANN, the inputs are evaluated by an activation function [95] before being multiplied by the weighted connections tying each neuron to every subsequent neuron. The chosen activation function for this study is the rectified linear unit (ReLu) with weight optimization performed using RMSprop, which features an adaptive learning rate formula [96]. Backpropagation, using the gradient of the previous iteration to train the weights of the ANN, was performed automatically by the Keras neural network framework written in Python to build and train the ANNs [158]. To minimize over-fitting, a dropout rate, i.e., the probability that any neuron and its connections will be temporarily excluded from the network, was incorporated into the ANN [97]. Additional details on ANNs can be found in [89] [11] [198].

Machine learning forest ensembles are based on decision trees that find logical splits in the data leading from one branch to the next until ending at the leaf node [9] [87]. To reduce prediction inaccuracy and over-fitting, the predictions from a collection of decision trees are bagged [9] [159], termed ensembles. In the Random Forest (RF) ensemble, the best split of the data is determined by considering all of the input features and checking a criterion, such as Gini impurity, to select the most discriminative threshold [92] [87]. Each individual decision tree in the RF ensemble does not use the entire set of training data, but a bootstrap sample made from subsets of the training data with replacement [87] [159]. In the Extra Trees (ET) ensemble, the splits are drawn at random for each feature and the best split, as measured by the chosen criteria, is selected as the splitting rule [92] [87]. In the ET regression model, the entire dataset is incorporated into each individual tree [92]. The prediction results of the individual trees are averaged to produce the output prediction in the RF and ET regressions. In a Gradient Boosted Tree (GBT) ensemble, an initial tree is trained with the entire dataset. All subsequent trees in the forest are trained to minimize the residual between the predicted and actual values of the previous tree [9] [92] [93]. The final prediction is calculated as the weighted sum of the predictions of each tree. For each tree beyond the first, the prediction is multiplied by the learning rate, with typical values between 0.01 and 0.1 [9] [92]. A specialized form of the GBT is Extreme Gradient Boosted (XGB) tree [93]. XGB performs shrinkage and column subsampling techniques to prevent overfitting between boosted trees and additionally offers scalability through parallel tree boosting (efficient computing regardless of data size) [93]. Additional details on ensemble methods can be found in [9] [87] [166].

177

7.4.2 Data Processing

For the UD-FRC case, the sets of 125 images for each volume fraction were shuffled together such that the selection of images for testing and training would include every volume fraction. A similar process was adopted for the sand mortar and the particulate reinforced cement paste case. The entire set of data was split such that 75% of the images were assigned to the training set and 25% of the images were assigned to the testing set. The output data points were normalized based on MinMaxScaler in scikit-learn [92] (Equation 8) before separation into the testing and training sets in order to avoid the ML weight assignments skewing in favor of the larger values [87].

$$z_{new} = \frac{z - z_{min}}{z_{max} - z_{min}} \tag{54}$$

where z_{new} is the value of the variable after transformation, z is the current value of the variable, z_{min} is the minimum value of that variable, and z_{max} is the maximum value of that variable. After training and predicting, the normalized test data is converted back to the original scale using the "inversetransform" function.

Training was performed by fitting the ML algorithms to the training data set. Model performance was evaluated using the testing data set, which the ML algorithms had not seen yet, and measuring the resulting errors. The mean squared error (MSE), given in Equation 9 was used as the objective function to be minimized.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (P_i - A_i)^2 \tag{55}$$

where n is the total number of data points, A_i is the actual value, and P_i is the predicted value. Other metrics tracked, but not used to train the models, were the mean absolute error (MAE) and the coefficient of determination (R^2), given as:

$$MAE = \frac{1}{n} \left(\sum_{i=1}^{n} \left| \frac{A_i - P_i}{A_i} \right| \right) \tag{56}$$

178

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - A_i)^2}{\sum_{i=1}^{n}(A_i - \overline{A})^2} = \left(\frac{n(\sum_{i=1}^{n} A_i * P_i) - (\sum_{i=1}^{n} A_i) * (\sum_{i=1}^{n} P_i)}{\sqrt{n(\sum_{i=1}^{n} A_i^2) - (\sum_{i=1}^{n} A_i)^2} * \sqrt{n(\sum_{i=1}^{n} P_i^2) - (\sum_{i=1}^{n} P_i)^2}} \right)^2 \tag{57}$$

Where $\overline{A}$ is the mean of the actual values.

Hyperparameter optimization was accomplished by utilizing the same number of epochs (i.e., each instance that the data is processed through an ML algorithm) to train MLs and compare their accuracy. Coarse optimization of the hyperparameters followed a random search pattern, found to be the most efficient method to optimize parameters [164], by generating 20 different random combinations of hyperparameters. The hyperparameters for random testing were chosen from the uniform distributions shown in Table 7-3. These values were chosen based on our previous works, where the optimization process is explained in detail [155]. As discussed earlier, the number of PCA components to be included as inputs also had to be tuned depending on the material used and property predicted.

For all the models, the parameters which minimized the 5-fold cross-validation MSE were used as the basis for the final models, with some additional fine-tuning. In every ML model the number of PCA components to include as inputs had to be tuned. For the linear regression model, the degree of the basis function was optimized. The parameters to optimize in the ANN models were the number of hidden layers, the number of neurons in each hidden layer, and the dropout rate. ReLu activation function with a learning rate of 0.001 and an RMSprop optimization scheme was used. For the RF, ET, and GBT models, the number of trees in the forest, the maximum depth of the trees, the minimum number of samples before splitting, and the minimum number of samples per leaf were tuned. For the XGB models, there are many hyperparameters available to tune, ranging

from structure-based, such as the depth of the trees or the number of GBTs, to how splits

are made via the subsample and colsample_bytree parameters, or even how big the leaf

groups can be via min_child_weight. Three parameters, the number of trees, the

maximum depth, and the learning rate, were found via grid searching to be the most

influential and were used for optimization in this study. Detailed breakdown of the

allowed ranges and significance of each of these hyperparameters are given in the XGB

code documentation [93].

Table 7-3. Hyperparameters tuned based on a uniform distribution range of potential values.

Model	Hyperparameter	Uniform Distribution Range
All	# principal components (inputs)	[6, 20]
ANN	# hidden layers	[1, 3]
	# starting neurons	[9, 50]
	Drop rate	[0, 0.1]
Random Forest (RF), Extra Trees (ET) Forest, and Gradient Boosted Trees (GBT)	# of trees	[50, 200]
	Maximum depth	[2, 10]
	Minimum# of samples before split	[2, 10]
	Minimum # of samples on leaf	[2, 10]
XGB	# of trees	[20, 100]
	Maximum depth	[2, 20]
	Learning Rate	[0.01, 0.3]

To test the accuracy of the predictions under each set of test parameters, an n-fold cross-

validation technique was employed [9, 87, 11]. A 5-fold cross-validation, deemed

sufficient for the size of the data sets, was performed using the following steps: (i)

randomizing the data set and splitting into 5 folds, (ii) training the model with selected

parameters using 4 of the folds, (iii) testing the model using the remaining fold, (iv)

repeating steps (ii) and (iii) until each fold has been used for testing once, acquiring 5

independent performance measures, and (v) averaging the individual metrics to obtain the

cross-validation errors. The parameters which minimized the cross-validation error was

used as a basis for the final models with some additional fine-tuning searches. Final

hyperparameters for each ML and dataset is listed in Table 7-4.

Table 7-4. Final hyperparameters used for each ML model and dataset.

Model	Hyperparameter	UD-FRC		UHP mortar and MPR paste		
		E_{22} (GPa)	ν_{23}	E_{22} (GPa)	ν_{23}	σ_{mat} (MPa)
All	# principal components (inputs)	15	8	10	10	10
Linear	Degree	1	1	2	1	2
ANN	# hidden layers	1	8	3	3	2
	# starting neurons	25	3	40	28	20
	Drop rate	0.05	30	0.1	0.1	0.05
Random Forest (RF)	# of trees	123	145	120	181	122
	Maximum depth	9	8	2	10	5
	Min # samples before split	9	7	4	7	2
	Minimum # of samples on leaf	4	2	10	3	10
Extra Trees (ET) Forest	# of trees	122	75	130	181	53
	Maximum depth	8	8	2	7	9
	Min # samples before split	5	7	4	4	3
	Minimum # of samples on leaf	2	2	8	4	5
Gradient Boosted Trees (GBT)	# of trees	159	130	152	177	82
	Maximum depth	2	6	4	4	4
	Min # samples before split	5	2	4	6	4
	Minimum # of samples on leaf	10	5	4	6	6
XGB	# of trees	36	65	10	86	10
	Maximum depth	3	12	46	9	24
	Learning Rate	0.3	0.1	0.06	0.027	0.05

The number of PCs to include as inputs was determined first as part of the

hyperparameter optimization, and later verified with an ANOVA one-way analysis to

determine if changing the number of PCs for the final ML hyperparameter options would

change the MSE results of the five-fold cross validation. Using 1 to 15 PCs for all the

models, the average RMSE values from the 5-fold cross validation are shown in Figure

7-5 for the UD-FRC case. Only the linear regression models were found to improve

significantly with respect to the number of PCs included above the minimum 5 to

represent 95% of the variance for UD-FRC. One-way ANOVA at a 5% level of

significance revealed the first time the linear regression had a non-significant difference

between five consecutive means was at 8 PCs for v_{23} and 15 for E_{22}.

| (a) | (b) |
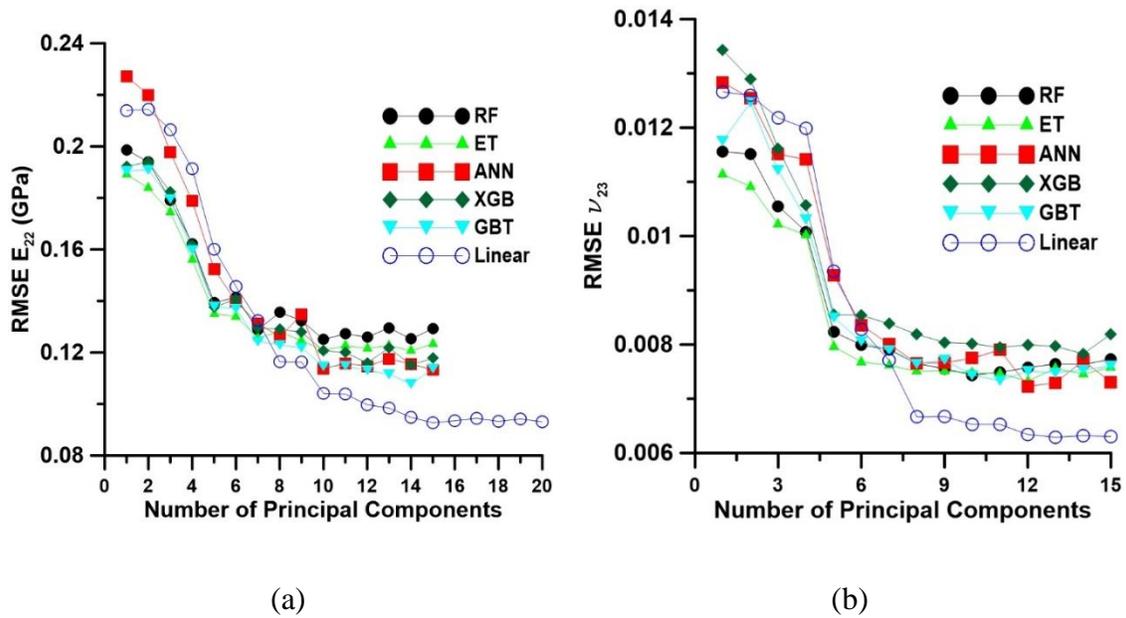
Figure 7-5: The 5-fold cross validation RMSE for each ML type using the UD-FRC
dataset as a function of the number of PCs included for: (a) E_{22} (GPa), and (b) v_{23}.

For the UHP mortar and MPR paste dataset, the principal component sensitivity analysis

did not present a clear pattern of decreasing error as the number of PCs increased, as

indicated in Figure 7-6. ML overfitting [97] means the generalizability to predict accurate

results on the testing data set is reduced as a result of over-training or memorizing the

outputs of the training data set. Overfitting can also be caused by incorporating many inputs such that the ML does not have enough data to accurately train the internal weights [90]. In the case of these microstructures, including more PCs becomes detrimental to the testing results, as especially seen in the linear ML for E_{22} analysis in Figure 7-6(a). Results from TPC analysis of the percentage variance as a function of the number of PCs, found that 7 PCs accounted for ~95% of the variance in the UHP mortar and MPR paste dataset. ANOVA analysis did not discern a single number of PCs where a non-significant difference between five consecutive means existed for all the MLs at the same time. Following the previous dataset in which the optimal number of PCs to include was a little more than the number needed to account for ~95% of the variance, the number of PCs to include in the UHP mortar and MPR paste dataset was chosen as 10 for all three of the mechanical outputs found from FEA.

(a) (b) (c)
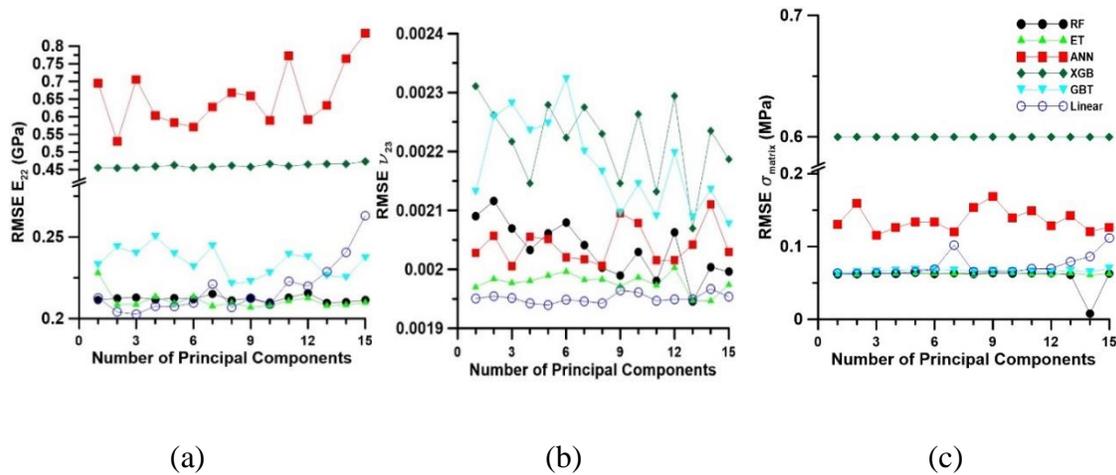
Figure 7-6: The 5-fold cross validation RMSE for each ML type using the UHP mortar and MPR paste dataset as a function of the number of PCs included for: (a) E_{22} (GPa), (b) ν_{23}, and (c) σ_{matrix} (MPa).

7.5 RESULTS AND DISCUSSIONS

The predictive efficiency of different ML models using TPC with principal component analysis as inputs to predict the elastic modulus, Poisson's ratio, and/or average matrix stress of two-phase microstructures is reported in this section. Each of the five ML algorithms (ANN, RF, ET, GBT, XGB) discussed above were implemented on two datasets. The first 750 point dataset was comprised of 125 randomly generated images for each of the 6 volume fractions (0.35, 0.4, 0.45, 0.50, 0.55 and 0.60) of carbon fiber in the UD-FRC studied. The second data set containing 500 randomly generated images was comprised of 125 microstructures of UHP mortars with three different size distributions (but with the same volume fraction, ϕ), as well as 125 images of the UHP mortar and MPR paste. A discussion of the results examines the applicability of the ML regression approach to analyze two-phase microstructures.

7.5.1 Unidirectional Fiber Reinforced Composites (UD-FRC)

The ML regression results for predicting the elastic modulus (E_{22}) of the microstructures is given in Table 7-5. From Table 7-5, it is clear that all of the ML methods performed well, where all 5-fold cross-validation average R^2 values exceeded 0.95, and the RMSE were all below 0.15 GPa. Example plots of the experimental vs. predicted E_{22} values are shown in Figure 7-7(a) for linear regression, the ML with the lowest MSE, and in Figure 7-7(b) for random forest, the ML producting the highest MSE. In these figures with $\pm$ 5% bounds shown, the prediction accuracy is best for the low and high extremes of volume fractions with 0.35, 0.55, and 0.60. Volume fractions near the middle of the range studied between 0.40 and 0.50 appear to have the highest errors which tend to underpredict in the linear model Figure 7-7(a), but equally under- and over-predict in the random forest

model Figure 7-7(b). Fundamentally, the weighting system of machine learners will favor

the extreme values [97] [89] because these values are the most sensitive to changes in the

weights and splits and are reflected in the MSE used to update the models.

Table 7-5. ML 5-fold cross-validation average and standard deviation values for predicting E_{22} for the UD-FRC dataset using 15 PCs. The most accurate ML model is shown in **bold**.

Model Type	RMSE E_{22} (GPa)	MAE E_{22} (GPa)	R^2
Linear	**9.38E-02 ±3.24E-02**	**7.40E-02 ±4.77E-03**	**0.990 ± 0.001**
Artificial Neural Network (ANN)	1.12E-01 ±4.36E-02	8.78E-02 ±8.06E-03	0.986 ± 0.003
Random Forest (RF)	1.26E-01 ±4.81E-02	1.00E-01 ±7.66E-03	0.982 ± 0.003
Extra Trees Forest (ET)	1.20E-01 ±4.78E-02	9.61E-02 ±8.02E-03	0.984 ± 0.003
Gradient Boosted Forest (GBT)	1.12E-01 ±2.43E-02	8.90E-02 ±3.16E-03	0.986 ± 0.001
Extreme Gradient Boosting (XGB)	1.13E-01 ±3.66E-02	9.10E-02 ±4.86E-03	0.986 ± 0.002
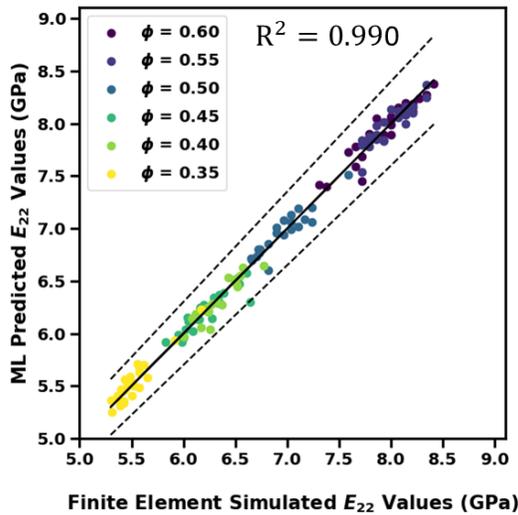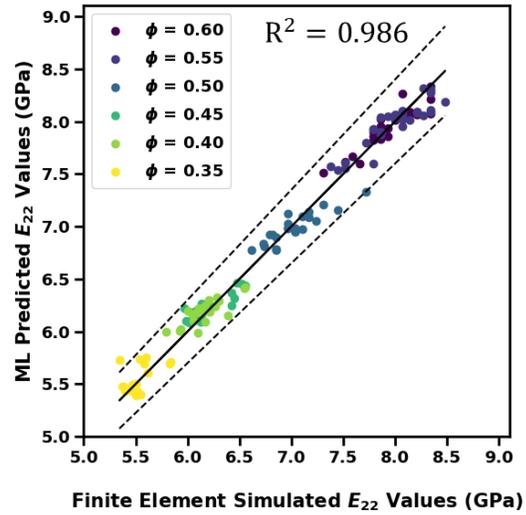
(a)   (b)

Figure 7-7: Test results for 1 fold (150 images) to predict the E_{22} (GPa) value for UD-FRC dataset using the machine learning techniques of: (a) Linear regression, and (b) Random Forest. The solid line represents the line of ideality, and the dashed lines represent a ± 5% bound. Note from Table 7-5 that linear regression had the lowest 5-fold average MSE of all MLs and Random Forest had the highest MSE.

Although these are complex and random two-phase microstructures, the fact that the linear regression model was the most accurate is not surprising as the experimental and finite element study of the T800 carbon fiber strands in a F3900 matrix found the approximate equation relating the elastic modulus in the 1-1 direction to the volume fraction (v_f) to be:

$$E_{11}(v_f) = 8 * 10^7 * v_f^2 + 4 * 10^{11} * v_f + 3 * 10^9 \tag{58}$$

This is in the transverse direction to the E_{22} studied, but the equation offers insight into how the volume fraction of carbon fibers has the highest influence on the elastic properties.

The results of ML regression of microstructural images to predict the Poisson's ratio v_{23} are shown in Table 7-6. The continued results of R^2 over 0.90 and a RMSE less than 8E-3 suggest that once again, all the machine learners were well-suited to relating the PCs of TPC to the mechanical properties. Figure 7-8(a) plots the experimental vs. predicted v_{23} values generated from for linear regression, while the Figure 7-8(b) results were generated from the extreme gradient boosting ML. In unidirectional fibers the major Poisson's ratio ($v_{12} = v_{13}$) obeys the rule of mixtures, and is therefore bounded between the Poisson's ratios of the matrix and inclusion, while the transverse Poisson's ratio (v_{23}) studied here is instead estimated as a function of the ratio of the transverse elastic modulus and transverse shear modulus [202] [203]. Therefore, it is not surprising that the finite element and ML estimation of v_{23} ranged from 0.45 to 0.57. The linear regression and extreme gradient boosting MLs were the best and worst, respectively, in terms of the 5-fold cross-validation average RMSE and R^2 values. One difference between the two models is that the linear regression evenly over and under predicts across all volume

fractions with an emphasis on under predicting the 0.60 volume fraction, while the

extreme gradient boosting features more extreme errors, almost cyclically under and

over-predicting beginning from 0.60 down to 0.35 volume fraction. Regardless, as shown

by the $\pm$ 5% bounds in Figure 7-8, these errors are within a reasonable tolerance for every

ML method studied.

Table 7-6. ML 5-fold cross-validation average and standard deviation values for predicting ν_{23} for the UD-FRC composite dataset using 8 PCs. The most accurate ML model is shown in **bold**.

Model Type	RMSE ν_{23}	MAE ν_{23}	R^2
Linear	**6.67E-03 ±2.33E-03**	**5.28E-03 ±2.91E-04**	**0.946 ± 0.006**
Artificial Neural Network (ANN)	7.38E-03 ±2.50E-03	5.96E-03 ±3.76E-04	0.934 ± 0.005
Random Forest (RF)	7.46E-03 ±2.58E-03	5.89E-03 ±4.12E-04	0.933 ± 0.005
Extra Trees Forest (ET)	7.40E-03 ±2.00E-03	5.84E-03 ±3.02E-04	0.934 ± 0.006
Gradient Boosted Forest (GBT)	7.48E-03 ±2.00E-03	5.93E-03 ±2.46E-04	0.932 ± 0.005
Extreme Gradient Boosting (XGB)	7.91E-03 ±2.42E-03	6.29E-03 ±2.85E-04	0.924 ± 0.007
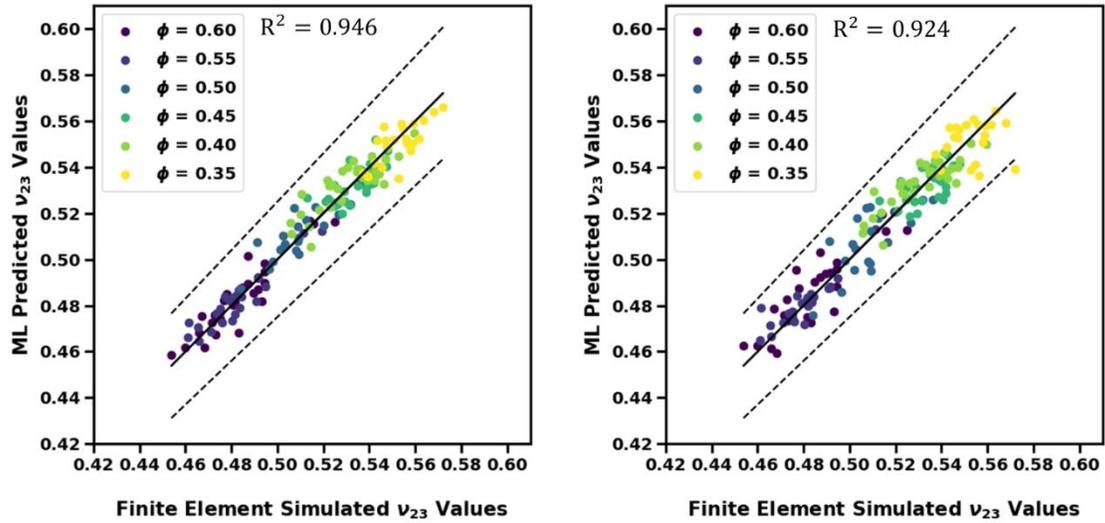
(a)   (b)

Figure 7-8: Test results for 1 fold (150 images) to predict the ν_{23} value of the UD-FRC dataset using the machine learning techniques of: (a) Linear regression, and (b) Extreme Gradient Boosting. The solid line represents the line of ideality, and the dashed lines represent a $\pm 5\%$ bound. Note from Table 7-6 that linear regression had the lowest 5-fold average MSE of all MLs and Extreme Gradient Boosting had the highest MSE.

Compared to the E_{22} results, the ν_{23} results had lowered R^2 values, but a reduction in the variance from the mean cross-validated results. The Poisson's ratio in composites systems is affected by the clustering of the strands [174]. In image statistical descriptors, the lineal clustering or connectedness between features is best modelled by the lineal-path function, not the TPC [204]. Instead, the TPC function focuses on the short-range information about different clusters as well as morphological information larger than the maximum cluster size [93]. The TPC function provides an overall description of the strand locations that generated quite accurate predictions of the ν_{23}, as shown by the ML results, but TPC alone may not be the best image descriptor for the ν_{23} mechanical property.

7.5.2 Ultra High Performance (UHP) Mortar and Metallic Particulate Reinforced Cement (MPR) Paste

The second dataset consisting of UHP mortar and MPR paste with inclusions of different sizes, shapes, and volume fractions, tests the ML algorithms' ability to learn from two-phase systems with multiple microstructural differences. Table 7-7 and Table 7-8 display the consistently accurate results of all the ML classifiers to predict the Young's modulus (E_{22}) and average stress in the matrix (σ_{mat}) values, respectively. Unlike the UD-FRC microstructures where the simple linear ML was sufficient to link the TPC and PCA to the microstructural properties, these UHP mortar and MPR paste microstructures with diverse particle sizes and shapes were best represented by ensemble ML methods.

As shown in Figure 7-9, the E_{22} (Figure 7-9 (a)) and σ_{mat} (Figure 7-9 (b)) ML prediction results were close to the plane strain linear-elastic FEA values across all of the given input microstructures. It is known that the particle size distribution of the inclusions in two-phase composites such as mortars have a significant effect on the average microstructural stress and the stress-path within the material [4]. Yet as seen in Figure 7-9, there was not a significant difference in the FEA simulation values, and therefore in the ML predictions, between the three sand-mortar distributions. This similarity in FEA output values are an effect of the small compressive strain applied to the microstructure. Future work will address ML prediction of the non-linear microstructural response, where the effects of particle size distribution, orientation, and shape produces significant mechanical differences among the microstructures.

Table 7-7. ML 5-fold cross-validation average and standard deviation values for predicting E_{22} for the UHP mortar and MPR paste dataset using 10 PCs. The most accurate ML model is shown in **bold**.

Model Type	RMSE E_{22} (GPa)	MAE E_{22} (GPa)	R^2
Linear	2.09E-01 ±1.11E-01	1.49E-01 ±1.90E-02	0.999 ±3.931E-04
Artificial Neural Network (ANN)	5.89E-01 ±3.61E-01	3.88E-01 ±9.79E-02	0.990 ±3.654E-03
Random Forest (RF)	2.10E-01 ±9.32E-02	1.48E-01 ±1.56E-02	0.999 ±2.993E-04
Extra Trees Forest (ET)	**2.08E-01 ±9.98E-02**	**1.48E-01 ±1.69E-02**	**0.999 ±3.118E-04**
Gradient Boosted Forest (GBT)	2.28E-01 ±1.02E-01	1.55E-01 ±1.51E-02	0.998 ±3.911E-04
Extreme Gradient Boosting (XGB)	4.66E-01 ±1.19E-01	4.19E-01 ±1.63E-02	0.993 ±1.042E-03

Table 7-8. ML 5-fold cross-validation average and standard deviation values for predicting σ_{mat} for the UHP mortar and MPR paste dataset using 10 PCs. The most accurate ML model is shown in **bold**.

Model Type	RMSE σ_{mat} (MPa)	MAE σ_{mat} (MPa)	R^2
Linear	6.56E-02 ±2.72E-02	4.73E-02 ±4.76E-03	0.999 ± 0.000
Artificial Neural Network (ANN)	1.40E-01 ±6.55E-02	1.10E-01 ±1.07E-02	0.994 ± 0.001
Random Forest (RF)	**6.26E-02 ±3.33E-02**	**4.45E-02 ±5.24E-03**	**0.999 ± 0.000**
Extra Trees Forest (ET)	6.31E-02 ±3.46E-02	4.45E-02 ±6.21E-03	0.999 ± 0.000
Gradient Boosted Forest (GBT)	6.67E-02 ±3.92E-02	4.74E-02 ±7.35E-03	0.999 ± 0.001
Extreme Gradient Boosting (XGB)	6.00E-01 ±6.72E-02	5.95E-01 ±3.42E-03	0.888 ± 0.009
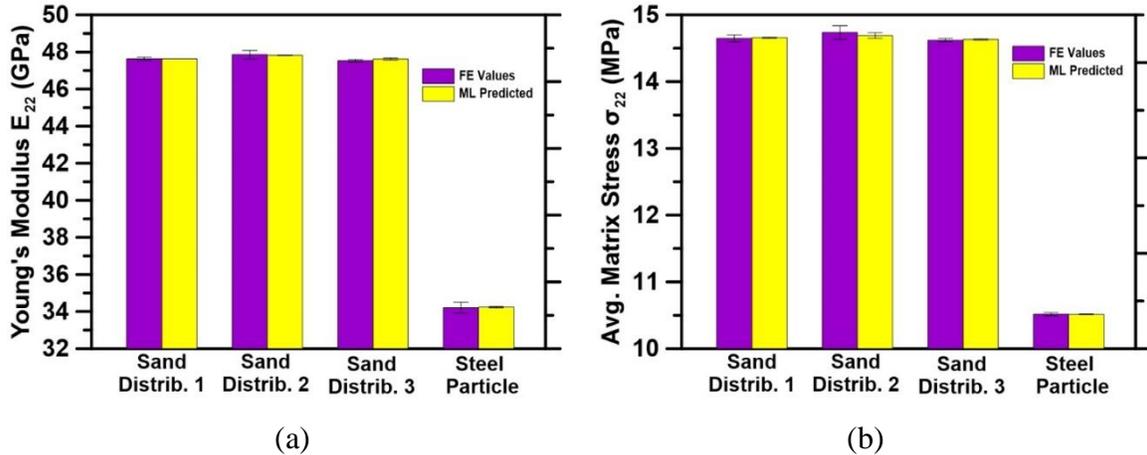
Figure 7-9: Bar charts of the UHP mortar and MPR paste dataset results from one fold of cross-validation testing comparing the FEA simulation results to the ML predictions for the outputs of: (a) E_{22} (GPa) predicted using ET ML, and (b) σ_{mat} (MPa) predicted using RF ML.

The analysis of these microstructures suggests that TPC with PCA inputs were able to distinguish between the microstructures and the ML was accurately able to learn and predict the resulting mechanical properties. This general applicability of the TPC with PCA means if a similar strain is applied to any two-phase composite system, the MLs can produce initial insights into the microstructural properties without extensive laboratory testing and imaging. An important caveat to any ML algorithm is that the model can only predict within the domain which it was trained on, such as the applied displacement or the mechanical properties of the matrix and inclusions.

Table 7-9 demonstrates the effects of non-significantly different output values being used to train the ML algorithms, where the results of the ML algorithms on the Poisson's ratio values are shown. For the entire 500 point UHP mortar and MPR paste dataset, the average ν_{23} was 0.316 ± 0.003. As a result, the ML algorithm was training to predict only the noise in the FEA results, not directly related to the differences in the microstructures. The average 5-fold RMSE results are 3 times smaller than those of Table 7-6 for the UD-

192

FRC dataset, however the coefficient of determination is reduced to around ~0.65. Such

small errors combined with such a low R^2 value suggest that the ML was overfitting

instead of learning meaningful results from the TPC and PCA.

Table 7-9: ML 5-fold cross-validation average and standard deviation values for predicting v_{23} for the UHP mortar and MPR paste dataset using 10 PCs. The most accurate ML model is shown in **bold**.

Model Type	RMSE v_{23}	MAE v_{23}	R^2
Linear	**1.96E-03 ±1.11E-03**	**1.33E-03 ±1.92E-04**	**0.667 ± 0.125**
Artificial Neural Network (ANN)	2.08E-03 ±9.95E-04	1.39E-03 ±1.17E-04	0.632 ± 0.088
Random Forest (RF)	2.03E-03 ±1.18E-03	1.37E-03 ±2.30E-04	0.644 ± 0.137
Extra Trees Forest (ET)	1.99E-03 ±1.14E-03	1.34E-03 ±2.08E-04	0.659 ± 0.127
Gradient Boosted Forest (GBT)	2.15E-03 ±1.16E-03	1.46E-03 ±2.36E-04	0.602 ± 0.138
Extreme Gradient Boosting (XGB)	2.26E-03 ±1.18E-03	1.58E-03 ±1.65E-04	0.557 ± 0.144

7.5.3 Patterns of PCs for Microstructural Analysis

Principal component analysis alone does not have physical meaning because it is a

transformation of the input data into a new basis [181] [182]. However, when comparing

and contrasting the PCs for each of the microstructure types addressed in this study,

patterns in the feature maps emerge, that can be used as a basis for understanding the

most significant differences between microstructural features.

Figure 7-10 shows a grid of the first three principal components of each microstructure

type described in Table 7-1. When developing an ML with TPC and PCA to learn all the

two-phase microstructures, only 5 PCs are needed to explain 95% of the variance. The

most prominent explanation of the variance between all the microstructures, represented

in PC 1 of group (a), stems from the particle shape since 125/1250 of the images are

ellipses and the remaining images are circular. In this regard, the PC patterns for group (c) result from the changes in both the particle size and, unique to any of the other microstructures, the ellipse-shaped inclusion orientation. Note how the boundaries between areas of light and dark are blurred and less well-defined in group (c) as compared to the other groups. This signifies that there are more possible variations in the particle location and distance relationship to other particles as compared to circular systems in which orientation does not matter.

PC 2 in group (a) for all 1250 of the microstructures is approximately the sum of PC 1 of group (b) for the UD-FRC and PC 1 of group (c) for the sand mortars. This means that PC 2 for (a), PC 1 for (b) and PC 1 for (c) occur along the same principal component axis. The microstructures of group (b) only vary in volume fraction, while the microstructures in (c) vary only in their particle size distributions, therefore this PC with a brighter (more positive) circle surrounded by a dark ring may represent the center to center distance between the particles. The likelihood of the autocorrelation to meet an inclusion phase beyond a certain distance (radius) from the center of the circular particles is a direct result of the packing (ϕ) and the available sizes of the inclusions.

Finally, the PC 3 of the 1250 full dataset (a) is identical in shape and magnitude to the PC 2 of the UD-FRC (b). This PC results from the fact that most of the dataset (700/1250) is the UD-FRC with varying ϕ. After establishing the shape of the particles, and the center to center distribution of the particles, the next axis of greatest variance is one unique to the problem of purely varying the volume fraction. Alone, the feature map representations of the PCs are not easily related to their original datasets, but when comparing datasets, as the ML algorithms inherently do when training, the PCs are an

194

excellent visual reference and source for meaningful comparison between the axis of

greatest variance.
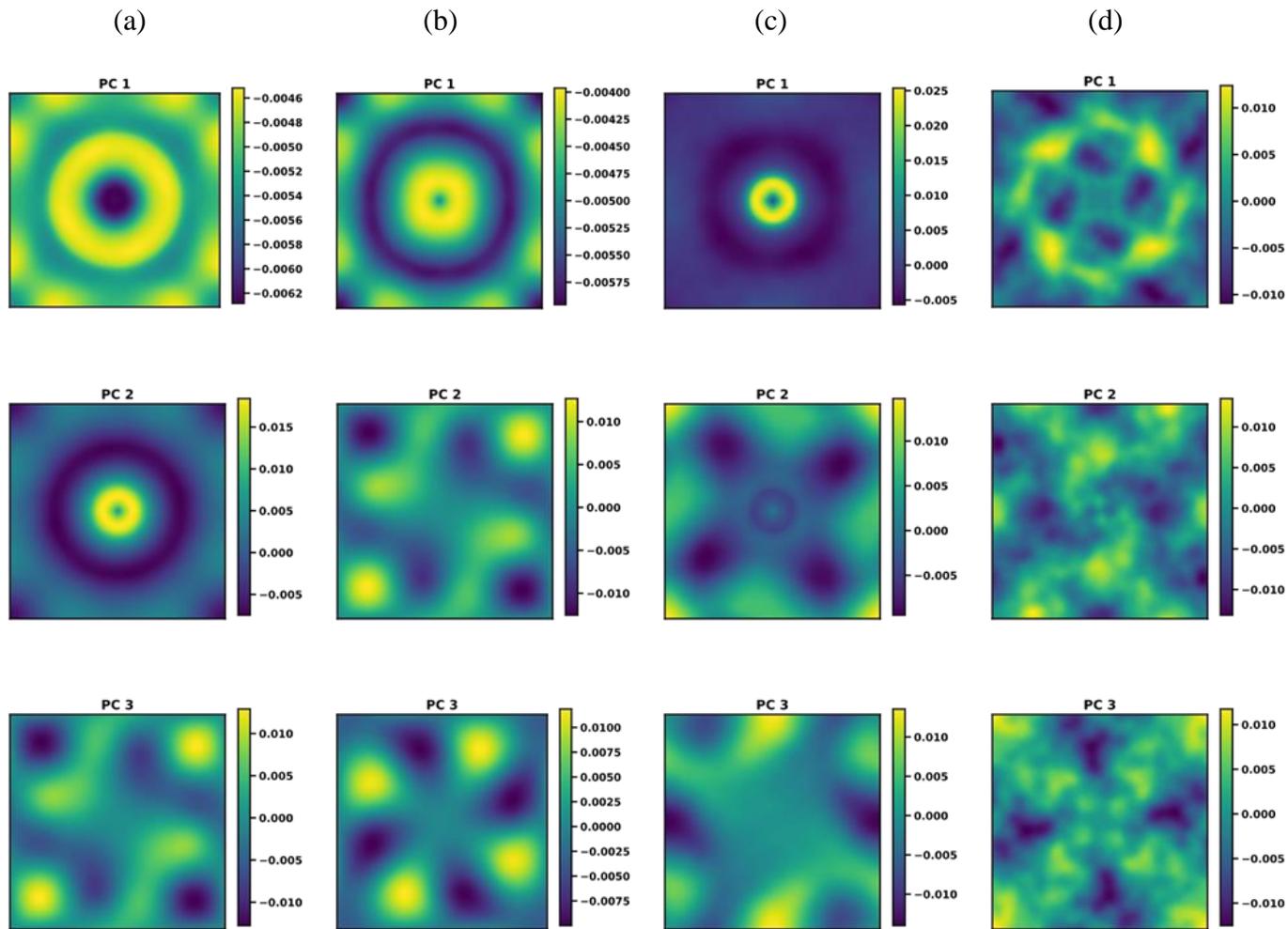
Figure 7-10: Feature maps of the first 3 principal components of: (a) all 1250 microstructural images examined in this study, (b) 700 UD-FRC images, (c) 375 cement paste with UHP mortar images, and (d) 125 MPR paste images.

7.6 SUMMARY

Although extensive finite element (FE) work has been dedicated to predicting the macroscale response of composite systems, it requires computationally intensive analysis. To speed up the process of evaluating the mechanical properties of a composite system given the properties of the individual components and a microstructure describing the phase morphology and distribution, image-based machine learning techniques can be applied. This work generated multiple machine learning techniques to analyze synthetic two-phase composite microstructural images to predict the mechanical properties. The "ground truth" homogenized mechanical property values for each image was determined using finite element analysis on FE models verified using available experimental results from previous studies. Two-point correlation functions (TPC) of the microstructures were determined, followed by dimensionality reduction using principal component analysis (PCA). These pre-processed images were used as inputs into the artificial neural network and ensemble machine learning algorithms. A sensitivity analysis using ANOVA one-way testing at a 5% level of significance was performed to determine the least number of PCs required as inputs to produce efficient and accurate ML results.

The linear regression models for the UD-FRC were more sensitive to the number of PCs beyond the 95% variance level, requiring 3-8 more PCs before the change in mean MSE became insignificant. For datasets containing microstructures with multiple variations, i.e. change in particle shape and change in particle size distribution, the number of PCs to achieve sufficient precision in prediction was not as evident. All the ML models produced highly accurate regressions of the E_{22}, ν_{23}, and σ_{mat} mechanical properties.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

8.1 OVERALL CONCLUSIONS

This study discussed the microscale mechanical and chemical characterization of UHP binders containing high volumes of commonly available cement replacement materials. Focus was placed on relating the constituent components to the resulting microstructural behavior, then predicting the behavior of cementitious systems through machine learning techniques. Machine learners to classify the cementitious phases and predict of the nanoindentation stiffness were developed. Finally, a visual machine learning methodology was established for two-phase composites as a first step to utilizing images to predict the microstructural behavior of cementitious systems. It was shown that coupling the nanomechanical properties to chemical intensities provides fundamental insights into novel, ultra-high performance binder systems, guiding their synthesis and offering a preview of the microstructural features that are key towards material performance. Conclusions for each chapter are presented in the following sections.

8.2 CONCLUSIONS FROM NANOMECHANICAL STUDIES

Chapter 3 discussed the nanoscale mechanical characterization of UHP cementitious matrices containing high volumes of commonly available cement replacement materials. Nanoindentation experiments coupled with a Bayesian information criterion were carried out to determine the average modulus and hardness of the statistical phases. The microstructure was found to be highly heterogeneous, attributable mainly to the low w/b used and multitude of cement replacement materials of differing size ranges. This study

explored how nanomechanical analysis reveals the enhanced influence of physical packing of particles in low w/b systems such as UHPC.

Key findings included:

- LD and HD C-S-H along with an UHS phase was identified for the fly ash-based UHP paste (w/b ~ 0.20) after 30 days of hydration, while only the UHS phase was observed for the fly ash-based paste after 90 days of hydration.

- For the microsilica-based UHP paste, the volume fractions of LD and HD C-S-H phases were so low that the points were clustered together with the UHS phase at all ages.

- The UHS phase was shown to be a nanocomposite of HD C-S-H phase and nanoscale CH. The reduced w/b in UHP pastes as well as the presence of several fine materials such as limestone powder and microsilica was postulated to enhance the formation of such a product through the absence of capillary spaces that would likely favor CH nanocrystal formation. The formation of such a product was supported using thermogravimetric analysis and nanogranular C-S-H packing density calculations.

- Both the UHP pastes contained several mixed/composite phases as determined from the M-H response, which is not uncommon in packed systems with multiple blends of fine, incompletely reacted particles. This renders the indentation-based analysis of such heterogeneous systems more cumbersome; however, the presence of phases stiffer than the hydrates function as micro-inclusions that enhance the properties of the UHPC.

199

- Analytical homogenization models based on Eshelby's solution for inclusions embedded in a matrix were used to upscale the elastic response of the individual phases in UHP paste to the elastic response of UHPC. For the fly ash-based UHPC, the homogenized E was in good agreement with that determined experimentally using strain controlled compression tests on UHPC specimens.

- For the microsilica-based UHPC, the upscaled E was found to be higher than that from experiments, likely due to potential agglomeration of microsilica in the UHP paste. In the presence of coarse aggregates, these agglomerates would be broken down by the shearing action of aggregates, reducing the volume of what were identified as mixed phases. The local heterogeneity in the microstructure in such cases are more likely to be significant in upscale property prediction. While increasing the number of grids and indents might appear to be a straightforward solution, it is unlikely to account for effects that are fundamentally different in pastes and concrete – e.g., ultra-fine particle agglomeration.

Nanomechanical properties and analysis of the intensities of relevant chemical species at the indentation locations for two UHP pastes containing multiple starting materials were reported in Chapter 4. Nanoindentation experiments were carried out on multiple grids at different ages, while the chemical analysis of the indentation points was carried out using qualitative EDS analysis, after performing a grid alignment procedure. Statistical clustering analysis of the mechanical and chemical data was performed assuming a Gaussian distribution by minimizing Bayes Information Criterion (BIC). The relationships between M, H, and the different chemical intensities were used to infer the fundamental nature of the reaction products in systems with high degree of heterogeneity,

aided by the very low w/b and the presence of multiple cement replacement materials/fillers. It is shown that coupling the nanomechanical properties to chemical intensities provides fundamental insights into novel, ultra-high performance binder systems, guiding their synthesis and offering a preview of the microstructural features that are key towards material performance.

Key findings included:

- The UHP pastes consisted mostly of HD C-S-H and an UHS phase, along with mixed phases comprising of partly reacted starting materials and some reaction products. The normalized chemical intensities and ratios of Ca, Si, and Al species, along with the mechanical property description provided by nanoindentation, allowed for further insights into the microstructure of complex, heterogeneous systems such as UHPC pastes.

- The scatter in the Ca/Si intensity plots for the UHPC pastes were found to be higher than those reported for conventional OPC pastes. The HD C-S-H and UHS phases were observed to have similar Ca and Si intensities, even though the spread was larger for the UHS phase, attributable to its lower Ca/Si ratio. However, the Al incorporation was higher in the UHS phase.

- For both the UHP pastes, the Ca/Si ratio for the UHS phase demonstrated an elliptical spread, with reduced heterogeneity being visually identifiable for the ML paste. The reduced heterogeneity in this mixture (fewer starting materials, and the presence purer chemical species) also restricted the spread of the Ca/Si ratio.

- The relationship between M and H of the reaction product phases in both the UHP pastes were found to be very similar, irrespective of significant changes in mixture composition. In general, M and H were found to be not highly dependent on the Ca/(Si+Al) ratio for the considered mixtures.

- The Ca/Si ratio was found to be an important parameter influencing the incorporation of Al in the gel, with its reduction promoting the incorporation of more Al into the C-A-S-H gel. In fact, Al incorporation is less dependent on the total amount of aluminates in the system, and more on the reduction in Ca/Si ratio.

8.3 CONCLUSIONS FROM ML FOR MICROSTRUCTURAL CLASSIFICATION AND REGRESSION

Chapter 5 presented a novel approach to accurately predict cement hydration phases from chemical intensity maps, using ML methods. Chemical intensity data from SEM-EDS for different UHP cement paste datasets representing multiple cementing materials and hydration ages were combined. Micromechanical information from nanoindentation as well the elemental intensities from qualitative EDS maps were then coupled with Bayesian statistical clustering. With the phase labels (e.g., LD or HD C-S-H, clinker etc.) thus identified, different ML classification techniques based on Artificial Neural Networks (ANN) and forest ensemble methods were implemented on the dataset. The classification algorithms were implemented on the 5-input dataset (chemical intensities of Ca, Si, Al, and Fe, along with the cube of the brightness of the BSE image (γ^3)), and 7-input dataset (the above 5 inputs, and M and H from nanoindentation). The area under the Receiver Operator Characteristic curve (ROC-AUC) was chosen as the indicator of

model performance. It is shown that chemical intensity mapping of microstructures, coupled with machine learning, can be used to accurately (in the case of common cementitious microstructures) classify the microstructural phases, which can lead to *apriori* property (e.g., stiffness) predictions. ML models can thus classify the cementitious component phase at locations in a microstructure to facilitate real-time characterization and first-order estimation of bulk properties.

Key findings included:

- For the combined dataset of the UHP pastes, the ROC-AUC values were higher than 0.90 for both the 7-input and 5-input datasets. The removal of nanoindentation information from the dataset did impact the efficiency of classification, as noted from the accuracy and F1 values.

- Confusion matrices demonstrated that the removal of nanoindentation information resulted in misidentification of some of the microstructural labels, especially where the chemical intensity data overlapped between multiple phases due to the unique composition of the UHP pastes. It was shown that, in such complex systems, the use of additional inputs in the form of nanomechanical properties help classification significantly.

- The same approach was also used on two less complex microstructures (i.e., fewer starting materials and more complete hydration), one of a plain OPC paste and the other a paste with 20% OPC replaced using a highly reactive natural pozzolan. Here, normalized intensities of just the three chemical species (Ca, Si, and Al) were deemed sufficient (without nanoindentation data) to generate a highly accurate classifier.

203

Chapter 6 presented a novel approach to predict the nanoindentation stiffness of cement hydration phases from chemical intensity maps, using ML methods, for the first time. To demonstrate the methodology and application, nanoindentation modulus and normalized chemical intensity data from EDS for four different data sets representing various cementitious mixtures with multiple cement replacement materials, w/b, hydration ages, and microstructural complexities were chosen. Four different classes of ML methods including Support Vector Regression (SVR), Gaussian Process Regression (GPR), Artificial Neural Networks with backpropagation (ANN), and ensemble methods (e.g., RF, ET, and GBT) were implemented on data sets comprising of ~400 to 2400 data records, each corresponding to a distinct nanoindentation point in the microstructure. The models were trained using 75% of the data and tested on the remaining 25% of the data. Established procedures for data analysis, including hyperparameter optimization, were implemented. EDS maps were obtained for the microstructures so that the local chemistry, which is easier to obtain than nanoindentation modulus, could be used to predict the hydration phase modulus using ML models. The relative intensity of the major chemical species (Ca, Si, Al, Fe), the mixture ages, as well as the local density of the microstructure represented by the cube of the brightness coefficient from backscattered SEM were used as the inputs. Estimating the mechanical properties of complex, heterogeneous cementitious mixes from ML will minimize the expensive and time required for testing as well as shift the focus of mix design from the macroscopic to the fundamental microscopic properties.

Key findings included:

- For the NP mixture, all ML models were able to predict the phase stiffness with an R^2 value ~ 0.90 using just three chemical intensities (Ca, Si, Al) as inputs, all of which appear in the chemical description of C-A-S-H gel. This was shown to be a result of the more homogeneous microstructure in this mixture, owing to a highly reactive cement replacement material, higher w/b, and a longer curing duration. GPR was shown to be the best model for this mixture in terms of the metrics (RMSE, MAE, R^2), attributed to its ability to work on sparse data sets with fewer inputs.

- When data sets belonging to NP, OPC, and SF mixtures were combined, the relative predictive efficiency dropped irrespective of the ML model used. The ensemble models showed better predictive ability in this case, attributable to the randomness of the bagging and input feature selection, generating ensembles with greater generalization.

- For the UHP pastes demonstrating complex microstructures as a result of multiple binding materials of varying chemical constitution and reactivities, and incomplete reaction due to low w/b, all the ML models were found to have lower predictive abilities. Specifically, it was noticed that the prediction quality was better for the C-S-H phases, but the large scatter in the data for the high modulus phases and the mixed reactant/product phases reduced the overall predictive ability.

- Potential options to tide over such inaccuracies include: increased representation of the mixed/unreacted clinker phases in the data sets to reduce skew-sensitivity,

205

more precise identification of the constituents of the mixed phase, additional

inputs such as information from neighboring indents, improved ML models that

account for imbalanced data, and/or image-based ML models.

Finally, Chapter 7 generated multiple machine learning techniques to analyze synthetic

two-phase composite microstructural images to predict the mechanical properties. The

"ground truth" homogenized mechanical property values for each image was determined

using finite element analysis on FE models verified using available experimental results

from previous studies. Two-point correlation functions (TPC) of the microstructures were

determined, followed by dimensionality reduction using principal component analysis

(PCA). These pre-processed images were used as inputs into the artificial neural network

and ensemble machine learning algorithms. A sensitivity analysis using ANOVA one-

way testing at a 5% level of significance was performed to determine the least number of

PCs required as inputs to produce efficient and accurate ML results. This work

demonstrated that ML can accurately predict mechanical properties for any two-phase

composite system, regardless of the inclusion volume fraction, shape, or size distribution.

Such TPC with PCA analysis is a first step to developing ML algorithms capable of

evaluating the mechanical properties of complex, multi-phase microstructural

composites, such as ultra-high performance binders, based on microstructural images.

Key findings included:

- Changing only the volume fraction, the ANN and ensembles models had a

  significant reduction in MSE as more principal components (PCs) were included,

  but only up to the number of principal components required to explain 95% of the

  variance. On the other hand, the linear regression models were more sensitive to

the number of PCs beyond the 95% variance level, requiring 3-8 more PCs before the change in mean MSE became insignificant.

- For datasets containing microstructures with multiple variations, i.e. change in particle shape and change in particle size distribution, the number of PCs to achieve sufficient precision in prediction was not as evident.

- All the ML models produced highly accurate regressions of the E_{22}, ν_{23}, and σ_{mat} mechanical properties.

- Comparing the PCs of each microstructural group studied, visual patterns emerge that can be used as references when determining the most importance influences of microstructures on the outputs.

REFERENCES

[1]   D. A. López, W. H. Schreiner, S. R. de Sánchez and S. N. Simison, "The influence of carbon steel microstructure on corrosion layers: An XPS and SEM characterization," *Applied Surface Science,* vol. 207, p. 69–85, 2003.

[2]   M. Li, J. Füssl, M. Lukacevic, J. Eberhardsteiner and C. M. Martin, "Strength predictions of clear wood at multiple scales using numerical limit analysis approaches," *Computers and Structures,* vol. 196, pp. 200-216, 2018.

[3]   Z. B. Haber, I. De la Varga, B. A. Graybea, B. Nakashoji and R. El-Helou, "FHWA-HRT-18-036 Properties and behavior of UHPC-class materials," U.S. Department of Transportation Federal Highway Administration, McLean, 2018.

[4]   B. Mobasher, A. Arora, M. Aguayo, F. Kianmofrad, Y. Yao and N. Neithalath, "FHWA-AZ-19-745 Developing Ultra High-Performance Concrete Mix Designs for Arizona Bridge Element Connections," Arizona Department of Transportation, Phoenix, 2019.

[5]   A. P. Lampropoulos, S. A. Paschalis, O. T. Tsioulou and S. E. Dritsos, "Strengthening of reinforced concrete beams using ultra high performance fibre reinforced concrete (UHPFRC)," *Engineering Structures,* vol. 106, pp. 370-384, 2016.

[6]   A. Arora, A. Almujaddidi, F. Kianmofrad, B. Mobasher and N. Neithalath, "Material design of economical ultra-high performance concrete (UHPC) and evaluation of their properties," *Cement and Concrete Composites,* vol. 104, p. 103346, 2019.

[7]   M. Alkaysi and S. El-Tawil, "Effects of variations in the mix constituents of ultra high performance concrete (UHPC) on cost and performance," *Materials and Structures,* vol. 49, pp. 4185-4200, 2016.

[8]   F. E. Bock, R. C. Aydin, C. J. Cyron, N. Huber, S. R. Kalidindi and B. Klusemann, "A Review of the Application of Machine Learning and Data Mining Approaches in Continuum Materials Mechanics," *Frontiers in Materials,* vol. 6, 2019.

[9]   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.

[10] M. A. DeRousseau, E. Laftchiev, J. R. Kasprzyk, B. Rajagopalan and W. V. Srubar III, "A comparison of machine learning methods for predicting the compressive strength of field-placed concrete," *Construction and Building Materials,* vol. 228, p. 116661, 2019.

[11] J.-S. Chou, C.-K. Chiu, M. Farfoura and I. Al-Taharwa, "Optimizing the Prediction Accuracy of Concrete Compressive Strength Based on a Comparison of Data-Mining Techniques," *Journal of Computing in Civil Engineering,* vol. 25, no. 3, pp. 242-253, 2011.

[12] D. V. Dao, H. Adeli, H.-B. Ly, L. M. Le, V. M. Le, T.-T. Le and B. T. Pham, "A Sensitivity and Robustness Analysis of GPR and ANN for High-Performance Concrete Compressive Strength Prediction Using a Monte Carlo Simulation," *Sustainability,* vol. 12, no. 830, pp. 1-22, 2020.

[13] K. T. Nguyen, Q. D. Nguyen, T. A. Le, J. Shin and K. Lee, "Analyzing the compressive strength of green fly ash based geopolymer concrete using experiment and machine learning approaches," *Construction and Building Materials,* vol. 247, p. 118581, 2020.

[14] J. Zhang, Y. Huang, Y. Wang and G. Ma, "Multi-objective optimization of concrete mixture proportions using machine learning and metaheuristic algorithms," *Construction and Building Materials,* vol. 253, p. 119208, 2020.

[15] E. P. Koumoulos, K. Paraskevoudis and C. A. Charitidis, "Constituents Phase Reconstruction through Applied Machine Learning in Nanoindentation Mapping Data of Mortar Surface," *Journal of Composites Science,* vol. 3, no. 63, 2019.

[16] G. Konstantopoulos, E. P. Koumoulos and C. A. Charitidis, "Testing Novel Portland Cement Formulations with Carbon Nanotubes and Intrinsic Properties Revelation: Nanoindentation Analysis with Machine Learning on Microstructure Identification," *nanomaterials,* vol. 10, p. 645, 2020.

[17] S. Abbas, M. L. Nehdi and M. A. Saleem, "Ultra-High Performance Concrete: Mechanical Performance, Durability, Sustainability and Implementation Challenges," *International Journal of Concrete Structures and Materials,* vol. 10, no. 3, pp. 271-295, 2016.

[18] A. Arora, M. Aguayo, H. Hansen, C. Castro, E. Federspiel, B. Mobasher and N. Neithalath, "Microstructural packing- and rheology-based binder selection and characterization for Ultra-high Performance Concrete (UHPC)," *Cement and Concrete Research,* vol. 103, pp. 179-190, 2018.

[19] R. Yu, P. Spiesz and H. Brouwers, "Effect of nano-silica on the hydration and microstructure development of Ultra-High Performance Concrete (UHPC) with a

low binder amount," *Construction and Building Materials,* vol. 65, pp. 140-150, 2014.

[20] M. Alkaysi, S. El-Tawil, Z. Liu and W. Hansen, "Effects of silica powder and cement type on durability of ultra high performance concrete (UHPC)," *Cement and Concrete Composites,* vol. 66, pp. 47-56, 2016.

[21] S. Zhao and W. Sun, "Nano-mechanical behavior of a green ultra-high performance concrete," *Construction and Building Materials,* vol. 63, pp. 150-160, 2014.

[22] S. Pyo, M. Tafesse, B.-J. Kim and H.-K. Kim, "Effects of quartz-based mine tailings on characteristics and leaching behavior of ultra-high performance concrete," *Construction and Building Materials,* vol. 166, pp. 110-117, 2018.

[23] M. C. Juenger, R. Snellings and S. A. Bernal, "Supplementary cementitious materials: New sources, characterization, and performance insights," *Cement and Concrete Research,* vol. 122, pp. 257-273, 2019.

[24] *ASTM C618-19 Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete,* West Conshohocken: ASTM, 2019, pp. 1-5.

[25] F. Pelisser, P. J. P. Gleize and A. Mikowski, "Effect of the Ca/Si Molar Ratio on the Micro/nanomechanical Properties of Synthetic CSH Measured by Nanoindentation," *The Journal of Physical Chemisty,* vol. 116, pp. 17219-17227, 2012.

[26] I. G. Richardson, "Tobermorite/jennite- and tobermorite/calcium hydroxide-based models for the structure of C-S-H: applicability to hardened pastes of tricalcium silicate, h-dicalcium silicate, Portland cement, and blends of Portland cement with blast-furnace slag, metakaol," *Cement and Concrete Research,* vol. 34, pp. 1733-1777, 2004.

[27] B. Lothenbach, K. Scrivener and R. D. Hooton, "Supplementary cementitious materials," *Cement and Concrete Research,* vol. 41, pp. 1244-1256, 2011.

[28] M. A. Qomi, K. Krakowiak, M. Bauchy, K. Stewart, R. Shahsavari, D. Jagannathan, D. Brommer, A. Baronnet, M. Buehler, S. Yip, F.-J. Ulm, K. Van Vliet and R.-.. Pellenq, "Combinatorial molecular optimization of cement hydrates," *Nature Communications,* pp. 1-10, 2014.

[29] D. Hou, H. Li, L. Zhang and J. Zhang, "Nano-scale mechanical properties investigation of C-S-H from hydrated tri-calcium silicate by nano-indentation and

molecular dynamics simulation," *Construction and Building Materials,* vol. 189, pp. 265-275, 2018.

[30] J. Rossen and K. Scrivener, "Optimization of SEM-EDS to determine the C–A–S–H composition in matured cement paste samples," *Materials Characterization,* vol. 123, pp. 294-306, 2017.

[31] J. E. Rossen, "Composition and morphology of C-A-S-H in pastes of alite and cement blended with supplementary cementitious materials," École Polytechnique Fédérale de Lausanne, Suisse, 2014.

[32] Y. K. Cho, S. H. Jung and Y. C. Choi, "Effects of chemical composition of fly ash on compressive strength of fly ash cement mortar," *Construction and Building Materials,* vol. 204, pp. 255-264, 2019.

[33] K. Vance, M. Aguayo, T. Oey, G. Sant and N. Neithalath, "Hydration and strength development in ternary portland cement blends containing limestone and fly ash or metakaolin," *Cement and Concrete Composites,* pp. 93-103, 2013.

[34] A. Arora, G. Sant and N. Neithalath, "Ternary blends containing slag and interground/blended limestone: Hydration, strength, and pore structure," *Construction and Building Materials,* vol. 102, pp. 113-124, 2016.

[35] G. Puerta-Falla, M. Balonis, G. Le Saout, G. Falzone, C. Zhang, N. Neithalath and G. Sant, "Elucidating the Role of the Aluminous Source on Limestone Reactivity in Cementitious Materials," *The American Ceramic Society,* vol. 98, no. 21, pp. 4076-4089, 2015.

[36] W. Huang, H. Kazemi-Kamyab, W. Sun and K. Scrivener, "Effect of cement substitution by limestone on the hydration and microstructural development of ultra-high performance concrete (UHPC)," *Cement and Concrete Composites,* vol. 77, pp. 86-101, 2017.

[37] E. B. Costa, F. A. Cardoso and V. M. John, "Influence of high contents of limestone fines on rheological behaviour and bond strength of cement-based mortars," *Construction and Building Materials,* vol. 156, pp. 1114-1126, 2017.

[38] J. Moon, S. Yoon, R. M. Wentzcovitch and P. J. Monteiro, "First-principles elasticity of monocarboaluminate hydrates," *American Mineralogist,* vol. 99, pp. 1360-1368, 2014.

[39] C.-J. Haecker, E. J. Garboczi, J. W. Bullard, R. B. Bohn, Z. Sun, S. P. Shah and T. Voigt, "Modeling the linear elastic properties of Portland cement paste," *Cement and Concrete Research,* vol. 35, pp. 1948-1960, 2005.

[40] D. Wang, C. Shi, N. Farzadnia, Z. Shi and H. Jia, "A review on effects of limestone powder on the properties of concrete," *Construction and Building Materials,* vol. 192, pp. 153-166, 2018.

[41] A. Arora, Y. Yao, B. Mobasher and N. Neithalath, "Fundamental insights into the compressive and flexural response of binder and aggregate-optimized ultra-high performance concrete (UHPC)," *Cement and Concrete Composites,* vol. 98, pp. 1-13, 2019.

[42] E. W. Washburn, "Note on a method of determining the distribution of pore sizes in a porous material," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 7, no. 4, pp. 115-116, 1921.

[43] H. Alghamdi, A. Dakhane, A. Alum, M. Abbaszadegan, B. Mobasher and N. Neithalath, "Synthesis and characterization of economical, multi-functional porous ceramics based on abundant aluminosilicates," *Materials and Design,* vol. 152, pp. 10-21, 2018.

[44] A. A. Liabastre and C. Orr, "An evaluation of pore structure by mercury penetration," *Journal of Colloid and Interface Science,* vol. 64, no. 1, pp. 1-18, 1978.

[45] D. Shi and D. N. Winslow, "Contact angle and damage during mercury intrusion into cement paste," *Cement and Concrete Research,* vol. 15, pp. 645-654, 1985.

[46] M. S. Kim, Y. Jun, C. Lee and J. E. Oh, "Use of CaO as an activator for producing a price-competitive non-cement structural binder using ground granulated blast furnace slag," *Cement and Concrete Research,* vol. 54, pp. 208-214, 2013.

[47] P. E. Stutzman, P. Feng and J. W. Bullard, "Phase Analysis of Portland Cement by Combined Quantitative X-Ray Powder Diffraction and Scanning Electron Microscopy," *Journal of Research of the National Institute of Standards and Technology,* vol. 121, pp. 47-107, 2016.

[48] V. Lilkov, O. Petrov, Y. Tzvetanova and P. Savov, "Mössbauer, DTA and XRD study of Portland cement blended with fly ash and silica fume," *Construction and Building Materials,* vol. 29, pp. 33-41, 2012.

[49] G. V. P. B. Singh and K. V. L. Subramaniam, "Quantitative XRD Analysis of Binary Blends of Siliceous Fly Ash and Hydrated Cement," *Journal of Materials in Civil Engineering,* vol. 28, no. 8, 2016.

[50] L. Sorelli, G. Constantinides, F.-J. Ulm and F. Toutlemonde, "The nano-mechanical signature of Ultra High Performance Concrete by statistical

nanoindentation techniques," *Cement and Concrete Research,* no. 38, pp. 1447-1456, 2008.

[51] C. Hu and Z. Li, "A review on the mechanical properties of cement-based materials measured by nanoindentation," *Construction and Building Materials,* no. 90, pp. 80-90, 2015.

[52] G. Constantinides and F.-J. Ulm, "The effect of two types of C-S-H on the elasticity of cement-based materials: Results from nanoindentation and micromechanical modeling," *Cement and Concrete Research,* no. 34, pp. 67-80, 2004.

[53] P. Mondal, S. P. Shah and L. Marks, "A reliable technique to determine the local mechanical properties at the nanoscale for cementitious materials," *Cement and Concrete Research,* no. 37, pp. 1440-1444, 2007.

[54] H. M. Jennings, J. J. Thomas, J. S. Gevrenov, G. Constantinides and F.-J. Ulm, "A multi-technique investigation of the nanoporosity of cement paste," *Cement and Concrete Research,* no. 37, pp. 329-336, 2007.

[55] C. Hu, "Nanoindentation as a tool to measure and map mechanical properties of hardened cement pastes," *MRS Communications,* vol. 5, pp. 83-87, 2015.

[56] M. J. DeJong and F.-J. Ulm, "The nanogranular behavior of C-S-H at elevated temperatures (up to 700 °C)," *Cement and Concrete Research,* no. 37, pp. 1-12, 2007.

[57] M. Vandamme, F.-J. Ulm and P. Fonollosa, "Nanogranular packing of C–S–H at substochiometric conditions," *Cement and Concrete Research,* vol. 40, pp. 14-26, 2010.

[58] C. Hu, "Microstructure and mechanical properties of fly ash blended cement pastes," *Construction and Building Materials,* no. 73, pp. 618-625, 2014.

[59] C. Hu and Z. Li, "Property investigation of individual phases in cementitious composites containing silica fume and fly ash," *Cement & Concrete Composites,* no. 57, pp. 17-26, 2015.

[60] V. Z. Zadeh and C. P. Bobko, "Nanoscale mechanical properties of concrete containing blast furnace slag and fly ash before and after thermal damage," *Cement & Concrete Composites,* no. 37, pp. 215-221, 2012.

[61] C. Hu, Z. Li, Y. Gao, Y. Han and Y. Zhang, "Investigation on microstructures of cementitious composites incorporating slag," *Advances in Cement Research,* vol. 26, no. 4, pp. 222-232, 2014.

[62] M. Vandamme and F.-J. Ulm, "Nanoindentation investigation of creep properties of calcium silicate hydrates," *Cement and Concrete Research,* no. 52, pp. 38-52, 2013.

[63] M. Vandamme, "The Nanogranular Origin of Concrete Creep: A Nanoindentation Investigation of Microstructure and Fundamental Properties of Calcium-Silicate-Hydrates," Massachusetts Institute of Technology, 2008.

[64] K. Velez, S. Maximilien, D. Damidot, G. Fantozzi and F. Sorrentino, "Determination by nanoindentation of elastic modulus and hardness of pure constituents of Portland cement clinker," *Cement and Concrete Research,* vol. 31, pp. 555-561, 2000.

[65] W. Oliver and G. Pharr, "An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments," *Journal of Materials Research,* vol. 7, no. 6, 1992.

[66] W. Oliver and G. Pharr, "Measurement of hardness and elastic modulus by instrumented indentation: Advances in understanding andrefinements to methodology," *Journal of Materials Research,* vol. 19, no. 1, pp. 1-20, 2004.

[67] P. Stutzman, "Scanning electron microscopy imaging of hydraulic cement microstructure," *Cement and Concrete Composites,* vol. 26, pp. 957-966, 2004.

[68] P. E. Stutzman, J. W. Bullard and P. Feng, "Quantitative Imaging of Clinker and Cement Microstructure," *National Institute of Standards and Technology Technical Note,* vol. 1877, 2015.

[69] J. J. Chen, L. Sorelli, M. Vandamme, F.-J. Ulm and G. Chanvillard, "A Coupled Nanoindentation/SEM-EDS Study on Low Water/Cement Ratio Portland Cement Paste: Evidence for C–S–H/Ca(OH)2 Nanocomposites," *Journal of the American Ceramic Society,* vol. 93, no. 5, 2010.

[70] W. Wilson, J. M. Rivera-Torres, L. Sorelli, A. Durán-Herrera and A. Tagnit-Hamou, "The micromechanical signature of high-volume natural pozzolan concrete by combined statistical nanoindentation and SEM-EDS analyses," *Cement and Concrete Research,* vol. 91, pp. 1-12, 2017.

[71] W. Wilson, L. Sorelli and A. Tagnit-Hamou, "Unveiling micro-chemo-mechanical properties of C–(A)–S–H and other phases in blended-cement pastes," *Cement and Concrete Research,* vol. 107, pp. 317-336, 2018.

[72] W. Wilson, L. Sorelli and A. Tagnit-Hamou, "Automated coupling of NanoIndentation and Quantitative Energy-Dispersive Spectroscopy (NI-QEDS): A comprehensive method to disclose the micro-chemo-mechanical properties of cement pastes," *Cement and Concrete Research,* vol. 103, pp. 49-65, 2018.

[73] P. E. Stutzman, "Microscopy of Clinker and Hydraulic Cements," *Reviews in Mineralogy & Geochemistry,* vol. 74, pp. 101-146, 2012.

[74] C. G. Hoover and F.-J. Ulm, "Experimental chemo-mechanics of early-age fracture properties of cement paste," *Cement and Concrete Research,* vol. 75, pp. 42-52, 2015.

[75] H. J. Butt, R. Berger, E. Bonaccurso, Y. Chen and J. Wang, "Impact of atomic force microscopy on interface and colloid science," *Advances in Colloid and Interface Science,* vol. 133, pp. 91-104, 2007.

[76] M. Miller, C. Bobko, M. Vandamme and F.-J. Ulm, "Surface roughness criteria for cement paste nanoindentation," *Cement and Concrete Research,* vol. 38, pp. 467-476, 2008.

[77] V. Z. Zadeh and C. P. Bobko, "Nanoscale mechanical properties of concrete containing blast furnace slag and fly ash before and after thermal damage," *Cement and Concrete Composites,* vol. 37, pp. 215-221, 2013.

[78] Z. Luo, W. Li, K. Wang and S. P. Shah, "Research progress in advanced nanomechanical characterization of cement-based materials," *Cement and Concrete Composites,* vol. 94, pp. 277-295, 2018.

[79] J. E. Sader, J. W. M. Chon and P. Mulvaney, "Calibration of rectangular atomic force microscope cantilevers," *Review of Scientific Instruments,* vol. 70, no. 10, pp. 3967-3969, 1999.

[80] H. Yamada, C. Liu, S. Wu, Y. Koyama, S. Ju, J. Shiomi, J. Morikawa and R. Yoshida, "Predicting Materials Properties with Little Data Using Shotgun Transfer Learning," *American Chemical Society Central Science,* vol. 5, no. 10, pp. 1717-1730, 2019.

[81] P. Gardner, X. Liu and K. Worden, "On the application of domain adaptation in structural health monitoring," *Mechanical Systems and Signal Processing,* vol. 138, p. 106550, 2020.

[82] 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,* 2016.

[83] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks,* vol. 61, pp. 85-117, 2015.

[84] P. Acar, "Machine Learning Approach for Identification of Microstructure–Process Linkages," *American Institute of Aeronautics and Astronautics Journal,* vol. 57, no. 8, pp. 3608-3614, 2019.

[85] S. Ye, B. Li, Q. Li, H.-P. Zhao and X.-Q. Feng, "Deep neural network method for predicting the mechanical properties of composites," *Applied Physics Letters,* vol. 115, p. 161901, 2019.

[86] Z. Yang, Y. C. Yabansu, R. Al-Bahrani, W.-k. Liao, A. N. Choudhary, S. R. Kalidindi and A. Agrawal, "Deep learning approaches for mining structure-property linkages in high contrast composites from simulation datasets," *Computational Materials Science,* vol. 151, pp. 278-287, 2018.

[87] T. Oey, S. Jones and G. Sant, "Machine learning can predict setting behavior and strength evolution of hydrating cement systems," *Journal of the American Ceramic Society,* vol. 103, no. 1, pp. 480-490, 2020.

[88] S. Mahdinia, H. Eskandari-Naddaf and R. Shadnia, "Effect of cement strength class on the prediction of compressive strength of cement mortar using GEP method," *Construction and Building Materials,* vol. 198, pp. 27-41, 2019.

[89] B. R. Prasad, H. Eskandari and B. V. Reddy, "Prediction of compressive strength of SCC and HPC with high volume fly ash using ANN," *Construction and Building Materials,* vol. 23, pp. 117-128, 2009.

[90] W. B. Chaabene, M. Flah and M. L. Nehdi, "Machine learning prediction of mechanical properties of concrete: Critical review," *Construction and Building Materials,* vol. 260, p. 119889, 2020.

[91] Y. Pu, D. B. Apel, J. Chen and C. Wei, "A Gaussian process machine learning model for cemented rockfill strength prediction at a diamond mine," *Neural Computing and Applications,* 2019.

[92] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, "Scikit-learn: Machine

Learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825-2830, 2011.

[93]   T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[94]   G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Systems With Applications,* vol. 73, pp. 220-239, 2017.

[95]   X. Li, Z. Liu, S. Cui, C. Luo, C. Li and Z. Zhuang, "Predicting the effective mechanical property of heterogeneous materials by image based modeling and deep learning," *Computer Methods Applied Mechanics and Engineering Methods,* vol. 347, pp. 735-753, 2019.

[96]   T. Tieleman and G. Hinton, "Lecture 6.5 RmsProp: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning,* 2012.

[97]   N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research,* vol. 15, pp. 1929-1958, 2014.

[98]   M. M. Norhasri, M. Hamidah, A. M. Fadzil and O. Megawati, "Inclusion of nano metakaolin as additive in ultra high performance concrete (UHPC)," *Construction and Building Materials,* vol. 127, pp. 167-175, 2016.

[99]   W. Meng and K. H. Khayat, "Mechanical properties of ultra-high-performance concrete enhanced with graphite nanoplatelets and carbon nanofibers," *Composites Part B,* vol. 107, pp. 113-122, 2016.

[100] C. Wang, C. Yang, F. Liu, C. Wan and X. Pu, "Preparation of Ultra-High Performance Concrete with common technology and materials," *Cement and Concrete Composites,* vol. 34, pp. 538-544, 2012.

[101] D.-Y. Yoo, M. J. Kim, S.-W. Kim and J.-J. Park, "Development of cost effective ultra-high-performance fiber-reinforced concrete using single and hybrid steel fibers," *Construction and Building Materials,* vol. 150, pp. 383-394, 2017.

[102] P. Li, Q. Yu and H. Brouwers, "Effect of coarse basalt aggregates on the properties of Ultra-high Performance Concrete (UHPC)," *Construction and Building Materials,* vol. 170, pp. 649-659, 2018.

[103] L. Xu, F. Wu, Y. Chi, P. Cheng, Y. Zeng and Q. Chen, "Effects of coarse aggregate and steel fibre contents on mechanical properties of high performance concrete," *Construction and Building Materials,* vol. 206, pp. 97-110, 2019.

[104] F.-J. Ulm, M. Vandamme, C. Bobko, J. A. Ortega, K. Tai and C. Ortiz, "Statistical Indentation Techniques for Hydrated Nanocomposites: Concrete, Bone, and Shale," *The American Ceramic Society,* vol. 90, no. 9, pp. 2677-2692, 2007.

[105] K. J. Krakowiak, W. Wilson, S. James and F.-J. Ulm, "In situ Chemo Mechanical Characterization of Cementitious Microstructures with Coupled X Ray Microanalysis and Indentation Technique," in *Mechanics and Physics of Creep, Shrinkage, and Durability of Concrete: A Tribute to Zdenek P. Bažant: Proceedings of the Ninth International Conference on Creep, Shrinkage, and Durability Mechanics*, Cambridge, 2013.

[106] J. Fu, "Nano-indentation experiment for determining mechanical properties of typical cement phases at nano/micro-scale," *IOP Conference Series: Materials Science and Engineering,* vol. 439, 2018.

[107] F.-J. Ulm, M. Vandamme, H. M. Jennings, J. Vanzo, M. Bentivegna, K. J. Krakowiak, G. Constantinides, C. P. Bobko and K. J. Van Vliet, "Does microstructure matter for statistical nanoindentation techniques?," *Cement and Concrete Composites,* vol. 32, pp. 92-99, 2010.

[108] P. Lura, P. Trtik and B. Münch, "Validity of recent approaches for statistical nanoindentation of cement pastes," *Cement & Concrete Composites,* vol. 33, no. 4, pp. 457-465, 2011.

[109] K. J. Krakowiak, W. Wilson, S. James, S. Musso and F.-J. Ulm, "Inference of the phase-to-mechanical property link via coupled X-ray spectrometry and indentation analysis: Application to cement-based materials," *Cement and Concrete Research,* vol. 67, pp. 271-285, 2015.

[110] G. Constantinides, K. R. Chandran, F.-J. Ulm and K. V. Vliet, "Grid indentation analysis of composite microstructure and mechanics: Principles and validation," *Materials Science and Engineering A,* vol. 430, pp. 189-202, 2006.

[111] J. Němeček, V. Šmilauer and L. Kopecký, "Nanoindentation characteristics of alkali-activated aluminosilicate materials," *Cement and Concrete Composites,* vol. 33, pp. 163-170, 2011.

[112] G. Constantinides and F.-J. Ulm, "The nanogranular nature of C–S–H," *Journal of the Mechanics and Physics of Solids,* vol. 55, pp. 64-90, 2007.

[113] P. J. Monteiro and C. T. Chang, "The Elastic Moduli of Calcium Hydroxide," *Cement and Concrete Research,* vol. 25, no. 8, pp. 1605-1609, 1995.

[114] F. Lavergne, A. B. Fraj, I. Bayane and J. F. Barthélémy, "Estimating the mechanical properties of hydrating blended cementitious materials: An investigation based on micromechanics," *Cement and Concrete Research,* vol. 104, pp. 37-60, 2018.

[115] N. K. Dhami, A. Mukherjee and M. S. Reddy, "Micrographical, minerological and nano-mechanical characterisation of microbial carbonates from urease and carbonic anhydrase producing bacteria," *Ecological Engineering,* vol. 94, pp. 443-454, 2016.

[116] Z. Yu, A. Zhou and D. Lau, "Mesoscopic packing of disk-like building blocks in calcium silicate hydrate," *Scientific Reports,* pp. 1-8, 2016.

[117] V. L. Bonavetti, V. F. Rahhal and E. F. Isassar, "Studies on the carboaluminate formation in limestone filler-blended cements," *Cement and Concrete Research,* vol. 31, pp. 853-859, 2001.

[118] G. S. Barger, J. Bayles, B. Blair, D. Brown, H. Chen, T. Conway, P. Hawkins, R. A. Helinski, F. A. Innis, M. D. Luther, W. C. McCall, D. Moore, W. O'Brien, E. R. Orsini, M. F. Pistilli, D. Suchorski and O. Tavares, "Ettringite Formation and the Performance of Concrete," *Portland Cement Association,* pp. 1-16, 201.

[119] A. Ipavec, R. Gabrovsek, T. Vuk, V. Kaucic, J. Macek and A. Meden, "Carboaluminate Phases Formation During the Hydration of Calcite-Containing Portland Cement," *Journal of the American Ceramic Society,* vol. 94, no. 4, pp. 1238-1242, 2011.

[120] D. Hou, T. Zhao, H. Ma and Z. Li, "Reactive Molecular Simulation on Water Confined in the Nanopores of the Calcium Silicate Hydrate Gel: Structure, Reactivity, and Mechanical Properties," *The Journal of Physical Chemistry C,* vol. 119, pp. 1346-1358, 2015.

[121] P. Mondal, S. P. Shah, L. D. Marks and J. J. Gaitero, "Comparative Study of the Effects of Microsilica and Nanosilica in Concrete," *Transportation Research Record,* vol. 2141, pp. 6-9, 2010.

[122] W. da Silva, J. Němeček and P. Štemberk, "Application of multiscale elastic homogenization based on nanoindentation for high performance concrete," *Advances in Engineering Software,* Vols. 62-63, pp. 109-118, 2013.

[123] Y. Li, P. Wang and Z. Wang, "Evaluation of elastic modulus of cement paste corroded in brine solution with advanced homogenization method," *Construction and Building Materials,* vol. 157, pp. 660-609, 2017.

[124] X. Gao, Y. Wei and W. Huang, "Effect of individual phases on multiscale modeling mechanical properties of hardened cement paste," *Construction and Building Materials,* vol. 153, pp. 25-35, 2017.

[125] S. Das, P. Yang, S. S. Singh, J. C. Mertens, X. Xiao, N. Chawla and N. Neithalath, "Effective properties of a fly ash geopolymer: Synergistic application of X-ray synchrotron tomography, nanoindentation, and homogenization models," *Cement and Concrete Research,* vol. 78, pp. 252-262, 2015.

[126] L. Brown, P. G. Allison and F. Sanchez, "Use of nanoindentation phase characterization and homogenization to estimate the elastic modulus of heterogeneously decalcified cement pastes," *Materials and Design,* vol. 142, pp. 308-318, 2018.

[127] W. Ashraf, J. Olek and N. Tian, "Multiscale characterization of carbonated wollastonite paste and application of homogenization schemes to predict its effective elastic modulus," *Cement and Concrete Composites,* vol. 72, pp. 284-298, 2016.

[128] D. P. Bentz, J. Arnold, M. J. Boisclair, S. Z. Jones, P. Rothfeld and P. E. Stutzman, "NIST Technical Note 1963: Influence of Aggregate Characteristics on Concrete Performance," National Institute of Standards and Technology, 2017.

[129] M. Hori and S. Nemat-Nasser, "Double-inclusion model and overall moduli of multi-phase composites," *Mechanics of Materials,* vol. 14, pp. 189-206, 1993.

[130] S. Nemat-Nasser and M. Hori, "Section 10 Self-Consistent, Differential, and Related Average Methods," in *Micromechanics Overall Properties of Heterogeneous Materials*, North-Holland Series in Applied Mathematics and Mechanics, 1993, pp. 325-366.

[131] C. C. Yang and R. Huang, "Double inclusion model for approximate elastic moduli of concrete material," *Cement and Concrete Research,,* vol. 26, no. 1, pp. 83-91, 1996.

[132] P. Trtik, B. Münch and P. Lura, "A critical examination of statistical nanoindentation on model materials and hardened cement pastes based on virtual experiments," *Cement and Concrete Composites,* vol. 31, no. 10, p. 705–714, 2009.

220

[133] Z. Wu, K. H. Khayat and C. Shi, "Changes in rheology and mechanical properties of ultra-high performance concrete with silica fume content," *Cement and Concrete Research,* vol. 123, p. 105786, 2019.

[134] N. Hernandez, J. Lizarazo-Marriaga and M. A. Rivas, "Petrographic characterization of Portlandite crystal sizes in cement pastes affected by different hydration environments," *Construction and Building Materials,* vol. 182, pp. 541-549, 2018.

[135] H. Jennings, "The Colloid/Nanogranular Nature of Cement Paste and Properties," *Nanotechnology in Construction,* vol. 3, pp. 27-36, 2009.

[136] W. Shen, W. Zhang, J. Wang, C. Han, B. Zhang, J. Li and G. Xu, "The microstructure formation of C-S-H in the HPC paste from nano-scale feature," *Journal of Sustainable Cement-Based Materials,* vol. 0, no. 0, pp. 1-15, 2019.

[137] E. Masoero, E. Del Gado, R. J.-M. Pellenq, S. Yip and F.-J. Ulm, "Nano-scale mechanics of colloidal C–S–H gels," *Royal Society of Chemisty,* vol. 10, pp. 491-499, 2014.

[138] B. Gathier, *Multiscale strength homogenization: application to shale nanoindentation,* Cambridge, MA: Massachuesetts Institute of Technology, 2008.

[139] J. A. Ortega, *Microporomechanical Modeling of Shale,* Cambridge, MA: Massachusetts Institute of Technology, 2010.

[140] G. Habert, E. Denarié, A. Šajna and P. Rossi, "Lowering the global warming impact of bridge rehabilitations by using Ultra High Performance Fibre Reinforced Concretes," *Cement and Concrete Composites,* vol. 38, pp. 1-11, 2013.

[141] S. Pyo, M. Tafesse, H. Kim and H.-K. Kim, "Effect of chloride content on mechanical properties of ultra high performance concrete," *Cement and Concrete Composites,* vol. 84, pp. 175-187, 2017.

[142] C. Hu, Y. Gao, Y. Zhang and Z. Li, "Statistical nanoindentation technique in application to hardened cement pastes: Influences of material microstructure and analysis method," *Construction and Building Materials,* vol. 113, pp. 306-316, 2016.

[143] W. da Silva, J. Němeček and P. Štemberk, "Methodology for nanoindentation-assisted prediction of macroscale elastic properties of high performance cementitious composites," *Cement & Concrete Composites,* vol. 45, pp. 57-68, 2014.

[144] M. Vandamme and F.-J. Ulm, "Viscoelastic solutions for conical indentation," *International Journal of Solids and Structures,* vol. 43, pp. 3142-3165, 2006.

[145] L. P. Singh, S. R. Karade, S. K. Bhattacharyya, M. M. Yousuf and S. Ahalawat, "Beneficial role of nanosilica in cement based materials – A review," *Construction and Building Materials,* vol. 47, pp. 1069-1077, 2013.

[146] E. L. Ford, A. Arora, B. Mobasher, C. G. Hoover and N. Neithalath, "Elucidating the nano-mechanical behavior of multi-component binders for ultra-high performance concrete," *Construction and Building Materials,* vol. 243, p. 118214, 2020.

[147] E. Kapeluszna, Ł. Kotwica, A. Rózycka and Ł. Gołek, "Incorporation of Al in C-A-S-H gels with various Ca/Si and Al/Si ratio: Microstructural and structural characteristics with DTA/TG, XRD, FTIR and TEM analysis," *Construction and Building Materials,* vol. 155, pp. 643-653, 2017.

[148] M. J. A. Qomi, F.-J. Ulm and R. J.-M. Pellenq, "Evidence on the Dual Nature of Aluminum in the Calcium-Silicate-Hydrates Based on Atomistic Simulations," *Journal of the American Ceramic Society,* vol. 95, no. 3, pp. 1128-1137, 2012.

[149] G. Geng, R. J. Myers, J. Li, R. Maboudian, C. Carraro, D. A. Shapiro and P. J. M. Monteiro, "Aluminum-induced dreierketten chain cross-links increase the mechanical properties of nanocrystalline calcium aluminosilicate hydrate," *Scientific Reports,* vol. 7, 2017.

[150] Y. Suda, T. Saeki and T. Saito, "Relation between Chemical Composition and Physical Properties of C-S-H Generated from Cementitious Materials," *Journal of Advanced Concrete Technology,* vol. 13, pp. 275-290, 2015.

[151] H. Manzano, J. Dolado and A. Ayuela, "Elastic properties of the main species present in Portland cement pastes," *Acta Materialia,* vol. 57, pp. 1666-1674, 2009.

[152] P. T. Durdziński, C. F. Dunant, M. B. Haha and K. L. Scrivener, "A new quantification method based on SEM-EDS to assess fly ash composition and study the reaction of its individual components in hydrating cement paste," *Cement and Concrete Research,* vol. 73, pp. 111-122, 2015.

[153] C. Rößler, F. Steiniger and H.-M. Ludwig, "Characterization of C–S–H and C–A–S–H phases by electron microscopy imaging, diffraction, and energy dispersive X-ray spectroscopy," *Journal of the American Ceramic Society,* vol. 100, pp. 1733-1742, 2017.

[154] E. L. Ford, C. G. Hoover, B. Mobasher and N. Neithalath, "Relating the nano-mechanical response and qualitative chemical maps of multi-component ultra-high performance cementitious binders," *Construction and Building Materials,* vol. 260, p. 119959, 2020.

[155] E. L. Ford, S. Kailas, K. Maneparambil and N. Neithalath, "Machine learning approaches to predict the micromechanical properties of cementitious hydration phases from microstructural chemical maps," *Construction and Building Materials,* vol. 265, p. 120647, 2020.

[156] P. E. Stutzman and J. R. Clifton, "Specimen Preparation for Scanning Electron Microscopy," *Proceedings from the Twenty-First Internation Conference on Cement Microscopy,* pp. 10-22, 1999.

[157] G. Konstantopoulos, E. P. Koumoulos and C. A. Charitidis, "Classification of mechanism of reinforcement in the fiber-matrix interface: Application of Machine Learning on nanoindentation data," *Materials and Design,* vol. 192, p. 108705, 2020.

[158] F. Chollet, "Keras," GitHub repository, 2015. [Online]. Available: https://keras.io/api/.

[159] B. F. Huang and P. C. Boutros, "The parameter sensitivity of random forests," *BMC Bioinformatics,* vol. 17, p. 331, 2016.

[160] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics,* 2018.

[161] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters,* vol. 27, pp. 861-874, 2006.

[162] H. J. David and R. J. Till, "A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems," *Machine Learning,* vol. 45, pp. 171-186, 2001.

[163] C. Su, S. Ju, Y. Liu and Z. Yu, "Improving Random Forest and Rotation Forest for highly imbalanced datasets," *Intelligent Data Analysis,* vol. 19, pp. 1409-1432, 2015.

[164] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research,* vol. 13, pp. 281-305, 2012.

[165] Y. Wei, X. Gao and S. Liang, "A combined SPM/NI/EDS method to quantify properties of inner and outer C-S-H in OPC and slag-blended cement pastes," *Cement and Concrete Composites,* vol. 85, pp. 56-66, 2018.

[166] T. Han, A. Siddique, K. Khayat, J. Huang and A. Kumar, "An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete," *Construction and Building Materials,* vol. 244, p. 118271, 2020.

[167] S. Bishnoi, S. Singh, R. Ravinder, M. Bauchy, N. N. Gosvami, H. Kodamana and N. A. Krishnan, "Predicting Young's modulus of oxide glasses with sparse datasets using machine learning," *Journal of Non-Crystalline Solids,* vol. 524, p. 119643, 2019.

[168] M. S. Caywood, D. M. Roberts, J. B. Colombe, H. S. Greenwald and M. Z. Weiland, "Gaussian Process Regression for Predictive But Interpretable Machine Learning Models: An Example of Predicting Mental Workload across Tasks," *Frontiers in Human Neuroscience,* 2017.

[169] D. Duvenaud, H. Nickisch and C. E. Rasmussen, "Additive Gaussian Processes," *arXiv,* 2011.

[170] K. Lee, H. Cho and I. Lee, "Variable selection using Gaussian process regression-based metrics for high-dimensional model approximation with limited data," *Structural and Multidisciplinary Optimization,* vol. 59, pp. 1439-1454, 2019.

[171] H. Son, C. Kim and C. Kim, "Automated Color Model–Based Concrete Detection in Construction-Site Images by Using Machine Learning Algorithms," *Journal of Computing in Civil Engineering,* vol. 26, no. 3, pp. 421-433, 2012.

[172] D. Prayogo, "Metaheuristic-Based Machine Learning System for Prediction of Compressive Strength based on Concrete Mixture Properties and Early-Age Strength Test Results," *Civil Engineering Dimension,* vol. 20, no. 1, pp. 21-29, 2018.

[173] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning, Cambridge: The MIT Press, 2016.

[174] S. Swaminathan, S. Ghosh and N. J. Pagano, "Statistically Equivalent Representative Volume Elements for Unidirectional Composite Microstructures: Part I – Without Damage," *Journal of Composite Materials,* vol. 40, no. 7, pp. 583-604, 2006.

[175] H. Kumar, C. L. Briant and W. A. Curtin, "Using microstructure reconstruction to model mechanical behavior in complex microstructures," *Mechanics of Materials,* vol. 38, pp. 818-832, 2006.

[176] S. Das, A. Maroli, S. S. Singh, T. Stannard, X. Xiao, N. Chawla and N. Neithalath, "A microstructure-guided constitutive modeling approach for random heterogeneous materials: Application to structural binders," *Computational Materials Science,* vol. 119, pp. 52-64, 2016.

[177] M. M. Shahzamanian, T. Tadepalli, A. M. Rajendran, W. D. Hodo, R. Mohan, R. Valisetty, P. W. Chung and J. J. Ramsey, "Representative Volume Element Based Modeling of Cementitious Materials," *Journal of Engineering Materials and Technology,* vol. 136, p. 011007, 2014.

[178] L. Akand, M. Yang and Z. Gao, "Characterization of pervious concrete through image based micromechanical modeling," *Construction and Building Materials,* vol. 114, pp. 547-555, 2016.

[179] H. Zhang, B. Šavija, S. C. Figueiredo, M. Lukovic and E. Schlangen, "Microscale Testing and Modelling of Cement Paste as Basis for Multi-Scale Modelling," *Materials,* vol. 9, no. 11, p. 907, 2016.

[180] M. Luković, E. Schlangen and G. Ye, "Combined experimental and numerical study of fracture behaviour of cement paste at the microlevel," *Cement and Concrete Research,* vol. 73, pp. 123-135, 2015.

[181] M. V. Pathan, S. A. Ponnusami, J. Pathan, R. Pitisongsawat, B. Erice, N. Petrinic and V. L. Tagarielli, "Predictions of the mechanical properties of unidirectional fibre composites by supervised machine learning," *Nature Scientific Reports,* vol. 9, p. 13964, 2019.

[182] A. Choudhury, Y. C. Yabansu, S. R. Kalidindi and A. Dennstedt, "Quantification and classification of microstructures in ternary eutectic alloys using 2-point spatial correlations and principal component analyses," *Acta Materialia,* vol. 110, pp. 131-141, 2016.

[183] C. Rao and Y. Liu, "Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization," *Computational Materials Science,* vol. 184, p. 109850, 2020.

[184] M. Tschopp, "Synthetic Microstructure Generator," MATLAB Central File Exchange, 2020. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/25389-synthetic-microstructure-generator. [Accessed 24 9 2020].

[185] M. A. Tschopp, G. B. Wilks and J. E. Spowart, "Multi-scale characterization of orthotropic microstructures," *Modelling and Simulation in Materials Science and Engineering,* vol. 16, p. 065009, 2008.

[186] T.-S. Han, X. Zhang, J.-S. Kim, S.-Y. Chung, J.-H. Lim and C. Linder, "Area of lineal-path function for describing the pore microstructures of cement paste and their relations to the mechanical properties simulated from micro-CT microstructures," *Cement and Concrete Composites,* vol. 89, pp. 1-17, 2018.

[187] M. V. Karsanina, K. M. Gerke, E. B. Skvortsova and D. Mallants, "Universal Spatial Correlation Functions for Describing and Reconstructing Soil Microstructure," *PLOS ONE,* vol. 10, no. 5, p. e0126515, 2015.

[188] J. Lißner and F. Fritzen, "Data-Driven Microstructure Property Relations," *Mathematical and Computational Applications,* vol. 24, no. 2, p. 57, 2019.

[189] S. Torquato, Random Heterogeneous Materials, New York: Springer-Verlag, 2002.

[190] S. Torquato, "Statistical Description of Microstructures," *Annual Review of Materials Research,* vol. 32, pp. 77-111, 2002.

[191] S. Torquato, "Predicting transport characteristics of hyperuniform porous media via rigorous microstructure-property relations," *Advances in Water Resources,* vol. 140, p. 103565, 2020.

[192] D. T. Fullwood, S. R. Niezgoda and S. R. Kalidindi, "Microstructure reconstructions from 2-point statistics using phase-recovery algorithms," *Acta Materialia,* vol. 56, pp. 942-948, 2008.

[193] M. S. Sumanasooriya, D. P. Bentz and N. Neithalath, "Planar Image-Based Reconstruction of Pervious Concrete Pore Structure and Permeability Prediction," *ACI Materials Journal,* vol. 107, no. 4, pp. 413-421, 2010.

[194] M. G. Rozman and M. Utz, "Uniqueness of Reconstruction of Multiphase Morphologies from Two-Point Correlation Functions," *Physical Review Letters,* vol. 89, no. 13, p. 135501, 2002.

[195] A. Cecen, T. Fast and S. R. Kalidindi, "Versatile algorithms for the computation of 2-point spatial correlations in quantifying material structure," *Integrating Materials and Manufacturing,* vol. 5, pp. 1-15, 2016.

[196] H. L. Frisch and F. H. Stillinger, "Contribution to the Statistical Geometric Basis of Radiation Scattering," *The Journal of Chemical Physics,* vol. 38, p. 2200, 1963.

[197] D. . B. Brough, D. Wheeler and S. R. Kalidindi, "Materials Knowledge Systems in Python—a Data Science Framework for Accelerated Development of Hierarchical Materials," *Integrating Materials and Manufacturing Innovation,* vol. 6, pp. 36-53, 2017.

[198] A. U. Ozturk and O. Onal, "Identification of Water/Cement Ratio of Cement Pastes, Basing on the Microstructure Image Analysis Data and using Artificial Neural Network," *KSCE Journal of Civil Engineering,* vol. 17, no. 4, pp. 763-768, 2013.

[199] A. Chowdhury, E. Kautz, B. Yener and D. Lewis, "Image driven machine learning methods for microstructure recognition," *Computational Materials Science,* vol. 123, pp. 176-187, 2016.

[200] N. Khatavkar, S. Swetlana and A. K. Singh, "Accelerated prediction of Vickers hardness of Co- and Ni-based superalloys from microstructure and composition using advanced image processing techniques and machine learning," *Acta Materialia,* vol. 196, pp. 295-303, 2020.

[201] C. Kunselman, V. Attari, L. McClenny, U. Braga-Neto and R. Arroyave, "Semi-supervised learning approaches to class assignment in ambiguous microstructures," *Acta Materialia,* vol. 188, pp. 49-62, 2020.

[202] Z. Hashin, "NASA-CR-1974: Theory of Fiber Reinforced Materials," National Aeronautics and Space Administration, Washington, D.C., 1972.

[203] M. S. Mohammadi, M. Komeili, A. B. Phillion and A. S. Milani, "Toward better understanding of the effect of fiber distribution on effective elastic properties of unidirectional composite yarns," *Computers and Structures,* vol. 163, pp. 29-40, 2016.

[204] D. Chen, X. He, Q. Teng, Z. Xu and Z. Li, "Reconstruction of multiphase microstructure based on statistical descriptors," *Physica A,* vol. 415, pp. 240-250, 2014.

[205] R. D. Singh, A. Mittal and R. K. Bhatia, "3D convolutional neural network for object recognition: a review," *Multimedia Tools and Applications,* vol. 78, pp. 15951-15995, 2019.