



Article

Constituents Phase Reconstruction through Applied Machine Learning in Nanoindentation Mapping Data of Mortar Surface

Elias P. Koumoulos ^{1,2,*}, Konstantinos Paraskevoudis ¹ and Costas A. Charitidis ² 

¹ Innovation in Research & Engineering Solutions (IRES), Boulevard Edmond Machtens 79/22, 1080 Brussels, Belgium

² RNANO Lab—Research Unit of Advanced, Composite, Nano Materials & Nanotechnology, School of Chemical Engineering, National Technical University of Athens, GR-15773 Zographos Athens, Greece

* Correspondence: epk@innovation-res.eu

Received: 30 May 2019; Accepted: 14 June 2019; Published: 26 June 2019



Abstract: In the present study, data generated from nanoindentation were used in order to reconstruct the surface constituent phases of mortar grids through machine learning algorithms. Specifically, the K-Means algorithm (unsupervised learning) was applied to two 49 measurement (7×7) datasets with information about the modulus (E) and hardness (H) in order to discover the underlying structure of the data. The resulting clusters from K-Means were then evaluated and values range assigned so as to signify the various constituent phases of the mortar. Furthermore, another dataset from nanoindentation containing information about E, H, and the surface colour of the measured area (obtained from an optical microscope) was used as the training set in order to develop a random forests model (supervised learning), which predicts the surface colour from the E and H values. Colour predictions on the two 7×7 mortar grids were made and then possible correlations between the clusters, signifying constituent phases, and the predicted colours were examined. The groupings of data in the clusters (phases) corresponded to a unique surface colour. Finally, the constituent phases of the mortar grids were reconstructed in contour plots by assigning the corresponding cluster of the K-Means algorithm to each measurement (position in the grid).

Keywords: nanoindentation grid; mapping; machine learning; mortar contour reconstruction; data clustering

1. Introduction

Instrumentation requirements in the characterization field are changing rapidly in terms of quality assurance, desired accuracies, and measurement speed. Therefore, novel and innovative approaches regarding the design of instruments and characterization protocols are needed in order to address these new demands [1]. The nanoindentation technique can be considered as one credible and fast approach for rapid assessment for the mapping of localized surface mechanical integrity [2,3]. In particular, a specimen surface is placed under a diamond indenter with specific design and geometrical dimensions. Through an applied force, the indenter is driven into the measured surface while sensors and actuators record the values of the applied force and of the displacement [4]. Via this method, the loading and unloading curve of the probed material are generated. Using the theory of contact mechanics and assuming an elastic behavior, the hardness (referring to the contact pressure) and the elastic modulus of the material can then be calculated from the initial part of the unloading curve [5]. Furthermore, information on elastic recovery, plastic deformation [6], fracture [7], sensitivity of strain-rate and strain-hardening, the effect of residual stresses [8], and size [9] in plasticity locally are provided, making the nanoindentation technique a rich information source of the mechanics of a probed material

with a growing range of applications. An important and recent field of application of the nanoindenter is the assessment of in-situ properties of fibrous composite constituents [10,11]. However, the effects of sample preparation, neighboring constituents, residual stress, pile-up, time-dependent deformation, and hydrostatic stress on the important nanoindentation parameters and properties introduce new challenges, which need to be taken into consideration and further investigated [12]. As quick in-line and real-time testing can be considered as the new demands in manufacturing and industrial applications, the need for upscaling the nanomechanical assessment and mechanical properties mapping over large areas in composites and hybrids is growing [4,13,14].

Cement is a heterogeneous mixture of multiple phases. As a result, its structure may vary between colloidal nanometric units (nanometers) and large capillary pores (micrometers). The colloidal nature of the main hydrate phase in cement paste, also described as calcium silicate hydrate (C-S-H), plays a significant role in the overall mechanical properties of cement and is a subject of increasing research across materials scientists. Its nanogranular nature consists of nanometric (~5 nm) building blocks [15,16], which bind together the entire cement matrix when aggregated [17]. However, the extensive use of additives has resulted in further complexity inside the cement structure [18,19]. Chemical mixtures can be considered as a decisive factor regarding materials' properties improvement. For example, specific organic products are used in order to retard or accelerate the hydration kinetics or as a superplasticizer to give more fluidity without adding excess water. Furthermore, they can be used as viscosity modifiers for controlling the plasticity of the paste. Additionally, inorganic products are also used as additives. For instance, silica-based products, also called "supplementary cementitious materials" (SCM), exhibit pozzolanic properties and are used in order to decrease CO₂ emissions per ton in cementitious materials [20]. Specifically, for cement admixtures, several other nanoparticles are used as additives and can deliver special properties [21]. All the above admixtures (organic, inorganic, nanoparticles) alter the structure of the materials and affect the morphology of hydrated phases, including C-S-H [22–25]. The C-S-H phase, as well as the entire material, is also affected by these admixtures in their mechanical properties. For example, the nanoscale mechanical properties, the structure, and the composition at the nanoscale have to be studied together, as their relation is a very important prerequisite in order to understand the mechanical performance of a cement-based structure [26]. As a result, research focused on providing the industry with methods for nanoscale (or multiscale) characterization of cement-based materials is required, while industrial applications for cement and concrete formulation are getting more and more complex [27,28].

The interfacial transition zone (ITZ) is considered as one important area which affects the overall mechanical performance of concrete. However, the ITZ shows weak mechanical properties and, as a result, has been in the center of research and study focus [29]. The nanoindentation technique's use for characterizing cement-based materials is growing rapidly as it can help identify the local elastic properties and the hardness at micro-scale and at nanoscale [30,31]. This rapid growth can be further justified from the fact that nanoindentation is highly localized and, furthermore, provides a big number of elastic, plastic, and fracture properties without being destructive [32]. Specifically, as of the ITZ, the nanoindentation technique can provide important information about its micro-mechanical properties [31]. However, the relationship between the micro-mechanical properties of ITZ and its microstructures has not been adequately studied and research. Research around this relationship could greatly help to better understand the ITZ.

Data science and machine learning ease the efficient mining and potential for further processing of large material data sets, resulting in the extraction and identification of high-value material knowledge, towards design, quality, and manufacturing. This is accomplished by using linkages of process-structure-property (PSP) information, with the main focus of data transformations to be in the forward direction (process → structure → properties). Therefore, as high-value information requires linkage with the manufacturing and product design routes, the main challenge is, starting from a proper data management plan, to design and build the needed database stems (tackling challenging issues such as rich internal materials structures that span multiple length scales). Data science foresight

into the advanced characterisation of materials mitigates the inherent risk to a large extent, not only by making decisions more concrete (e.g., in design and manufacturing), but also by capturing failures and successes. Information from this is then useful and processable to and from other disciplines. An effective mitigation plan based on data management is strongly based on the availability of data and the use of data-driven protocols, as the uncertainty associated with the information and knowledge used in making decisions (in materials development workflows) is then quantifiable.

The nanoindentation technique is a paradigm of tip-sample interactions, where an indenter penetrates a material by an applied load and the values of displacement and load are returned. The signals (displacement, load) are highly dependent from many parameters, making generated nanoindentation data complex and not easy to understand. Machine learning exhibits a great potential in converting nanoindentation data into rich and readable information. Furthermore, mining nanoindentation data can provide valuable information on the behavior of materials and can reveal hidden patterns and underlying structures.

The implication for the design is as follows: In conventional cements, even if the calcium-to-silica ratio ranges (from about 1.2 to 2.2), the resulting molecular structures are not discussed and assessed in detail so as to feed the design process. Apart from building a database, machine learning could be exploited in qualification of the molecular structure of the hardened material transition from a tightly ordered crystalline structure to a disordered glassy structure, depending on the various ranges. Apart from mechanical properties, as emissions related to concrete production are estimated from 5% to 10% of industrial greenhouse-gas emissions, this approach could potentially be used in achieving a reduction in calcium content in the cement mix, directly impacting the CO₂ with a reduction in carbon emissions of up to 60 percent. Additionally, using machine learning, potential alternate materials available in nature could be investigated to replace Portland cement (with a great potential being sustainable and longer-lasting alternatives), without requiring a huge amount of energy to manufacture.

2. Machine Learning Principles

The growing generation of data in the characterization field brings up new opportunities and challenges. The huge spreading of machine learning applications now offers great tools for the exploitation of generated characterization data. As a result, materials scientists turn their focus on the computational applications that can deal with the complexity of these data and deliver important results [33]. Data-driven approaches can provide valuable information regarding the underlying structure and behavior of materials [34]. To name a few, opportunities fall within the areas of the material's property prediction, structure prediction, automated determination of phase diagrams using high-throughput combinatorial experiments, advanced and smart materials modelling and discovery, and artificial intelligence in the characterization field. More specifically, machine learning in materials science has been used for structure-property relationship prediction, crystal structure prediction, micrograph analysis, and descriptors identification [34–36]. Furthermore, supervised regression algorithms can be used for the prediction of mechanical properties of constructive materials [37]. Another application of machine learning in materials science is the inverse design of materials, in which used algorithms help identify materials that satisfy desired properties [38].

2.1. Supervised and Unsupervised Machine Learning

Machine Learning is a subsection of artificial intelligence which applies computational methods using prior knowledge and data in order to improve performance or make accurate predictions [39,40]. Various algorithms are used to go through data and, from it, learn patterns, rules, or underlying structures. These algorithms are then used in order to make predictions or determinations. Specifically, most algorithms aim to build a function out of some input variables and improve it while looking at more and more data [41]. Machine learning algorithms can be divided into two main categories, supervised and unsupervised. In supervised learning, there are labels (Y) for each input variable (x) and the algorithm aims to learn the mapping function from the input to the output $Y = f(x)$. In a

given dataset, data can be represented as a big table of instances with columns devoting different attributes (features). Every instance has a set of feature values which is also known as the input data (x). The features can be continuous, categorical, or binary. In supervised learning, instances are also labeled, which is that they have an output variable (Y). Learning from labeled input data (x with known y), the goal is to approximate the mapping function accurate enough to make predictions on new input data. Therefore, the term “supervised” refers to the existence of a training set (labeled input data) which acts as the supervisor of the whole learning process [34,42].

Supervised learning can be further categorized into regression and classification. In regression problems, the output variable is a number, while regression algorithms learn from labeled data and aim to predict a real value. A classification problem has a categorized output variable, which has no numerical significance. In this case, the algorithm learns from labeled data in order to predict the class of new input data. In unsupervised learning, there are only input data (x) and no labels (y). The challenge in these problems relies on finding an accurate way to create classes that best group similar input data. Therefore, the goal is to model the underlying structure or distribution in the data. Unlike supervised learning, in unsupervised learning there is no training set and algorithms try to discover patterns and groupings without prior knowledge. Clustering is a main subcategory of unsupervised learning and aims to discover the natural groupings and useful classes of items [34,39].

2.2. K-Means Clustering

The K-Means algorithm is popular for cluster analysis in data mining and is part of the unsupervised machine learning algorithms, as no labels and no training sets exist. The K-Means algorithm aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The number of clusters is known and each point can belong only in one cluster. At first, random k cluster center points are generated and each data point is assigned to the cluster with the nearest center point (smallest Euclidean distance). Then, the mean of each cluster is calculated and the k cluster centers are replaced by the corresponding cluster mean. Again, each point is assigned to the nearest cluster, measured in Euclidean distance. This procedure loops until no further updates of the cluster centers can occur. In this case, the K-Means algorithm has converged [33].

In Figure 1, the K-Means algorithm is presented, while Figure 2 describes an indicative example of the algorithm with 3 clusters. In the first step are the raw unclustered data. Three random centers are assigned (2nd step). In step 3, the data points are assigned to the center with the smallest Euclidean distance and the corresponding clusters occur. The centers are updated by calculating the mean for every cluster (4th step). In step 5, the data points are assigned to the new updated centers from step 4 with the smallest Euclidean distance. The centers are updated (6th step) likewise and the data points are assigned to the corresponding cluster (7th step). In step 8, the algorithm K-Means has converged and the centers can no longer be updated.

Input: A set of d -dimensional points $\{x_1, x_2, \dots, x_m\}$

1. **Initialize the means $\mu_1, \mu_2, \dots, \mu_k$ randomly**
2. **Loop: (Find the label, whose mean is closest in Euclidean distance to the point)**
 - a. **Label each point as a mean closest to it Label for $x_i = \operatorname{argmin} \|x_i - \mu_\lambda\|^2$**
 - b. **For every label l : – Re-compute the mean μ_l as the average of all points that were assigned to it**

Figure 1. K-Means Algorithm steps.

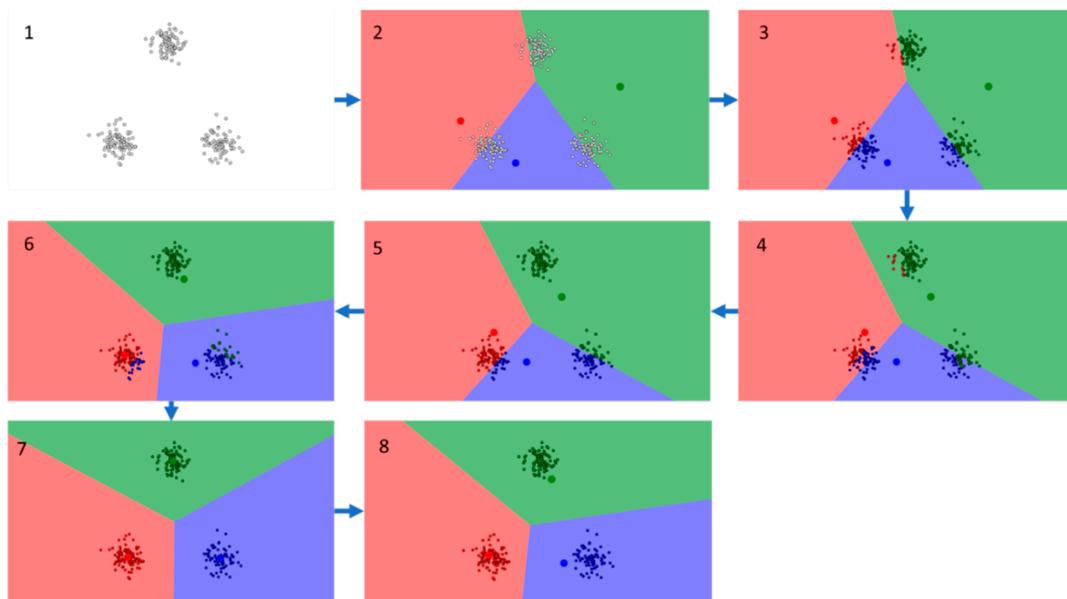


Figure 2. An example of K-Means clustering with 3 Clusters. In the first step are the raw unclustered data. Three random centers are then assigned (2nd step), while in step 3, the data points are assigned to the center with the smallest Euclidean distance and the corresponding clusters are revealed. This procedure continues in the next steps until the algorithm has converged and the cluster centers can no longer be updated (step 8).

2.3. Decision Trees and Random Forests

Decision trees in machine learning are supervised algorithms, as they signify a hierarchical data structure that represents data using a divide-and-conquer strategy and can be used as hypothesis class for non-parametric classification or regression (Figure 3). The main idea relies on learning a decision tree that best represents a given collection of examples. Then, applying the learned decision tree on new unlabeled input Data, the label can be predicted. Decision trees consist of nodes, branches and leaves. Nodes are tests for feature values, meaning there is one branch for every value that the feature can take. The leaves of the tree specify the class labels (output variable). The task is to develop a classification rule that can determine the class of any object from its values of the attributes [37]. The main decision in the algorithm is the selection of the next attribute to split on. Furthermore, a quite challenging aspect is to find the root node, which is the attribute to start making decisions on. The importance of each attribute relies on how it splits the examples. More important attributes split the data to sets that are relatively pure in one label. In order to address the above challenges, the information gain of each attribute is calculated. Information gain is the amount of information gained about a random variable or signal from observing another random variable [37]. An attribute with higher information gain will be tested (or split on) first. With the most important attribute being found, a new prediction starts at the root node (top node) and continues through the tree until reaching a final node (called leaf node). In any other node, the branch to follow is chosen by the value test, which tests the corresponding attribute [39].

A variation of decision trees are the random forests, which are an ensemble learning method. Ensembles use multiple learning algorithms and combine their output (e.g., the average) to give a final determination [39]. Random forests build multiple decision trees and combine them to improve the overall performance. The combined individual decision trees show an additional randomness compared to the original decision tree algorithm and this is the reason behind the name “random”. Here, the most important attribute to split on next is searched among a random subset of attributes. The final sub decision trees are merged together and their predictions are combined into a single one, giving better results.

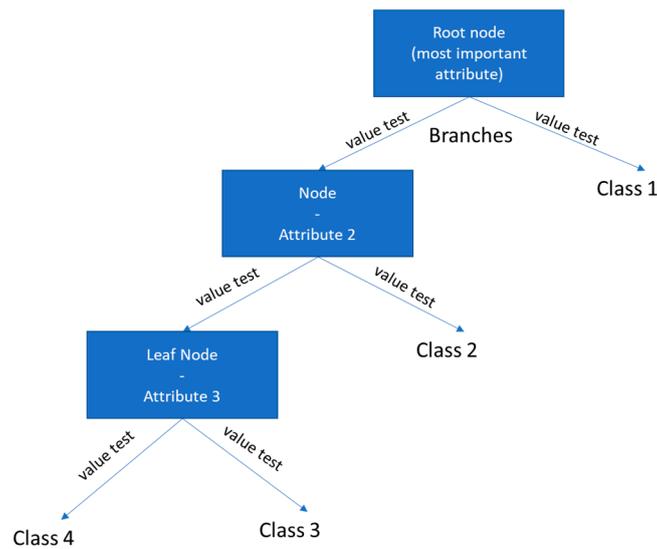


Figure 3. Example of representative overview of a decision tree.

3. The Dataset: Experimental Details and Methodology

The nanomechanical integrity of the materials was assessed through a Hysitron TriboLab[®] Nanomechanical Test Instrument (Minneapolis, MN, USA), which allows the application of loads from 1 μN to 30 mN and records the displacement as a function of applied loads with a high load resolution (1 nN) and a high displacement resolution (0.04 nm). In all measurements, a total of 10 indents were averaged, so as to determine the mean hardness (*H*) and elastic modulus (*E*) values for statistical reasons, with a 50 μm spacing, in a clean area environment with 45% humidity, and 23 °C ambient temperature. In order to operate under closed loop control, the feedback control option was selected. All measurements have been performed at 200 nm of displacement (so as for the indentation response not to be affected by the discrete nature of colloidal particles and microstructure/interaction of different phases), using the standard three-sided pyramidal Berkovich tip indenter, with an average curvature radius of 100 nm [43,44].

Considering the half-space elastic deformation theory, *H* and *E* values can be calculated from the experimental data (load displacement curves) using the Oliver–Pharr (O&P) model [6,9]. The derived expressions for calculating *E* extracted from indentation measurements are based on Sneddon’s elastic contact theory:

$$E_r = \frac{S \sqrt{\pi}}{2\beta \sqrt{A_c}} \tag{1}$$

where *S* is the unloading stiffness (initial slope of the unloading load-displacement curve at the maximum displacement of penetration (or peak load)), *A_c* is the projected contact area between the tip and the substrate, and β is a constant that depends on the geometry of the indenter (β = 1.167 for the Berkovich tip [45,46]). Typical nanoindentation hardness refers to the mean contact pressure required. This hardness, which is the contact hardness (*H_c*), is actually dependent upon the area geometry of the tip indenter (Equations (2)–(4)).

$$H_c = F/A \tag{2}$$

where,

$$A(h_c) = 24.5h_c^2 + a_1h_c + a_{1/2}h_c^{1/2} + \dots + a_{1/16}h_c^{1/16} \tag{3}$$

and

$$h_c = h_m - \varepsilon \frac{P_m}{S_m} \tag{4}$$

where h_m is the total penetration displacement of the indenter at peak load, P_m is the peak load at the indenter displacement h_m , and ε is an indenter geometry constant, equal to 0.75 for the Berkovich indenter [45–48]. Prior to indentation, the area function of the indenter tip was calibrated in a fused quartz/silica (standard material) [49].

Two 49 measurement datasets were used from nanoindentation experiments on mortar with information about the elastic modulus (E , GPa) and hardness (H , GPa), which are both continuous quantitative attributes. Each measurement corresponds to a position in the 7×7 grid. The goal was to use machine learning algorithms to cluster the data and evaluate the phase of each measurement and also to discover a possible correlation between optical properties (colour) and the clustered phase from the first step. For the second part, training a model from other labeled data was required, as the two grids do not contain any information about the surface colour. For this purpose, another nanoindentation dataset with labeled surface colour for every instance was used as the training set in order to construct the classifier (using random forests algorithm). This classifier was then used to predict the surface colour for each measurement of the two 7×7 grids. Finally, contours were created in order to reconstruct the constituent phase of the mortar grid just from E and H measurements.

4. Results and Discussion

For the mortar sample (Figure 4), the brightness order through scanning electron microscopy imaging from high to low is reported as unhydrated cement clinker, followed by hydration products, and finally pores [29]. The phases, including clinker (3 types), unhydrated cement clinker/hydration products, and pores present in mortar sample, were identified in the optical images, as marked in the table below (Table 1).

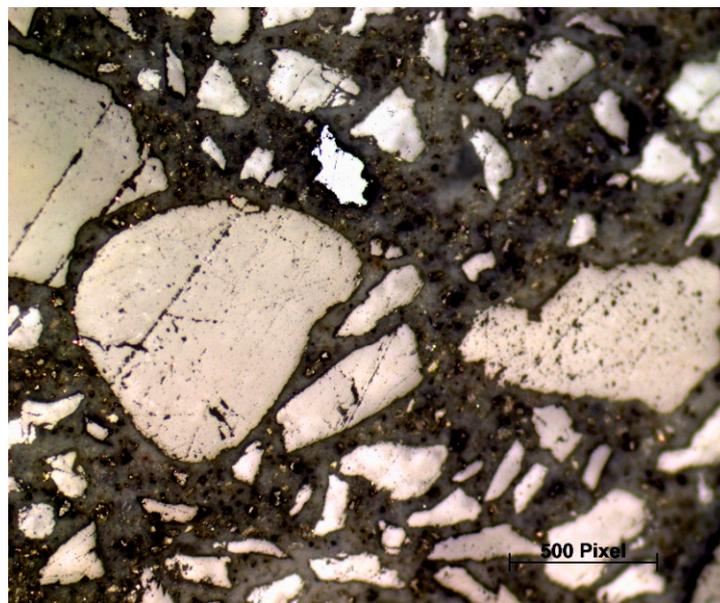
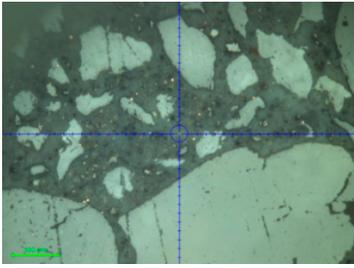
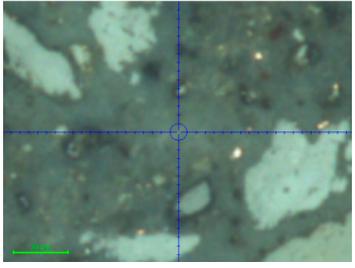
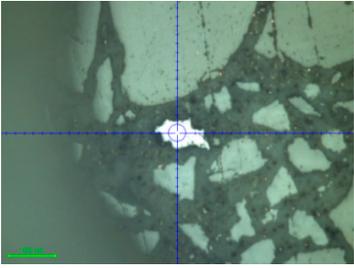
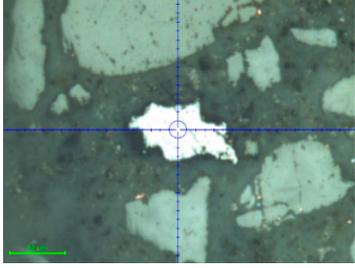
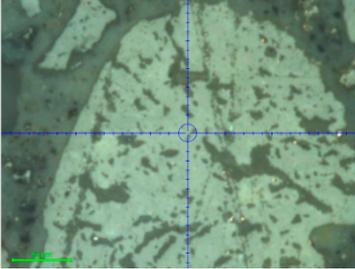
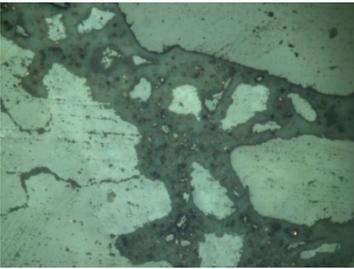


Figure 4. Optical image of the mortar sample.

Nanoindentation testing was performed to all identified areas using the optics system of the indenter (Table 1). The magnification is embedded in the optical images, while the blue cross denotes the center of the selected area where the measurements were conducted. In Figure 5, the load-unload curves of clinker identified constituents (gold, grey porous, and white, respectively) and pores (black) area are presented, exhibiting their different responses to the applied load (e.g., resistance to applied load, amount of energy stored during a nanoindentation load-unload cycle/elastoplastic deformation) and (for the case of the pores area) exhibiting energy releasing phenomena. In principle, this could be attributed to the possibility of the rapid production/energy release of high-pressure phases through fast

mechanical loading and unloading and energy sequential releasing. In our case, it could be attributed to sliding upon unloading, at the latter stage, in the pores area case. It should be noted that H and E are calculated in the max applied load and initial unloading parts, respectively, and thus do not influence our results.

Table 1. Typical optical image of the mortar sample showing the material phases/constituents (right images present left images in higher magnification).

| | | |
|---|---|---|
| <p>1. Cement (grey), containing low/high density C–S–H phase (LD/HD C–S–H) and portlandite (CH)</p> |  |  |
| <p>2. Clinker (white)</p> |  |  |
| <p>3. Clinker (left, gold) and Porous clinker (right, grey)</p> |  |  |
| <p>4. Pores (black)</p> |  |  |

Elastic modulus (E) (Figure 6a) and hardness (Figure 6b) values are presented in black (as 3D), while their projection is noted with blue points. The results depicted below correspond to 49 measurements points (grid was set to 7×7 measurements-dataset No. 2, in a selected area of $35 \times 35 \mu\text{m}^2$). On the left of the z-axis, the ranges of individual phase nanomechanical properties are mentioned for comparison. In Figure 7, E and H contour maps obtained through nanoindentation mapping protocol are presented.

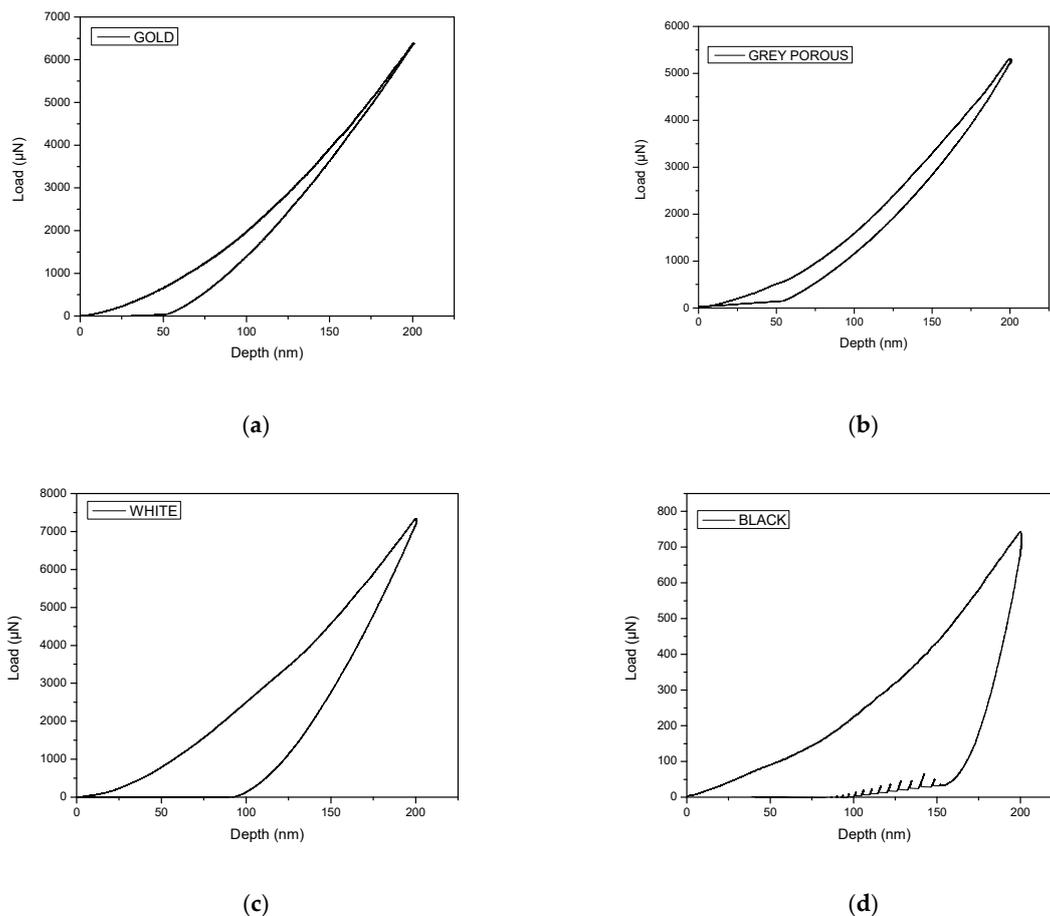
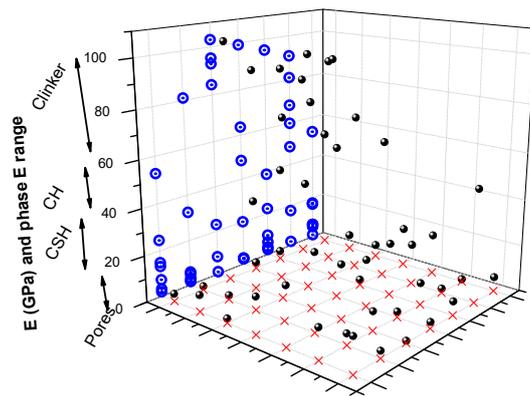


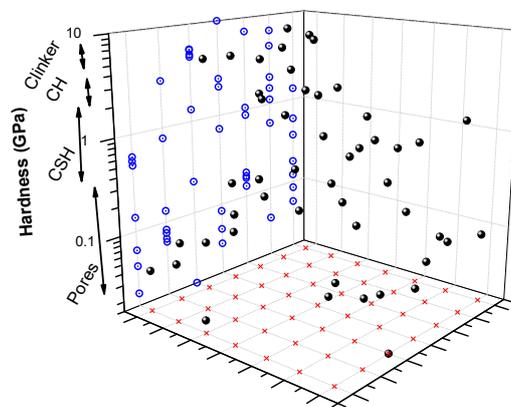
Figure 5. (a–d) Load-unload curves of clinker identified constituents (gold, grey porous, and white, respectively) and pores (black).

After measuring the different areas separately, i.e., porous, white, black and grey area, a nanoindentation mapping protocol was used in an area of mainly grey colour (cement phases area and clinker residues). For the cement phases, the phases below were considered, excluding the clinker constituent (of ratio E/H reported as 14.5 [50]). With a low stiffness/hardness phase of macroporosity (denoted as pores-MP), of which E was reported to vary from 4 to 6.7 GPa [29] and from 0 to 13 GPa [50], the ratio E/H was reported as 47.3. These very low properties were attributed to material regions dominated by the capillary pores volume fraction. The C–S–H phase E reported to vary from 10.9 to 29.7 GPa [29]; however, it is often reported as low- and high-density C–S–H phases separately.

The low-density C–S–H phase (denoted as LD C–S–H) present, revealed an E either varying from 13 to 26 GPa [49], or ~23.6 GPa [51–53]. Its hardness ranged from 0.4 to 0.8 GPa [50,53] and the ratio E/H was reported as 40.3 [52]. Afterwards, the high-density C–S–H phase present, denoted as HD C–S–H, revealed an E either varying from 26 to 39 GPa [50], or ~32 GPa [51,53]; its hardness ranged from 0.8 to 1.25 GPa [50,53] and the ratio E/H was reported as 35 [50]. Finally, a portlandite (denoted as CH) phase was present. E was reported as higher than 39 GPa [52,53], while CH /interface was also reported with E ranging from 35.1 to 42.9 GPa [29]. Its hardness ranged from 1.31 to 1.66 GPa [49,52] and the ratio E/H is reported as 30.9 [50]. Statistical analysis, such as a statistical deconvolution technique, has been reported and applied on hundreds of nanoindentation tests in order to evaluate the results and conclude the above assessment of ranges to the constituent phases [50]. In Figure 8, a histogram of the elastic modulus of mortar sample constituents is presented. The three constituents' phases are identified and set [29], according to literature.



(a)



(b)

Figure 6. (a) Modulus and (b) hardness obtained through nanoindentation mapping protocol, dataset No. 2 (49 data points: 7×7 mapping dataset). In red the grid measured points are noted, while clinker, pores, CSH and CH range values are presented in z axis.

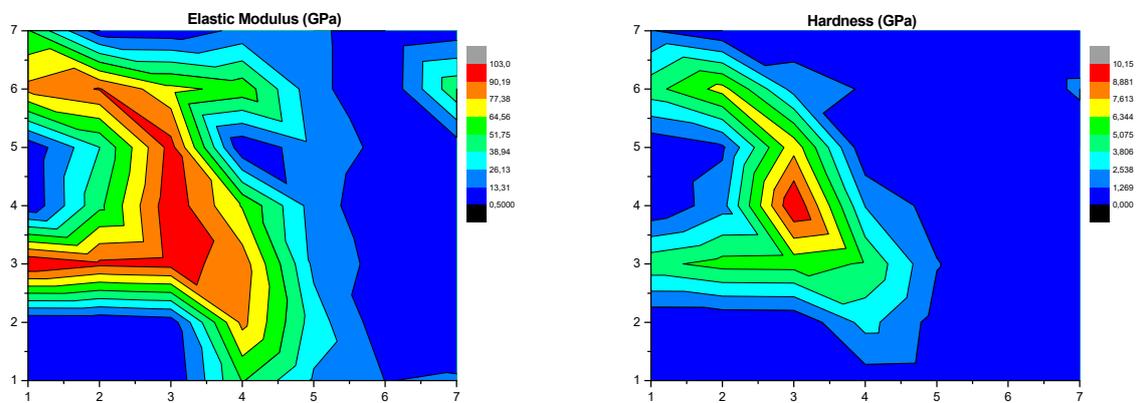


Figure 7. E and H contour map obtained through the nanoindentation mapping protocol, dataset No. 2 (49 data points: 7×7 mapping dataset).

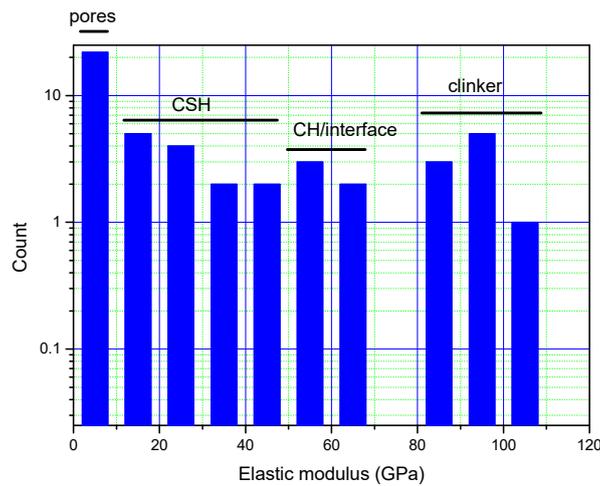


Figure 8. Histogram of the elastic modulus of mortar sample constituents, using the phase ranges from literature.

In Figure 9, a mortar specimen constituent phase identification (159 data points: Two 7×7 mapping datasets, individual mapping datasets) is presented through correlation of applied load to elastic modulus (a) and further grouping of values (b), based on the work in the literature [50–53]. The individual clinker constituent’s hardness to modulus correlation is presented in Figure 10 below, based on colour identification through nanomechanical instrument optics.

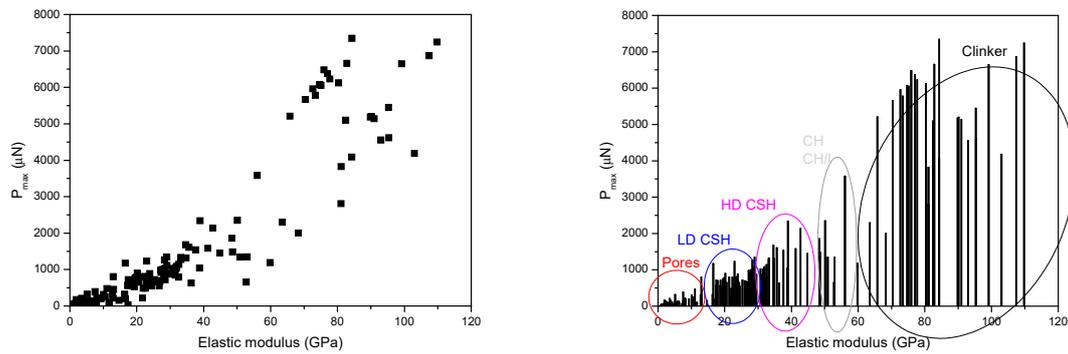


Figure 9. Mortar specimen constituent phase identification (159 data points: Two 7×7 mapping datasets, individual mapping datasets), based on P_{max} vs. modulus plotting.

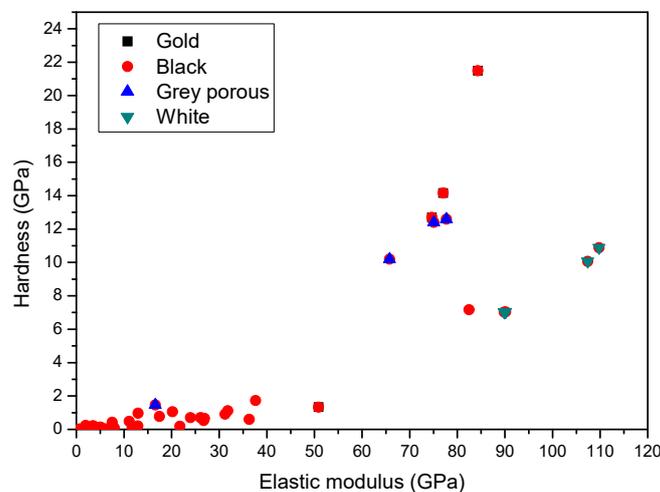


Figure 10. Individual clinker constituents’ hardness to modulus correlation.

In Figures 11 and 12 the mortar specimen constituent phase identification for the two datasets is presented, based on [53] phase ranges. As one means of quantifying the dependence of the heterogeneous response is the indentation modulus-to-hardness ratio [50], the following E/H contour map was generated for the dataset (No. 2 (49 data points: 7×7 mapping dataset)) in Figure 13.

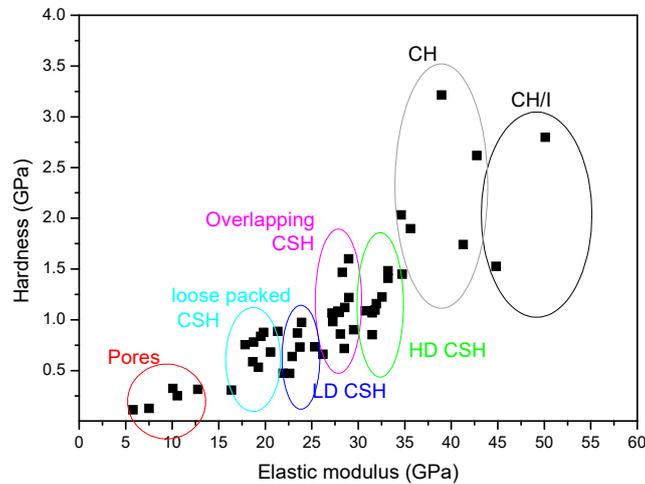


Figure 11. Mortar specimen constituent phase identification No. 1 (49 data points: 7×7 mapping dataset), based on [53] phase ranges.

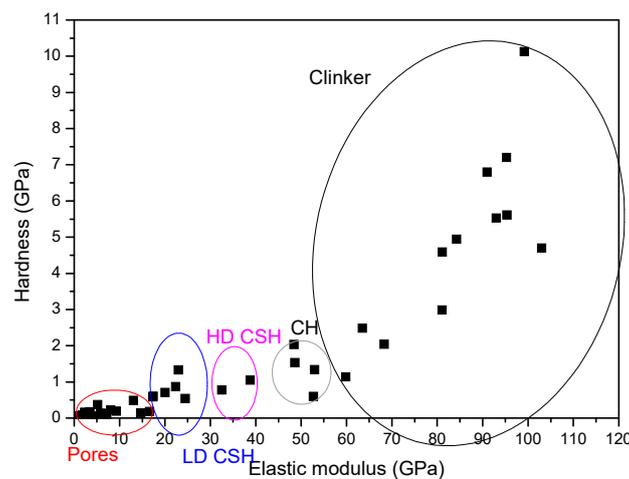


Figure 12. Mortar specimen constituent phase identification No. 2 (49 data points: 7×7 mapping dataset), based on [53] phase ranges.

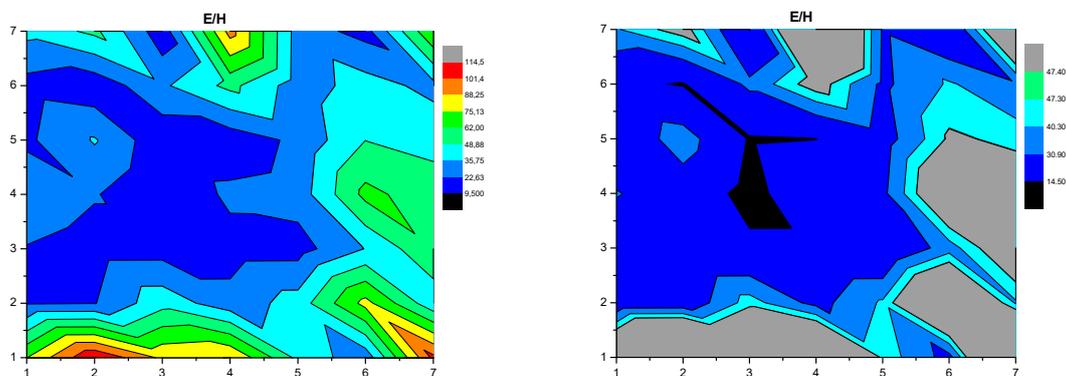


Figure 13. E/H contour map obtained through nanoindentation mapping protocol (on the right), phase- E/H set correlation using [52] ranges.

K-Means was applied on two 49-measurements grids in order to extract information about the underlying structure of the data. For the first grid, the optimum amount of clusters was set as 7. In Figures 14 and 15, the resulting clusters from K-Means on the 1st grid are presented. Each cluster signifies a constituent phase of the mortar, for each of the datasets, respectively.

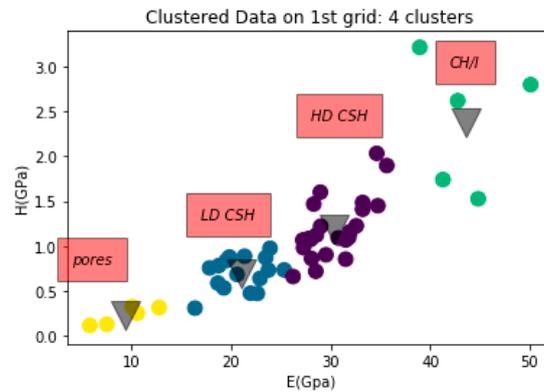


Figure 14. Resulting clusters from K-Means on 1st mapping grid. Each cluster signifies a constituent phase of the mortar.

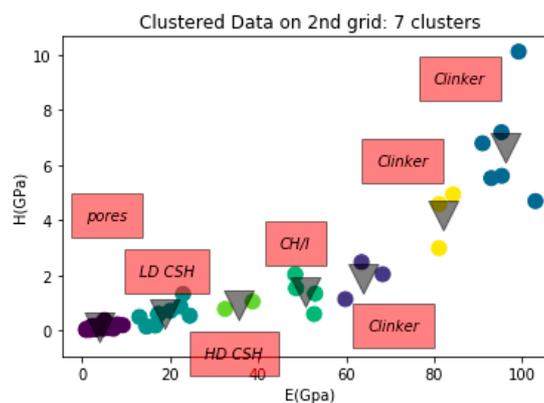


Figure 15. Resulting clusters from K-Means on 2nd mapping grid. Each cluster signifies a constituent phase of the mortar.

In Figure 14, the resulting clusters from K-Means on the 1st grid are presented with a different colour. K-Means successfully clustered the data as expected. Specifically, the first cluster (yellow) takes E values between 3–14 GPa. According to literature [29,50,51,53], this cluster signifies the pores inside the mortar. The next cluster (sea green) includes data points with E values in the range between 16 and 26 GPa and corresponds to the LD C-S-H phase. Blue coloured data belong to the third cluster, which signifies the HD C-D-H phase. Finally, the green cluster refers to data points belonging in the E values range, between 41 and 51, and can be considered as the CH/I phase.

On Figure 15, the resulting clusters from K-Means on the 2nd grid are presented. In this case, the resulting clusters are 7 and cover a bigger range of E values. As a result, clinker phases are also expected. Again, the first cluster takes values between 0.2 and 15 GPa and corresponds to the pores inside the mortar. The next three clusters are also similar with the K-Means clusters from 1st grid. LD C-S-H includes as expected values between 16 and 25 GPa. The third cluster (containing 2 points) shows the HD C-S-H phase and its E values are in the range between 30 and 40 GPa. The fourth cluster refers to CH/I and has as expected points with E values between 45 and 58 GPa. Clinker phases are the last 3 clusters and take values between 60 and 105 GPa. Clinker can be clustered into 3 groups, as follows: One with E values between 60 and 75 GPa, one with E values between 80 and 90 GPa, and one with E values above 90 GPa. It is clear that K-Means shows similarities in the cluster ranges

for both grids, although the number of total clusters is not the same. Furthermore, via this, the various ranges of E values have been assigned to a specific phase according to the bibliography and this has been validated throughout K-Means. In Table 2 below, the literature and this work’s values, based on the clustering method, are presented.

Table 2. E and H values from literature [29,50–53] and K-Means clustering for each mortar phase.

| Phases | E (GPa) | | H (GPa) | |
|----------|------------|---------------------|------------|---------------------|
| | Literature | Cluster (This Work) | Literature | Cluster (This Work) |
| Pores | 0–13 | 0.2–15 | 0.16–0.18 | 0.1–0.35 |
| LD C-S-H | 13–26 | 16–26 | 0.4–0.8 | 0.4–1.8 |
| HD C-S-H | 26–39 | 26–40 | 0.8–1.25 | 0.4–2.1 |
| CH-CH/I | 35.1–42.9 | 41–58 | 1.31–1.66 | 0.8–3.2 |
| Clinker | - | >60 | - | >1.8 |

In Figure 16, the predicted colours from the random forests algorithm in the 1st grid are presented. The 4 points (triangles) signify the clusters from the K-Means algorithm of Figure 14. The random forests algorithm predicted measurements with E values lower than 18 GPa as a black surface. All the rest of points (>18 GPa) were predicted as a grey surface. Via this, the 1st cluster of Figure 14 (pores) can be correlated with the black surface. In addition, clusters 2,3, and 4 from Figure 14 (LD CSH, HD CSH, CH/I) can be correlated with the gray coloured surface.

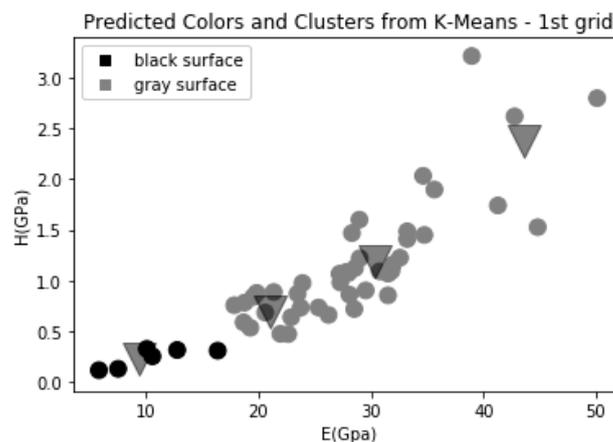


Figure 16. Predicted surface colours for each measurement of the 1st mapping grid from the random forests algorithm.

In Figure 17, the predicted colours from the random forests algorithm in the 2nd grid are presented. The 7 points (triangles) signify the clusters from the K-Means algorithm of Figure 15. The random forests algorithm predicted measurements with E values lower than 18 GPa as black surface. Points between 18 GPa and 58 GPa were predicted as grey surface. Via this, the 1st cluster of Figure 15 (pores) can be correlated with the black surface. In addition, clusters 2,3, and 4 from Figure 15 (LD CSH, HD CSH, CH/I) can be correlated with the gray coloured surface. Points with values between 60 GPa and 71 GPa were predicted as grey porous surface, which can be correlated with the cluster “Clinker” of Figure 15. Likewise, gold surfaces (80–88 GPa) and white surfaces (>90 GPa) can also be correlated with the cluster “Clinker” in Figure 15. Therefore, the cluster “Clinker” in Figure 15 contains points with 3 possible colours (grey porous, gold, white). In Table 3, the resulting ranges for E , H , and plasticity (integrated area of load-unload curve) values for each surface colour are presented.

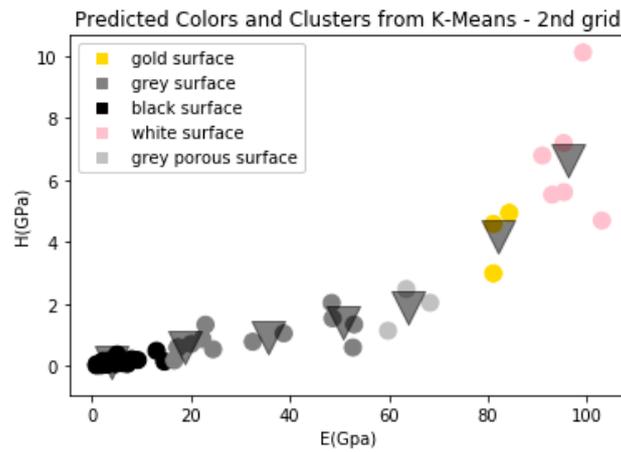


Figure 17. Predicted surface colours for each measurement of the 2nd mapping grid from the random forests algorithm.

Table 3. Resulting ranges for *E* and *H* values and plastic deformation for each surface colour.

| Surface Colour | <i>E</i> (GPa) | <i>H</i> (GPa) |
|----------------|----------------|----------------|
| black | <18 | <0.4 |
| grey | 18–58 | 0.5–3.1 |
| grey porous | 60–71 | 1–2.2 |
| gold | 80–88 | 2.4–5 |
| White | >90 | >4.2 |

In Figure 18, the reconstruction of the constituent phases of the resulting K-Means (Figure 14) in the 1st grid is presented. The deep blue colours signify pores in the corresponding areas. Light blue corresponds to LD CSH areas, while HD CSH areas are the deeper green areas and CH/I areas correspond to the light green colours. The reconstruction of the constituent phases of the resulting K-Means (Figure 15) in the 2nd grid is presented in Figure 19. Deep blue colours signify pores in the corresponding areas. Light blue corresponds to LD CSH areas, while HD CSH areas are the deeper green areas and CH/I areas correspond to the light green colours (yellow areas correspond to Clinker).

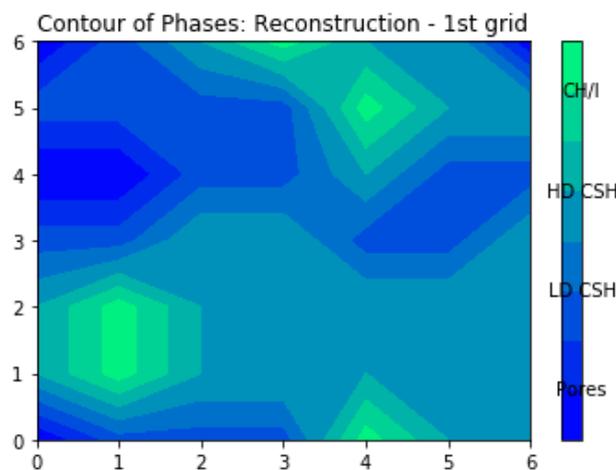


Figure 18. Reconstruction of constituent phases for the 1st mapping grid dataset from elastic modulus.

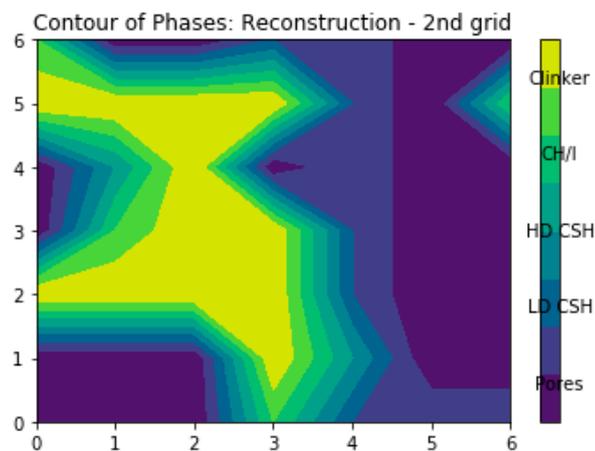


Figure 19. Reconstruction of constituent phases for the 2nd mapping grid dataset from elastic modulus.

5. Conclusions

In this work, constituent phase reconstruction through applied machine learning in nanoindentation mapping data of a mortar surface was conducted.

Firstly, the load-unload curves of the clinker identified the constituents (gold, grey porous, and white, respectively) and pores (black) areas exhibited different response to the applied load (e.g., resistance to applied load,) and (for the case of pores area) exhibited energy releasing phenomena. After measuring the different areas in separate, i.e., porous, white, black, and grey areas, a nanoindentation mapping protocol was used in an area of mainly grey colour (cement phases area and clinker residues). Through elastic modulus histogram creation, the three constituents' phases of cement were additionally revealed to the clinker residuals. The individual clinker constituent's hardness to modulus correlation was assessed, based on colour identification, through nanomechanical instrument optics. In order to quantify the dependence of the heterogeneous response to the indentation modulus-to-hardness ratio, an E/H contour map was generated.

Then, machine learning algorithms were used to cluster the data and evaluate the phase of each measurement. Additionally, the possible correlation between optical properties (colour) and the clustered phase from the first step was assessed. For the second part, training a model from other labeled data was required, as the two grids did not contain any information about the surface colour. For this purpose, another nanoindentation dataset, labeled with the surface colour for every instance, was used as training set in order to construct the classifier (using the random forests algorithm). This classifier was then used to predict the surface colour for each measurement of the two 7×7 grids. Finally, contours were created in order to reconstruct the constituent phase of the mortar grid just from E and H measurements.

Author Contributions: Conceptualization, E.P.K. and K.P.; Methodology, E.P.K. and K.P.; Software, K.P.; Validation E.P.K. and K.P.; Formal Analysis, E.P.K. and K.P.; Investigation, E.P.K. and K.P.; Resources, C.A.C., E.P.K.; Data Curation, E.P.K. and K.P.; Writing-Original Draft Preparation, C.A.C., E.P.K. and K.P.; Writing-Review & Editing, C.A.C., E.P.K. and K.P.; Visualization, C.A.C., E.P.K. and K.P.; Supervision, C.A.C., E.P.K. and K.P.; Project Administration, C.A.C., E.P.K.; Funding Acquisition, C.A.C., E.P.K.

Funding: This research was partially funded by European Union's Horizon 2020 Research and Innovation Programme MODCOMP project under grant number 685844 and partially by European Union's Horizon 2020 Research and Innovation Programme LORCENIS project under grant agreement no. 685445.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tofail, S.A.M.; Koumoulos, E.P.; Bandyopadhyay, A.; Bose, S.; O'Donoghue, L.; Charitidis, C. Additive manufacturing: Scientific and technological challenges, market uptake and opportunities. *Mater. Today* **2018**, *21*, 22–37. [[CrossRef](#)]
2. Leatherbarrow, A.; Wu, H. Mechanical behaviour of the constituents inside carbon-fibre/carbon-silicon carbide composites characterised by nano-indentation. *J. Eur. Ceram. Soc.* **2012**, *32*, 579–588. [[CrossRef](#)]
3. Urena, A.; Rams, J.; Escalera, M.D.; Sanchez, M. Characterization of interfacial mechanical properties in carbon fiber/aluminium matrix composites by the nanoindentation technique. *Compos. Sci. Technol.* **2005**, *65*, 2025–2038. [[CrossRef](#)]
4. Koumoulos, E.P.; Tofail, S.A.M.; Silien, C.; de Felicis, D.; Moscatelli, R.; Dragatogiannis, D.A.; Bemporad, E.; Sebastiani, M.; Charitidis, C.A. Metrology and nano-mechanical tests for nano-manufacturing and nano-bio interface: Challenges & future perspectives. *Mater. Des.* **2018**, *137*, 446–462. [[CrossRef](#)]
5. Koumoulos, E.P.; Charitidis, C. Surface analysis and mechanical behaviour mapping of vertically aligned cnt forest array through nanoindentation. *Appl. Surf. Sci.* **2017**, *396*, 681–687. [[CrossRef](#)]
6. Koumoulos, E.P.; Charitidis, C.A. Integrity of Carbon-Fibre Epoxy Composites through a Nanomechanical Mapping Protocol towards Quality Assurance. *Fibers* **2018**, *6*, 78. [[CrossRef](#)]
7. Koumoulos, E.P.; Jagdale, P.; Kartsonakis, I.A.; Giorelli, M.; Tagliaferro, A.; Charitidis, C.A. Carbon nanotube/polymer nanocomposites: A study on mechanical integrity through nanoindentation. *Polym. Compos.* **2015**, *36*, 1432–1446. [[CrossRef](#)]
8. Charitidis, C.A.; Dragatogiannis, D.A.; Koumoulos, E.P.; Kartsonakis, I.A. Residual stress and deformation mechanism of friction stir welded aluminum alloys by nanoindentation. *Mater. Sci. Eng. A* **2012**, *540*, 226–234. [[CrossRef](#)]
9. Charitidis, C.A.; Koumoulos, E.P.; Dragatogiannis, D.A. Nanotribological behavior of carbon-based thin films: Friction and lubricity mechanisms at the nanoscale. *Lubricants* **2013**, *1*, 22–47. [[CrossRef](#)]
10. Hu, Z.; Farahikia, M.; Delfanian, F. Fiber bias effect on characterization of carbon fiber-reinforced polymer composites by nanoindentation testing and modeling. *J. Compos. Mater.* **2015**, *49*, 3359–3372. [[CrossRef](#)]
11. Maurin, R.; Davies, P.; Baral, N.; Baley, C. Transverse properties of carbon fibres by nano-indentation and micro-mechanics. *Appl. Compos. Mater.* **2008**, *15*, 61. [[CrossRef](#)]
12. Hardiman, M.; Vaughan, T.J. A review of key developments and pertinent issues in nanoindentation testing of fibre reinforced plastic microstructures. *Compos. Struct.* **2017**, *180*, 782–798. [[CrossRef](#)]
13. Ulm, F.J.; Vandamme, M.; Bobko, C.; Ortega, J.A.; Tai, K.; Ortiz, C. Statistical indentation techniques for hydrated nanocomposites: Concrete, bone, and shale. *J. Am. Ceram. Soc.* **2007**, *90*, 2677–2692. [[CrossRef](#)]
14. Sebastiani, M.; Moscatelli, R.; Ridi, F.; Baglioni, P.; Carassitia, F. High-resolution high-speed nanoindentation mapping of cement pastes: Unravelling the effect of microstructure on the mechanical properties of hydrated phases. *Mater. Des.* **2016**, *97*, 372–380. [[CrossRef](#)]
15. Allen, A.J.; Thomas, J.J.; Jennings, H.M. Composition and density of nanoscale calcium–silicate–hydrate in cement. *Nat. Mater.* **2007**, *6*, 311–316. [[CrossRef](#)] [[PubMed](#)]
16. Chiang, W.S.; Ferraro, G.; Fratini, E.; Ridi, F.; Yeh, Y.Q.; Jeng, U.S.; Chen, S.H.; Baglioni, P. Multiscale structure of calcium-and magnesium-silicate-hydrate gels. *J. Mater. Chem. A* **2014**, *2*, 12991–12998. [[CrossRef](#)]
17. Shu, X.; Graham, R.K.; Huang, B.; Burdette, E.G. Hybrid effects of carbon fibers on mechanical properties of Portland cement mortar. *Mater. Des.* **2015**, *65*, 1222–1228. [[CrossRef](#)]
18. Soto-Pérez, L.; Lopez, V.; Hwang, S.S. Response Surface Methodology to optimize the cement paste mix design: Time-dependent contribution of fly ash and nano-iron oxide as admixtures. *Mater. Des.* **2015**, *86*, 22–29. [[CrossRef](#)]
19. Lothenbach, B.; Scrivener, K.; Hooton, R.D. Supplementary cementitious materials. *Cem. Concr. Res.* **2011**, *41*, 1244–1256. [[CrossRef](#)]
20. Lee, J.; Mahendra, S.; Alvarez, P.J.J. Nanomaterials in the construction industry: A review of their applications and environmental health and safety considerations. *ACS Nano* **2010**, *4*, 3580–3590. [[CrossRef](#)]
21. Chiang, W.C.; Fratini, E.; Ridi, F.; Lim, S.H.; Yeh, Y.Q.; Baglioni, P.; Choi, S.M.; Jeng, U.S.; Chen, S.H. Microstructural changes of globules in calcium–silicate–hydrate gels with and without additives determined by small-angle neutron and X-ray scattering. *J. Colloid Interface Sci.* **2013**, *398*, 67–73. [[CrossRef](#)] [[PubMed](#)]

22. Hajilar, S.; Shafei, B. Nano-scale investigation of elastic properties of hydrated cement paste constituents using molecular dynamics simulations. *Comput. Mater. Sci.* **2015**, *101*, 216–226. [[CrossRef](#)]
23. Nochaiya, T.; Sekine, Y.; Choopun, S.; Chaipanich, A. Microstructure, characterizations, functionality and compressive strength of cement-based materials using zinc oxide nanoparticles as an additive. *J. Alloys Compd.* **2015**, *630*, 1–10. [[CrossRef](#)]
24. Le, H.T.; Müller, M.; Siewert, K.; Ludwig, H.M. The mix design for self-compacting high performance concrete containing various mineral admixtures. *Mater. Des.* **2015**, *72*, 51–62. [[CrossRef](#)]
25. Constantinides, G.; Ulm, F.J. The effect of two types of CSH on the elasticity of cement-based materials: Results from nanoindentation and micromechanical modeling. *Cem. Concr. Res.* **2004**, *34*, 67–80. [[CrossRef](#)]
26. Haoyang, S.; Jinyu, X.; Weibo, R. Experimental study on the dynamic compressive mechanical properties of concrete at elevated temperature. *Mater. Des.* **2014**, *56*, 579–588. [[CrossRef](#)]
27. Chen, S.J.; Duan, W.H.; Li, Z.J.; Sui, T.B. New approach for characterisation of mechanical properties of cement paste at micrometre scale. *Mater. Des.* **2015**, *87*, 992–995. [[CrossRef](#)]
28. Jiang, J.; Lu, Z.; Niu, Y.; Li, J.; Zhang, Y. Study on the preparation and properties of high-porosity foamed concretes based on ordinary Portland cement. *Mater. Des.* **2016**, *92*, 949–959. [[CrossRef](#)]
29. Gao, Y.; Hu, C.; Zhang, Y.; Li, Z.; Pan, J. Characterisation of the interfacial transition zone in mortars by nanoindentation and scanning electron microscope. *Mag. Concr. Res.* **2018**, *70*, 965–972. [[CrossRef](#)]
30. Constantinides, G.; Ulm, F.J.; van Vliet, K. On the use of nanoindentation for cementitious materials. *Mater. Struct.* **2003**, *36*, 191–196. [[CrossRef](#)]
31. Zadeh, V.Z.; Bobko, C.P. Nano-mechanical properties of internally cured kenaf fiber reinforced concrete using nanoindentation. *Cem. Concr. Compos.* **2014**, *52*, 9–17. [[CrossRef](#)]
32. Hintsala, E.D.; Hangen, U.; Stauffer, D.D. High-Throughput Nanoindentation for Statistical and Spatial Property Determination. *JOM* **2018**, *70*, 494–503. [[CrossRef](#)]
33. Kusne, A.G.; Gao, T.; Mehta, A.; Ke, L.; Nguyen, M.C.; Ho, K.M.; Takeuchi, I. On-the-fly machine-learning for high-throughput experiments: Search for rare-earth-free permanent magnets. *Sci. Rep.* **2014**, *4*, 6367. [[CrossRef](#)]
34. Mueller, T.; Kusne, A.G.; Ramprasad, R. Machine learning in materials science: Recent progress and emerging applications. *Rev. Comput. Chem.* **2016**, *29*, 186–273.
35. Chang, S.; Cohen, T.; Ostdiek, B. What is the machine learning? *Phys. Rev. D* **2018**, *97*, 056009. [[CrossRef](#)]
36. Brehmer, J.; Cranmer, K.; Louppe, G.; Pavez, J. Constraining Effective Field Theories with Machine Learning. *arXiv* **2018**, arXiv:1805.00013.
37. Nieves, J.; Santos, I.; Penya, Y.K.; Rojas, S.; Salazar, M.; Bringas, P.G. Mechanical properties prediction in high-precision foundry production. In Proceedings of the 2009 7th IEEE International Conference on Industrial Informatics, Cardiff, UK, 23–26 June 2009; pp. 31–36. [[CrossRef](#)]
38. Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkiathi, A.; Kim, C. Machine learning in materials informatics: Recent applications and prospects. *npj Comput. Mater.* **2017**, *3*, 54. [[CrossRef](#)]
39. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [[CrossRef](#)]
40. Awad, M.; Khanna, R. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Apress: New York, NY, USA, 2015.
41. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
42. Mannila, H. Data mining: Machine learning, statistics, and databases. In Proceedings of the 8th International Conference on Scientific and Statistical Data Base Management, Stockholm, Sweden, 18–20 June 1996; p. 2.
43. Koumoulos, E.P.; Dragatogiannis, D.A.; Charitidis, C.A. Nanomechanical properties and deformation mechanism in metals, oxides and alloys. In *Nanomechanical Analysis of High-Performance Materials*; Springer: Dordrecht, The Netherlands, 2014; pp. 123–152.
44. Koumoulos, E.P.; Charitidis, C.A.; Papageorgiou, D.P.; Papathanasiou, A.G.; Boudouvis, A.G. Nanomechanical and nanotribological properties of hydrophobic fluorocarbon dielectric coating on tetraethoxysilane for electrowetting applications. *Surf. Coat. Technol.* **2012**, *206*, 3823–3831. [[CrossRef](#)]
45. Oliver, W.C.; Pharr, G.M. An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *J. Mater. Res.* **1992**, *7*, 1564–1583. [[CrossRef](#)]
46. Sneddon, I.N. Boussinesq's problem for a rigid cone. *Math. Proc. Camb. Philos. Soc.* **2008**, *44*, 492–507. [[CrossRef](#)]

47. Huang, L.; Lu, J.; Troyon, M. Nanomechanical properties of nanostructured titanium prepared by SMAT. *Surf. Coat. Technol.* **2006**, *201*, 208–213. [[CrossRef](#)]
48. King, R.B. Elastic analysis of some punch problems for a layered medium. *Int. J. Solids Struct.* **1987**, *23*, 1657–1664. [[CrossRef](#)]
49. Bei, H.; George, E.P.; Hay, J.L.; Pharr, G.M. Influence of indenter tip geometry on elastic deformation during nanoindentation. *Phys. Rev. Lett.* **2005**, *95*, 045501. [[CrossRef](#)]
50. Constantinides, G.; Ulm, F.J. The nanogranular nature of C–S–H. *J. Mech. Phys. Solids* **2007**, *55*, 64–90. [[CrossRef](#)]
51. Zhu, W.; Hughes, J.J.; Bicanic, J.N.; Pearce, C.J. Nanoindentation mapping of mechanical properties of cement paste and natural rocks. *Mater. Charact.* **2017**, *58*, 1189–1198. [[CrossRef](#)]
52. Hu, C. Nanoindentation as a tool to measure and map mechanical properties of hardened cement pastes. *MRS Commun.* **2015**, *5*, 83–87. [[CrossRef](#)]
53. Howind, T.; Hughes, J.J.; Zhu, W.; Puertas, F.; Elizalde, S.G.; Hernandez, M.S.; Dolado, J.S. Mapping of mechanical properties of cement paste microstructures. In Proceedings of the 13th International Congress on the Chemistry of Cement, Madrid, Spain, 3–8 July 2011; p. 309.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).