VIBRO-ACOUSTIC TESTING AND MACHINE LEARNING FOR CONCRETE STRUCTURES DAMAGE DIAGNOSIS

By

Sarah Ann Miele

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

In

Civil Engineering

December 17th, 2022

Nashville, Tennessee

Approved:

Sankaran Mahadevan Ph.D.

Pranav M. Karve, Ph.D.

Douglas E. Adams, Ph.D.

D. Mitchell Wilkes, Ph.D.

Vivek Agarwal, Ph.D.

Dedicated in loving memory of Joseph W. Miele.

“Remember to look up at the stars and not down at your feet” -Stephan Hawking

ACKNOWLEGMENTS

I cannot express enough my gratitude to my research advisor and committee chair, Dr. Sankaran Mahadevan for his time, patience, feedback and motivation over the last few years. My academic journey would not be the same without the mentorship and support of my committee co-chair, Dr. Pranav Karve. An immense thank you also goes out to my other committee members, Dr. Douglas Adams, Dr. Mitchell Wilkes, and Dr. Vivek Agarwal for their suggestions and guidance.

I want to give special thanks to the incredible Civil and Environmental Engineering department staff, especially the Laboratory for Systems Integrity and Reliability (LASIR) and CEE Structures Laboratory staff at Vanderbilt University. I am so grateful for the experimental support of David Koester, Garrett Thorne, and Richard Teising. A large amount of my dissertation research included laboratory experiments and collaborations so I would also like to thank Dr. Eric Giannini, Patty Kyslinger and April Snyder from the R.J. Lee group, Dr. Jinying Zhu and Dr. Hongbin Sun at the University of Nebraska-Lincoln, and finally Nolan Hayes and the University of Tennessee, Knoxville. Thank you to Dr. Kyle Neal and Dr. Paromita Nath for their advice and for the work they did on the project before I began. I would also like to acknowledge my lab mates and other students who helped me collect data, Julia Finfrock, Gbandi Nikabou, and Yulin Guo. I am very grateful to have had some phenomenal mentors as professors, lab mates, and women in STEM to encourage and mentor me over the past years. The women I would like to thank are Dr. Erin DeCarlo, Dr. Ghina Absi, Dr. Hiba Baroud, and Dr. Lori Troxel; thank you for your guidance, support, and friendship over the past years. Also, this endeavor would not have been the same without my fellow graduate students at Vanderbilt.

I am privileged to thank my close friends, Mackenzie Whitman, Jordan Ash, Haley Adams,

Cole Brubaker, Gabrielle Piraino, and Colton Weaver, for always supporting me near or far. And finally, one of the biggest thankyous to my family; my parents, my grandparents, and all my sisters for showing me what it means to reach for your goals and their constant support and kindness. I am so grateful for my nephews, who never cease to bring a smile to my face.

Acknowledgement of Funding Support

This research was partly supported by funding from the U.S. Department of Energy, through the Office of Nuclear Energy's Light Water Reactor Sustainability (LWRS) Program. I would again like to thank project technical monitor Dr. Vivek Agarwal, LWRS program director Dr. Bruce Hallbert and Plant Modernization Pathway Lead Dr. Craig A. Primer for their technical and programmatic support.

I would also like to thank the Civil and Environmental Engineering Department at Vanderbilt University for supporting me financially through multiple teaching assistant opportunities over the past six years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

ANN	artificial neural network
ASR	alkali-silica reaction
BSE	backscatter electron
DIC	digital image correlation
DRI	damage rating index
EDS	energy-dispersive spectroscopy
FEA	finite element analysis
FEM	finite element model
FFT	fast Fourier transform
LS	linear spectrum
MCMC	Markov Chain Monte Carlo
ML	machine learning
MSE	mean squared error
NDE	non-destructive examination
NIRAS	nonlinear impact resonance-acoustic spectroscopy
NN	neural network
NPP	nuclear power plant
PIML	physics-informed machine learning
PMF	probability mass function
PSD	power spectral density
RBF	radial basis function (kernel)
SBSum	sideband sum
SHM	structural health monitoring
VAM	vibro-acoustic modulation

CHAPTER 1

Introduction

Structural health monitoring (SHM) is used to evaluate structural integrity and performance and guide decisions regarding maintenance, repair, and future operational limits of the structure. The diagnosis of damage in a structural system involves damage detection, localization, classification (of damage type), and/or quantification. In the past, both model-based and model-free methodologies have been used in SHM [1]. In general, SHM consists of four stages: detection, localization, assessment, and prediction [2]. The first stage, detection, indicates that damage exists in the structure. Localization next provides information on the location(s) of damage within the structure. In the next stage, assessment estimates the extent of current damage; this information is then used in the prediction stage by assessing the growth of damage and the system's strength and safety in the future. This work focuses on localization and crack extent estimation for alkali-silica reaction-induced internal cracks in concrete structures.

Alkali-silica reaction (ASR) is a critical degradation mechanism of concern for concrete structures (bridges, dams, primary/secondary containment in nuclear power plants (NPPs), reactor buildings in chemical plants, etc.). For example, as most existing nuclear power plants (NPPs) continue to operate beyond their initial license periods, the passive structures, systems, and components of NPPs as they continue to operate, suffer deterioration that influences structural integrity and performance. Monitoring the conditions of these elements of NPPs is essential for ensuring that their conditions meet performance and safety requirements over the entire expected plant lifespan.

ASR is a reaction in concrete between the alkali hydroxides (K^+ and Na^+) in the pore solution and

the reactive non-crystalline (amorphous) silica (S^{2+}) found in many common aggregates, given sufficient moisture, is called ASR. This reaction occurs over time and causes the expansion of the altered aggregate by the formation of a swelling gel of calcium silicate hydrate (C-S-H). Reactive silica is mainly provided by reactive aggregates and the alkalis by the cement clinker. Swelling from an ASR is a result of the relative volume increase between the product and reactant phases involved in the chemical reaction. First, the products expand in pores and microcracks of the cementitious matrix. Once this free expansion space is filled, the swelling is restrained, and the product phases exert local pressure on the surrounding concrete skeleton [3], [4]. With water presence, the ASR gel increases in volume and exerts an expansive pressure inside the material, causing spalling micro-to macro-cracks (due to nonhomogeneous swelling related to non-uniform moisture distribution). As a result, ASR reduces the stiffness and tensile strength of concrete, because these properties are particularly sensitive to micro-cracking. ASR also can cause serious cracking in concrete, resulting in critical structural problems that can even force the demolition of a particular structure. The serviceability of concrete structures includes the resistance to excessive deflections, as well as a host of other durability concerns that can shorten the service life of a structure. Large surface crack widths and deep penetration of open surface cracks promote ingress moisture and any dissolved aggressive agents, such as chlorides. Additionally, the loss of concrete stiffness and potential for reinforcement yield are concerns for concrete deflection capabilities.

ASR is a complex chemical phenomenon, the rate and extent of which depend on a number of material and environmental parameters, for which the interactions among parameters is not fully understood. This critical nature of ASR on premature concrete deterioration requires the quantitative assessment of ASR structural effects during service life (both in time and space). In particular, a combined experimental modeling investigation method is required to evaluate the impact of ASR on

the dimensional stability of concrete structures. Although ASR has been identified as a cause of deterioration of numerous concrete structures and research has yielded some understanding of the mechanism of the reaction, the structural effects of ASR and how to best assess the extent of damage to existing structures remain major topics of ongoing research. This is because the expansion and cracking patterns (the most obvious sign of distress) caused by ASR affect both the concrete and the reinforcing steel, but similar crack patterns can also be produced by other distress mechanisms (i.e., drying shrinkage and sulfate attack).

In the nuclear industry, a scoping study of ASR in concrete is performed to support future activities that include evaluating the effects of ASR on structural capacity. From a safety perspective, the remaining capacity of a structure exhibiting distress due to ASR is an important factor in operational and maintenance management decisions. This is a challenging task for various reasons. First, the extent of the degradation will vary throughout the element as a function of the moisture content and of the degree of restraint provided by steel reinforcement. Also, predicting the properties of the concrete using certain testing results taken from the structure may be difficult. The size of the defects caused by ASR may be large compared to a small structure such as the cylinder, resulting in anomalously-low tested strength, but the defects would be small compared to the larger structure (suggesting there may be sufficient capacity). In addition, there is no reliable nondestructive means of estimating the degree of the reaction in an existing concrete structure. To ensure satisfactory performance and safety over the expected lifespan of a structure, nondestructive evaluation (NDE) techniques are studied to diagnose the presence, location, and extent of ASR damage (or its effect, distributed micro/macro cracking) in concrete structures [5].

Damage localization and estimation are essential for structural health evaluation and management. In previous studies, NDE methods have focused only on the *detection* of ASR

damage in concrete structures. Model-based NDE methods utilize structural behavior models to simulate and/or learn the behavior of healthy and damaged structural systems. Model-free diagnosis methodologies aim to diagnose damage by post-processing/analyzing the sensor data without requiring a structural system behavior model. A simple and commonly used model-free method is a visual inspection of observed surface cracking and gel extrusion from damaged concrete components. This is a qualitative NDE method, and it can detect the ASR damage only after the damage progression is in an advanced stage; thus, it is not useful for early detection. Visual inspection does not monitor ASR-induced damage over time, thus limiting the possibility of repair. Quantitative NDE methods for ASR detection can be broadly classified as deformation measurement methods, vibration-based methods (acoustic methods), and thermal imaging methods (Figure 1) [5]. Deformation measurement techniques aim to measure the expansion of concrete caused due to the characteristic expansion of the ASR gel; this is a global measurement of ASR's effect. Acoustic methods detect damage by evaluating a damage-sensitive dynamic response characteristic of the concrete structure (e.g., resonance frequency shift). Thermal imaging methods aim to detect micro-cracking-induced thermal load path discontinuities as indicators of ASR damage. Figure 1 provides a simplified catalog of these methods.

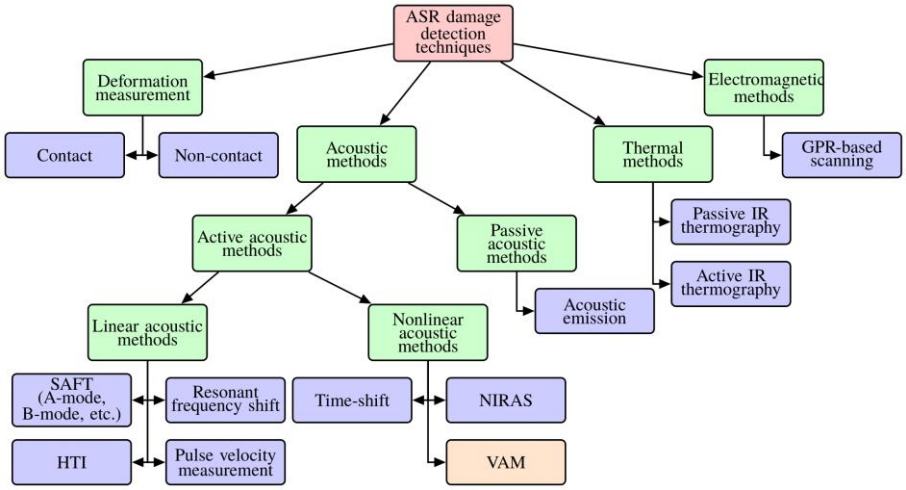


Figure 1. NDE techniques for detecting ASR in concrete structures.

Mechanical deformation measurements are useful for damage detection by measuring the expansion of concrete components but are not used to localize damage [6]. A concrete component's mechanical deformation can be measured using either a contact or a non-contact deformation measurement technique. Contact measurement techniques use calipers or an extensometer to measure the NDE specimen's deformation along multiple directions. Digital image correlation (DIC) is an optical, non-contact NDE technique capable of measuring the deformation, displacement, and strain of a structure [7]. Acoustic methods are the main focus of this work. Acoustic methods can be further classified into active and passive acoustic methods. Passive acoustic methods include acoustic emission to detect and classify cracks due to ASR [8]. Acoustic emission can be interpreted as transient stress waves produced by the sudden release of energy, such as crack formation or growth [9]. Active acoustic methods can be further classified as linear acoustic and nonlinear acoustic methods. In linear acoustic methods, the resonant frequency shift method is used to compute the concrete component's dynamic elastic modulus [10]. The pulse velocity method uses the experimentally-measured estimate of the compression stress wave speed

(using a pitch-catch set up) to detect ASR damage [11]. Together, the micro-cracking and loss of stiffness caused by ASR damage result in reducing the compression wave speed [11]–[13]. Khazanovich et al. [14] proposed using a Hilbert transform indicator (HTI) and an array of ultrasonic transducers to detect ASR damage in concrete slabs. In general, linear acoustic methods are less sensitive to ASR damage than nonlinear acoustic methods [15].

Nonlinear acoustic methods exploit the material (constitutive) or geometric nonlinearity in the damaged component to detect damage. Nonlinear impact resonance-acoustic spectroscopy (NIRAS) has been used to characterize ASR-induced cracks in concrete [15]–[17] by measuring the downward shift in the resonant frequency based on cracking-induced nonlinearity detection. It is a global vibrational response technique (i.e., it measures shifts in resonant frequency) and is better suited for small laboratory specimens than large structures. The material and/or geometric nonlinearity introduced by ASR damage promotes higher harmonics of the excitation frequency in the dynamic response spectrum [18]. The damage detection method based on harmonics generation evaluates the ratio of second harmonic amplitude to the square of the first harmonic amplitude, referred commonly in literature as the damage parameter β [19]–[21]. The material nonlinearity induced by ASR damage has been shown to increase the magnitude of the β. Laboratory experiments on composite plates with delamination have demonstrated that the higher-harmonic-based damage parameter (β) is less sensitive to micro-cracking-induced geometric nonlinearity than the wave-modulation-based damage parameter [22]. A technique that utilizes the shift in the arrival of high-frequency waves created by (impact-induced) low-frequency vibrations and nonlinear material behavior has also been shown to successfully detect the ASR damage in concrete specimen [23].

Thermal methods gauge the temperature distribution on a surface of the concrete component

using infrared (IR) thermography to estimate discontinuities (caused by cracking, spalling, delamination, etc.). Kobayashi and Banthia [24] combined induction heating with infrared thermography to detect corrosion in reinforced concrete. The technique is more effective with larger bar diameters and smaller cover depths. Bayesian methods have been used to optimize the image processing parameters and quantify uncertainty using infrared thermography [25]; however, the need for heat sources that can initiate a detectable temperature field in thick concrete structures may create practical difficulties in the field implementation of this method.

The vibro-acoustic modulation (VAM) method belongs to a family of NDE methods utilized as the primary indicator of damage to the spectral signature of a particular kind of nonlinear dynamic behavior occurring at a breathing, weakly bonded interface in a vibrating component. The nonlinear dynamic interactions at a crack or delamination interface can be induced using either a bi-harmonic excitation. The method depends on detecting peaks around the high-frequency in the collected dynamic response's spectra, referred to as sidebands. The bi-harmonic excitation induces forced vibrations for both the modulator (lower frequency) and modulated (higher frequency) waves. This method has shown particular promise in detecting ASR-related degradation (or damage) in concrete [16], [26], as well as cracks and debonding in metals and composites [19], [20], [27]–[33]. In past research, the VAM technique's effectiveness in localizing (in 3-dimensions) ASR damage in concrete specimens of varying sizes has yet to be studied. Furthermore, a VAM-based, probabilistic damage localization methodology has not been investigated. This research examines the application of VAM to multiple thick concrete specimens and determine the ideal damage metric for VAM-based damage localization in thick composites such as concrete. Main contributions of this research include:

- Localization of ASR-related damage in concrete specimens using VAM.

- The use of physics informed machine learning (PIML) to classify binary damage state at a sensor.
- Bayesian damage diagnosis for a ML prediction model built to give a posterior for damage depth.
- Investigation of transfer learning to combine PIML training data of multiple fidelities.
- Localization of damage in plain and reinforced concrete specimens with dispersed and unknown ASR damage inducement.

1.1. Research Objectives

The overall goal of the conducted research was to localize damage in concrete specimens. To do this, we first examined the application of VAM to a plain concrete specimen with known areas of ASR damage inducement. Next, we developed machine learning models using finite element analysis (FEA) data for training to build a better model for damage detection and localization in the plain concrete specimen. Since three-dimensional (3D) FEA models for mechanical vibrations are computationally expensive, we trained the diagnosis models with a large amount of two-dimensional (2D) FEA data. Next, we explored improving this model by using multi-fidelity physics-informed modeling to combine the 2D FEA data with a smaller amount of 3D FEA data and evaluated the resulting machine learning-based diagnosis model performance and the computational cost of simulating the training data. After this, the model was verified for the plain concrete specimen with known localized inducement of ASR damage. Lastly, we implemented these techniques to validate similarly constructed diagnosis models for different sized plain and reinforced concrete specimens with dispersed ASR damage inducement throughout.

The research is divided into four objectives:

1. Damage diagnosis using vibro-acoustic modulation for a plain concrete specimen with damage inducement at known locations;
2. Damage localization in a plain concrete specimen using 2D physics model-informed machine learning;
3. Multi-fidelity physics informed machine learning for damage localization using 2D and 3D finite element models; and
4. Damage localization in plain and reinforced concrete specimens with distributed damage inducement (unknown damage locations).

This dissertation is split into 6 chapters. Chapters 2-5 relate to the four research objectives. Chapter 2 develops the VAM testing procedure for ASR-damage localization in concrete specimens. This study is one of the first to localize this damage in concrete specimens. We examine the fusion of damage-metric localization results from multiple VAM tests by using both an averaging and a Bayesian technique. Both of these damage localization approaches are validated for a plain concrete specimen with localized ASR-induced damage. Chapter 3 extends this methodology by training a damage prediction model using machine learning techniques using input-output data from 2D FE simulations. The resulting artificial neural network (ANN) models provide a probabilistic value for the presence or absence of damage at each sensor location (lacking in the averaging approach from Chapter 2), do not depend on a damage threshold provided by an expert analyst (necessary for the Bayesian fusion approach from Chapter 2), and provide a means for depth prediction given a crack size and location in the X-Y plane. In Chapter 4 we explore improving the models built in Chapter 3. We train diagnosis models using 2D simulations, 3D simulations, and pursues a multi-fidelity approach that uses transfer learning to combine the 2D

FE training data with select 3D FE simulations for training. From this study, we discover that our computational resources are best expended for the creation of additional 2D FE simulations. In Chapter 5 we examine the fourth and final objective of applying the aforementioned localization techniques for specimens with dispersed ASR damage inducement (i.e., with unknown damage locations), with a different geometry (from the specimen used for validation in Chapters 2, 3 and 4) and rebar reinforcement. The machine learning models for VAM-based damage diagnosis are validated using experimental specimens that were destructively tested to determine the presence and extent of ASR-induced damage. The extent of the ASR-induced damage is quantified using a petrographic method, damage rating index (DRI). We find the models trained in this approach successful for ASR-induced damage localization of the specimens. Finally, in Chapter 6, we summarize the accomplishments of this study and discuss future research needs, including automation of the proposed methodology.

CHAPTER 2

Damage Diagnosis using Vibro-Acoustic Modulation for a Plain Concrete Specimen with Localized Damage Inducement

2.1. Motivation

Vibro-acoustic modulation (VAM), also known as nonlinear wave modulation spectroscopy, is a nondestructive evaluation (NDE) technique that relies on detecting nonlinear structural behavior's dynamic signature as the primary indicator of damage. Most of the previous work on VAM-based diagnosis has focused on *detecting* damage based on sidebands in the spectrum of the structure's dynamic response. Singh et al. [34] showed that VAM testing could be used for damage *localization* or damage mapping. They hypothesized that the effect of (geometric or material) nonlinearities is pronounced near the flaw's location. The relative magnitude of a sidebands-based damage index may enable localization of the flaw. Suppose the spatial distribution showing the variation of the damage index is obtained using a sensor grid. In that case, the damage is located in the neighborhood of sensors exhibiting higher magnitudes of the damage index. They tested their hypothesis using numerical simulations of VAM in delaminated composite plates. They studied damage indices based on various characteristics of the plate's dynamic response spectrum, establishing the feasibility of VAM-based damage localization. Pieczonka et al. [22] performed sidebands-based damage imaging to localize impact damage in a composite plate in a laboratory set-up. They used laser scanning vibrometry to collect the dynamic response on the surface of the composite plate. Their experimental damage mapping results show that sideband-based damage mapping performs better than the previously known higher-harmonics-based damage mapping.

Thus, the damage mapping scheme's utility has been studied for homogeneous, anisotropic, thin composite plates by performing numerical or laboratory experiments. However, the applicability of VAM-based damage mapping to detect and localize cracks in structural concrete components has not been investigated. In particular, thick, heterogeneous structural concrete components present significant challenges for VAM test set-up, data analytics, and damage mapping

The investigation of VAM for damage localization in concrete structures is the first objective of this research. The focus of past experimental studies on VAM-based damage localization was related to impact damage in thin composite plates [34]. Application of the VAM method for thick concrete specimens entails ensuring that the entire test specimen receives sufficiently strong vibration despite the high material and geometric attenuation and a (relatively) coarse grid of accelerometers to localize damage. We performed VAM tests in the laboratory using different test parameters. We applied a Bayesian data fusion algorithm to assimilate the diagnostic information obtained from multiple VAM tests and constructed a damage probability map. We used a cement slab with pockets of reactive aggregate (at known locations) and cured under aggressive conditions to promote accelerated ASR). We implemented damage mapping and data fusion algorithms on the VAM test data obtained from the laboratory experiments.

2.2. Relevant Background

In the VAM technique, the structural component of interest is excited simultaneously using two signals of specific frequencies. The dynamic response is measured at various locations using acoustic sensors (accelerometers). The low-frequency input is termed the “pump,” and the high-frequency input is termed the “probe” [32]. It is assumed that ASR-induced cracks introduce nonlinear structural behavior due to variable contact area at the surfaces of cracks. If the structure

undergoes damped, linear vibration, then the frequency spectrum of the steady state response shows peaks at the two forcing frequencies (f_{probe},f_{pump}) (Figure 2(a), (b)). The natural frequencies of the structure participate in the transient phase, but eventually die out due to damping. For a structure undergoing damped, nonlinear vibration, the frequency spectrum of the steady state response (Figure 2(c), (d)) shows additional peaks at higher harmonics of the pumping frequency ($i * f_{pump}, i = 1,2,3$ ...), and at sideband frequencies $f_{probe} \pm f_{pump}$ [18]. The peaks around the probe frequency in the frequency spectra of the measured response (referred to as sidebands, see Figure 2(d)) appear due to modulation of the probing frequency by the pumping frequency. For illustration, the linear spectrum (LS) of the recorded acceleration at a sensor is shown in Figure 2. The spectral magnitude at sideband frequencies (peaks in the LS at $f_{probe} \pm f_{pump}$) are denoted by AmpS1 and AmpS2.

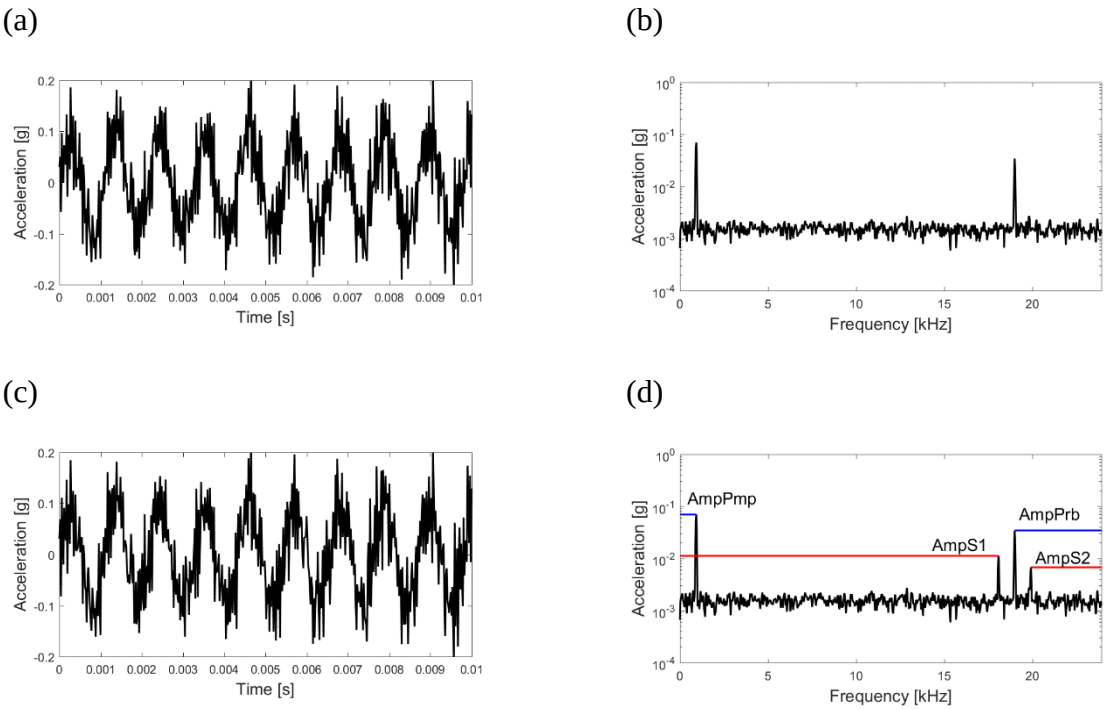


Figure 2: Time series data and linear spectrum (LS) plots for two different cases. (a): time history of acceleration for linear response (undamaged case); (b) LS of acceleration for linear response (undamaged case); (c) time history of acceleration for the nonlinear response (damaged case) with frequency modulation; (d) LS of acceleration for the nonlinear response (damaged case) with frequency modulation.

In this research, we discuss the mapping of ASR-induced cracks based on a damage parameter that is calculated using the sum of spectral amplitudes (AmpS1 and AmpS2) at sideband frequencies in the linear spectra of the recorded dynamic response (acceleration) can be defined as

$$SBSum = AmpS1 + AmpS2. \quad (1)$$

The use of VAM-based damage localization for thick concrete structures faces several challenges. The operating environment as well as test conditions for real-world concrete structures may introduce significant measurement noise in the test data. The damage diagnosis methodology should be able to perform well in the presence of measurement noise. Two data processing schemes that help alleviate the problem of noisy data in VAM tests are briefly described in subsection 2.3.1.

A detailed analysis of these methods for various signal-to-noise ratios can be found in Karve and Mahadevan [35]. In thick elastic concrete slabs with material damping, only the region in the vicinity of pump/probe actuators may receive sufficiently strong vibration to induce detectable SBSum values. Furthermore, the magnitude of SBSum may depend (among other things) on the amplitude of the wave motion experienced at the damaged (cracked) location. Thus, false negative (and false positive) results are expected when the damage location/size is not known and only a single VAM test is performed. Our numerical experiments on elastic slabs have shown that the sensitivity of the VAM test increases when the pump/probe are located near the damaged zone, and that VAM tests with a given probing frequency (wavelength) are sensitive to damage with characteristic size of about 70% of the probing wavelength [35]. In the real-world application of VAM tests, where the damage size and location are not known, there are advantages to performing VAM tests with multiple pump/probe locations and multiple probing frequencies (wavelengths). Furthermore, flaws of different sizes may show better resolution at different probing frequencies. Thus, different locations, frequencies, and amplitudes of the pumping and probing excitations provide different VAM test results, with varying information about the state of damage in the vicinity of each sensor. Consequently, a methodology is needed to fuse the results of multiple tests and provide the damage estimates at different sensor locations. Fusion of diagnostic information obtained from VAM tests with different test parameters can help alleviate false negative results due to inactivation of amplitude-dependent non-linear mechanisms associated with damage (the inactivation can occur for a few combinations of excitation frequencies, amplitudes, and pump/probe locations, and these combinations cannot be determined before conducting a VAM test as the damage state is unknown).

2.3. Methodology

The proposed methodology thus consists of four elements: a) VAM test data processing, b) damage mapping and localization using VAM test data, c) fusion of information obtained from multiple VAM tests with different test parameters, and d) validation of VAM-based damage localization using petrographic and chemical studies. The following subsections describe each of these four elements.

2.3.1. VAM data processing

Various practical aspects of the VAM methodology encountered during field or laboratory testing of concrete structures have not been studied previously. One of the key challenges in data processing for field or laboratory tests is distinguishing physically meaningful (nonlinearity-induced) sidebands (peaks) from noise-induced peaks at the specified frequencies ($f_{probe} \pm f_{pump}$) in the linear spectrum. For experimentally collected data, the linear spectra in neighborhood of the sidebands may not show near-zero amplitudes for all sensors, and an automated sideband detection algorithm might identify the *ambient noise* values at ($f_{probe} \pm f_{pump}$) as sidebands. This means the ordinates of LS at ($f_{probe} \pm f_{pump}$) need to be adjusted for the baseline value of LS in the neighborhood of sideband locations. In this study, we subtract the mean of the ambient noise in the LS ordinates in the neighborhood of the sidebands from the values calculated from the LS (Figure 3 (a)). The regions used to calculate the mean of LS amplitudes are shown by red boxes in Figure 3 (a). The width of the region is equal to the pump frequency. Thus, the peaks in the LS that are of the order of the mean LS amplitudes in the highlighted region are not counted as significantly high sideband values during data processing.

Furthermore, an automated SBSum calculation process may be misled by considering

ordinates at sideband frequencies that are not peaks (and hence do not indicate nonlinear structural behavior). To ensure that the values being used to calculate SBSum are physically meaningful sidebands, a method for finding local peaks in the data can be implemented. In this method, whether the sideband value for a given frequency is a local peak is determined first. Next, whether the sideband was a maximum value within a window (Figure 3 (b)) centered at the sideband frequency of interest is checked. As a result of the nonlinearity at the crack surface, a second pair of sidebands may be seen in the linear spectrum of the acceleration at frequency, $f = f_{probe} \pm 2 * f_{pump}$ [18]. The pump frequency can thus be safely used as the window size that does not interfere with the second pair of sidebands. In this work, a 1000–Hz-wide window (approximately equal to the pump frequency in our laboratory experiments) is used. The ordinate value at a given frequency is selected as a sideband (AmpS1 or AmpS2) only if it satisfies both the above-mentioned conditions. The utility of the aforementioned data processing techniques has been discussed by Karve and Mahadevan [35] in the context of numerical simulations of VAM tests. Here, these techniques are put to use for processing noisy data obtained from physical (laboratory) experiments.

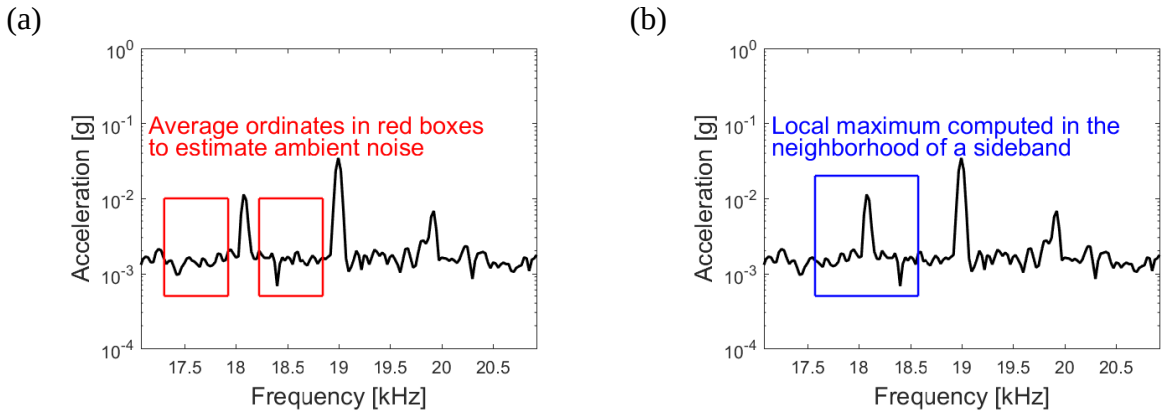


Figure 3: Linear Spectrum plot for a damaged structure: (a) the red boxes represent the window where the average of the linear spectrum was calculated; (b) the blue box shows the window used to identify local peaks around a sideband.

2.3.2. Damage mapping

In this section, a damage index (using the SBSum metric) that captures the essence of the key idea behind the damage localization method is defined. In VAM-based damage localization, sensors showing higher SBSum values (compared to other sensors in the same test) are assumed to be the sensors located near the damage zone. Hence, a damage index that classifies a given sensor as the one indicating or not indicating damage is defined here. Specifically, the damage index at sensor "*i*", D_i, can be computed as:

$$D_i = 1 \quad \text{if } SBSum_i \geq \frac{1}{N_{sens}} \sum_{i=1}^{N_{sens}} SBSum_i \,,$$

$$= 0, \text{otherwise}, \tag{2}$$

where N_{sens} denotes the total number for sensors used in a VAM test. If the SBsum value at a given sensor is higher than the mean SBsum value (for all sensors) for that VAM test, then the sensor is considered to be indicating damage [35]. The observed damage index represents, in essence, the key concept behind VAM-based damage localization. That is, the data informs that damage is *present* at/near a sensor if the SBSum value at that sensor is *relatively higher* than the SBSum values at other sensors. We remark that a numerical analysis of performance of the damage localization method for elastic slabs using the damage index defined in equation (2) has been discussed in Karve and Mahadevan [35].

2.3.3. VAM data fusion using Bayesian inference

For VAM testing of concrete structural components, it is desirable to conduct VAM tests for multiple values of the test parameters. In this section, a methodology for integrating VAM test data obtained from multiple tests to obtain a damage probability map is discussed. The proposed

method uses Bayesian updating to infer the regions that consistently show (relatively) high SBSum values, i.e., the regions that are most likely to be near a damaged zone. The Bayesian approach pursues a formal representation of the state of knowledge about the values of unknown parameters of interest through the use of probability distribution functions. The updated knowledge (i.e., the posterior distribution) about a parameter is obtained by combining prior knowledge (based on intuition, experience, model prediction, prior data, etc.) and observed data. The observed data is included in the (Bayesian) inference algorithm by computing the likelihood of observing the data for a given value of the parameter. In order to use Bayesian inference for VAM-based damage diagnosis, we need: a) definition of experimentally observed data and the associated measurement uncertainty, b) definition of the damage parameter that gives the state of damage in the neighborhood of each sensor, and the associated uncertainty due to lack of knowledge about its value (i.e., the prior distribution), and c) a method (typically, a computational model of the experimental procedure) to compute the likelihood of observing a given value of data for a given value of the damage parameter. These aspects of the Bayesian updating for VAM tests are discussed next.

For the Bayesian updating algorithm, the damage index D_i is the *data* obtained from experiments, as defined in equation (2). The observed damage index is a discrete random variable with two possible values ($D_i = 1$ or $D_i = 0$). If damage is indicated at a sensor, then $D_i = 1$, else $D_i = 0$. Similarly, the actual damage parameter at sensor "*i*" is defined as θ_i. θ_i is a discrete random variable that denotes whether damage is actually present ($\theta_i = 1$), or absent ($\theta_i = 0$) at/near the *i*-th sensor. With the above definitions, Baye's rule for the damage index θ_i at each sensor can be written as:

$$P(\theta_i|D_i) \propto P(D_i|\theta_i) * P(\theta_i), \tag{3}$$

where $P(\theta_i|D_i)$ is the posterior probability mass function (PMF) of variable θ_i, $P(\theta_i)$ is the prior PMF of θ_i, and $P(D_i|\theta_i)$ is the likelihood of observing the data ($D_i = 1$ or $D_i = 0$) for a given value of the unknown parameter ($\theta_i = 1$ or $\theta_i = 0$). The Bayesian update equations for a sensor where damage is indicated and for a sensor where damage is not indicated are:

$$P(\theta_i = 1|D_i = 1) = \frac{P(D_i = 1|\theta_i = 1) * P(\theta_i = 1)}{P(D_i = 1|\theta_i = 1) * P(\theta_i = 1) + P(D_i = 1|\theta_i = 0) * P(\theta_i = 0)}, \quad (4)$$

$$P(\theta_i = 1|D_i = 0) = \frac{P(D_i = 0|\theta_i = 1) * P(\theta_i = 1)}{P(D_i = 0|\theta_i = 1) * P(\theta_i = 1) + P(D_i = 0|\theta_i = 0) * P(\theta_i = 0)}. \quad (5)$$

Note that in the ideal case, $P(D_i = 1|\theta_i = 1) = 1$, and $P(D_i = 0|\theta_i = 1) = 0$, i.e., the test does not show any false positive or false negative results. However, in actual physical experiments, the VAM test may exhibit false positives as well as false negatives. If, for example, an antinode of structural vibration lies at/near a sensor location, the SBSum value at that sensor may be small even if it is located near the damaged zone. On the other hand, the wave component corresponding to sideband frequencies ($f_{probe} \pm f_{pump}$) may travel away from the damage and excite a sensor located away from the damage zone. Furthermore, measurement noise may create false positive and false negative data (in spite of the noise treatment strategies discussed above). In those cases, to compute the likelihood function, we must evaluate the sensitivity (true positive rate) and specificity (true negative rate) of VAM tests for thick elastic media containing breathing cracks. This can be done using numerical simulation of the test procedure.

In a numerical simulation of the VAM test, the value of the damage parameter (θ_i) is known and the value of the observed damage index can be obtained using the computed dynamic response (displacement time history). For the experiments reported in this article, numerical simulation of VAM tests implies modeling the nonlinear wave propagation in concrete slabs/blocks containing

cracks for multiple crack locations, pump and probe amplitudes, and probe frequencies. (The pump frequency is fixed at the first resonance frequency of the specimen). Modeling wave propagation in a heterogeneous medium like concrete (with cement paste and randomly distributed aggregate) is a computationally challenging task. Therefore, the material heterogeneity is ignored, and wave propagation in a homogeneous elastic specimen with smeared concrete-like properties [36] is simulated. Even with this simplification, simulating nonlinear wave propagation in three-dimensional media is computationally expensive. Geometric nonlinearity at the crack interface necessitates the use of an implicit time integration scheme for the simulations, which increases the computational burden significantly. To overcome this difficulty, wave propagation in a two-dimensional media is modeled, under plane strain assumptions. Our previous work [35] indicates that this model is suitable for computing the sensitivity and specificity of VAM tests with different test parameters as well as measurement noise. We remark that typically, the likelihood of observing the data for a given value of the parameter of interest is computed for each sensor using a numerical model of the underlying physical phenomenon. Since a two-dimensional model is to be used for computing the approximate likelihood for sensors installed on the (three-dimensional) slab, sensor-specific likelihood computation is not possible. Instead, the average values of true positive rate (sensitivity) and true negative rate (specificity) observed in our numerical simulations with different damage locations and VAM test parameters can be used as approximate likelihood values. For the first test, a uniform prior ($P(\theta_i = 1) = P(\theta_i = 0) = 0.5$) can be assumed, and the damage parameter at each sensor location can be updated using the experimental data for the first VAM laboratory test. For the subsequent tests, the posterior from the previous test can be used as the prior and obtain the posterior using the equations (2)-(4). In this manner, the observed damage index, damage parameter definitions (equation (1)), and an approximate likelihood function can

be utilized to perform Bayesian fusion of information obtained from VAM tests with different test parameters to arrive at the ASR damage probability maps for concrete test specimens.

2.3.4. Validation

The proposed damage mapping methodology is validated in multiple ways by performing dynamic, chemical and petrographic studies on a cement slab with pockets of reactive aggregate at specified locations. The slab is cast and cured in an aggressive environment conducive to ASR for an extended period of time. The onset of ASR in the slab can be ascertained using deformation measurements and NIRAS tests. The NIRAS tests indicate excitation-amplitude-dependent shifts in the resonant frequency of the slab when there is nonlinearity due to (ASR) damage. Further evidence of ASR initiation and progress can be collected by performing inductively coupled plasma-optical emission spectroscopy on the effluents seeping out of the slab. These preliminary tests establish the presence of ASR due to the combination of reactive aggregate, alkali loading, and aggressive curing environment.

Next, in order to accomplish and validate the localization, VAM tests are first performed on the specimen using the methodology discussed in sections 2.3.1 to 2.3.3, to obtain ASR damage maps. The damage maps highlight the high-damage-probability areas of the slab. Next, in order to validate the damage maps, core samples are taken from the locations where aggregates were placed. Petrographic examination of the cores is performed to establish the presence of ASR. Specifically, three tests are performed to confirm ASR damage in the core samples: a) optical imaging to confirm the presence of micro cracks in the aggregates, b) scanning electron microscopy (SEM) with backscatter electron (BSE) imaging to visualize the cracking, and c) SEM with energy dispersive spectroscopy (EDS) to confirm the presence of elements that are typically

present in the ASR (silicon, calcium, potassium, etc.). The results of the petrographic examination provide information regarding the presence of ASR byproducts and ASR-induced micro/macro cracking, and validate the damage localization results of the VAM methodology.

2.4. Laboratory Experiment

The cement slab specimen with localized inducement of ASR damage was cast and cured at Vanderbilt University. The specimen had known locations of reactive aggregate (possible sites of ASR initiation) (EXP-A). We first discuss the details of casting and curing the test specimen.

2.4.1. Cement slab with localized inducement of ASR damage

A plain cement slab specimen (EXP-A) of dimensions of 60.96 × 60.96 × 15.24 cm^3 (24 x 24 x 6 in^3) was cast and cured at the Laboratory for Systems Integrity and Reliability (LASIR) at Vanderbilt University. The water-to-cementitious material ratio of 0.4 was used for casting the slab to achieve adequate workability in the cement paste. Four types of aggregate, including three coarse aggregates known to be susceptible to alkali-silica reaction, as well as pure silica powder, were placed at known locations in (cylindrical) pockets of approximately 10 cm (4 in.) diameter in the four quadrants of the specimen, as shown in Figure 4. The aggregates were placed in pockets instead of being dispersed throughout the specimen so that the possible locations of ASR are known. The known locations of the aggregate pockets were used to obtain core samples that provided validation data for the VAM-based localization of ASR damage. The four types of aggregates are: (a) pure amorphous silica, (b) coarse aggregate from Maine (donated by The University of Alabama), (c) coarse aggregate from New Mexico (donated by The University of Alabama), (d) coarse aggregate from Ontario, Canada (donated by the Ontario Ministry of

Transportation). A boosting agent, sodium hydroxide (NaOH) was added to the mix water in the cement to boost the alkali loading to 5.25 kg/m^3 to accelerate ASR. EXP-A was cured in an environmental chamber maintained at 60 °C and ~100% relative humidity. After 10 months of curing, the onset and progress of ASR were confirmed by performing preliminary ASR detection tests (section 2.5.2). EXP-A was cured further for six additional months. After confirming the presence of ASR, VAM tests were performed on the specimen. The details of these tests are given in the next section.

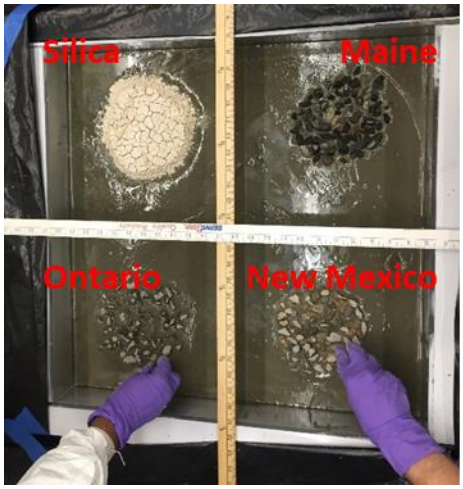


Figure 4: Pockets of aggregate in the specimen during casting.

2.4.2. VAM test framework

In this section, we discuss the setup used for conducting VAM tests on EXP-A. In VAM tests, we deliver the pump and probe excitations using piezo-stack actuators. We vary the locations of these actuators, as well as the frequencies at which they operate (i.e., the pump and probe frequencies). We measure the acceleration time history at various locations on the surface of the test specimen using a finite number of accelerometers. The relative magnitudes of sidebands at

various accelerometer locations are used to map the damage (ASR-induced cracks) in the component. We discuss the choice of various test parameters below.

2.4.2.1. Probe frequency

It has been suggested in the literature that the probe frequency should be 10 to 20 times the pump frequency [34]. This ensures that VAM-induced sidebands do not interfere with the first few natural frequencies, as well as higher harmonics of the pumping frequency. However, attenuation and test equipment (sampling frequency) capabilities may limit the highest usable probing frequency. Considering these factors, we used probe frequencies ranging from 10 kHz to 21 kHz (incremented by 1000 Hz) in our experiments, given that the pump frequency is 920 Hz. The probing frequency is related to the smallest detectable flaw size [35]. Typically, higher frequencies provide better resolution, but they also suffer from higher attenuation. In this study, we use a range of probing frequencies to perform the VAM tests on EXP-A.

2.4.2.2. Pump and probe excitation amplitudes

In VAM tests, the probe excitation amplitude should be much lower than the pump excitation amplitude [34]. An effective VAM is achieved when a low-amplitude probing wave rides the high-amplitude pumping wave. However, for a thick concrete specimen, if a low-amplitude probe excitation is used as an input, then the probing excitation energy at the damage site may be negligible due to high material attenuation experienced at the higher (probe) frequency. Thus, we experimented with pump-to-probe excitation amplitude ratios of 1, 2, 5, and 10. In our experiments, the pump and probe excitations were delivered using piezo-stack actuators. The ratio of pump and probe amplitudes was controlled by varying the voltage input given to the actuator. That is, we changed the peak-to-peak voltage output for the function generator used to excite the probing piezo stack actuator to experiment with different ratios of pump-to-probe amplitude. We

used four amplitudes for the peak-to-peak voltage difference provided by the probing signal generator: 500 mV, 250 mV, 100 mV, and 50 mV. The output voltage difference for the pumping signal generator was maintained at 500 mV. The pumping and probing signals were amplified using an amplifier and sent to the piezo-stack actuator. A constant amplification factor (+28 dB) was maintained for all tests, corresponding to the maximum capacity of the amplifier.

2.4.2.3. Pumping and probing excitation actuators

As mentioned in Section 2.3, for VAM-base damage localization in thick concrete specimens, material and geometric attenuation can pose a significant challenge. In addition, practical considerations such as equipment limitations or safety considerations limit the maximum pump and probe excitation amplitudes that can be used in a VAM test. In this work, we address these issues by applying the pump and probe excitations at multiple locations (referred to as configurations) on the accessible surface of EXP-A. The pump and probe actuators are placed close to each other to ensure that the region surrounding the actuators receives sufficient dual-frequency excitation. If sufficiently strong actuators are available, then the pump actuator could be placed at a fixed location near the centroid of the test surface, and the probe actuator could be moved to test different regions of a large concrete structural component.

2.4.2.4. Specimen instrumentation and data processing

To test the damage mapping capabilities of the VAM-based damage localization methodology, EXP-A was divided into two halves, and VAM tests were performed with a grid of 21 accelerometers (PCB Piezotronics accelerometers, model numbers PCB352C68 and PCB352C65) for each half. The accelerometers were calibrated before testing to ensure accuracy. Calibration was performed by applying excitations directly to the accelerometer using a handheld shaker and ensuring that the accelerometer reading was the same as the shaker output. The

sensitivity of all accelerometers, obtained by testing in the calibration shaker, was used in the data acquisition program to account for minor changes in the same. The accelerometers were attached to the specimen using wax. The pump and the probe actuators (PCB Piezotronics piezoelectric disk actuators, model number PCB712A02) were placed at the center of each quadrant and the center of the specimen (Figure 5). In Figure 5 (a), the blue and green circles represent the pump and probe actuators respectively. The pump and probe locations form the test configurations and are numbered 1 to 5 in Figure 5 (a). For example, in configuration 5, shown in Figure 5 (b), the pump and probe are located at the centroid of the top surface of the slab and the slab is instrumented with 21 accelerometers on the top half of the slab. For each configuration, the data is recorded by first populating the top half, and then populating the bottom half with accelerometers. At each accelerometer, acceleration response is measured five times, each (acceleration response) record was 0.2 s long, and the sampling frequency is 128 kHz. As the recorded acceleration time histories contain measurement noise, Welch's method [37] and a Nuttall-defined, 2048-point, four-term, symmetric, Blackman-Harris window [38] is used to compute the linear spectrum (LS) for each recorded acceleration time history. The LS for the five measurements is averaged and the values of AmpS1 and AmpS2 (Figure 6) were used to calculate the SBSum.

(a) (b)

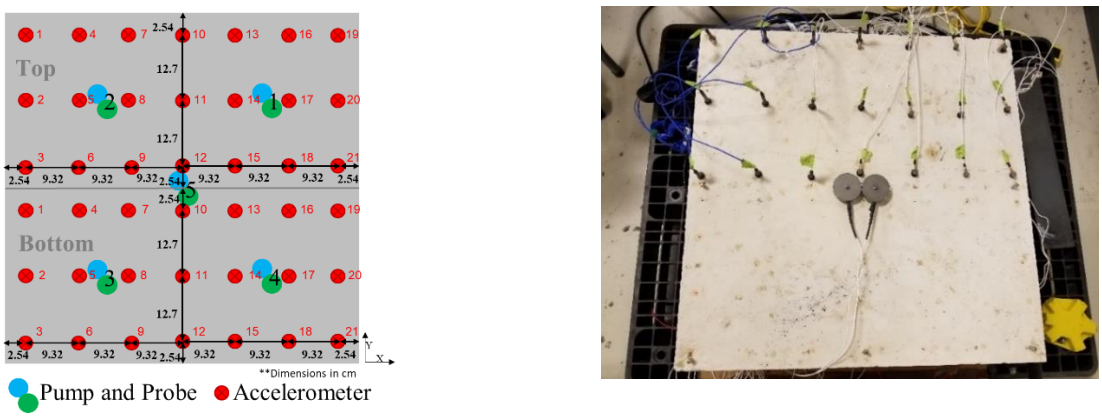


Figure 5: (a) Pump, probe, and accelerometer locations for EXP-A, (b) accelerometers on the top half of EXP-A (configuration 5).

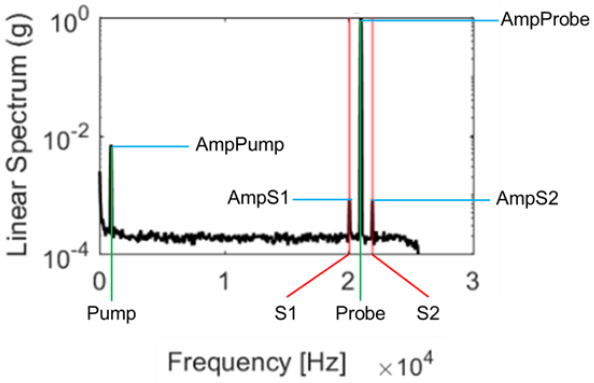


Figure 6. LS Plot for an accelerometer labeled to show the values used in calculating SBSum.

2.5. Results

In this section, we first discuss numerical simulation results used to compute the approximate likelihood function for Bayesian data fusion. Next, we discuss the results of preliminary tests performed to confirm the presence of ASR, petrographic examination, and VAM-based damage localization.

2.5.1. Likelihood computation for Bayesian data fusion

In this section, we discuss the details of the numerical modeling used to compute the model-based likelihood function. As discussed in Section 2.3.3, we use a two-dimensional domain with internal breathing cracks as the numerical model for computing the likelihood function. The domain geometry, the crack locations, and the pump and probe locations used in our numerical simulations are shown in Figure 7. The computational domain is 60.96 cm wide and 15.24 cm thick. It contains a crack of length 12.7 cm at mid-thickness (7.62 cm from the top surface). We use a commercial finite element program (Abaqus [39]) to perform numerical simulation. We discretize the domain using a structured mesh of 8-noded finite elements. We employ an implicit scheme for time integration of the governing equations and ensure that at least 10 computational nodes are available over the smallest wavelength. We model the interaction at the crack interface using a hard, frictionless contact condition. The material properties used in our simulations are given in Table 1 [36].

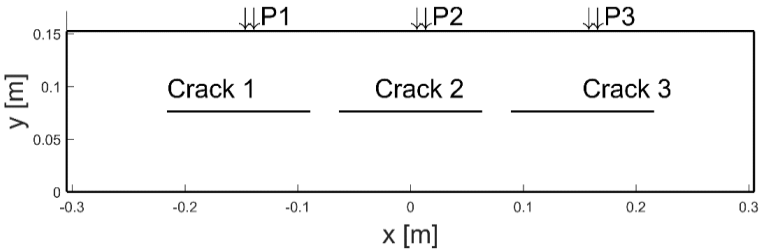


Figure 7: Two-dimensional domain showing crack locations and pump/probe locations used for likelihood computation.

Table 1: Material properties used in numerical simulation for likelihood computation.

Material property	Value
Young's modulus (E)	27 GPa
Density (ρ)	2400 kg/m^3
Poisson's ratio (ν)	0.15
Mass proportional Rayleigh damping parameter (a)	2120.04
Stiffness proportional Rayleigh damping parameter (b)	1.787×10^{-7}

We simulate the nonlinear wave propagation for different combinations of pump/probe locations (labeled as P1, P2, and P3 in Figure 7), crack locations, probe frequency, and probe/pump amplitude ratios (Amp_{probe} / Amp_{pump}). We apply point loads in the (negative) Y direction at two neighboring finite element nodes to simulate the dual frequency excitation provided by two actuators. We record the time history of displacements on the top surface at all computational nodes and compute the SBSum metric and the damage index for all nodes. In a given simulation, the nodes located directly above the crack are the nodes for which the damage parameter value θ_i = 1, otherwise θ_i= 0. Thus, we know the "truth" about the damage parameter and we compute the damage index (D_i) at each computational node. In this manner, we can arrive at false positives and false negative results at each node for each numerical VAM test. The matrix of VAM test parameters used in our numerical simulations are given in Table 2.

Table 2: VAM test parameters used for likelihood computation.

Variable	Values
Crack location	Crack 1, Crack 2, Crack 3 (Figure 7)
Probe frequency	10000Hz, 12000Hz, 14000Hz, 16000Hz, 18000Hz, 20000Hz
Amp_{probe} / Amp_{pump}	1, 0.5, 0.2, 0.1
Pump/Probe location	P1, P2, P3 (Figure 7)

In this manner, we compute the D_i values for known θ_i values for 216 VAM tests with different test parameters (three crack locations, three pump/probe locations, six probe frequencies,

and four probe to pump amplitude ratios). We count the false positives, false negatives, true positives, and true negatives for these tests to arrive at the following (approximate) values for the likelihood function displayed in Table 3.

Table 3: Approximate values for likelihood functions.

Likelihood Functions	Approximate Values
$P(D_i = 1 \mid \theta_i = 1)$	0.653
$P(D_i = 0 \mid \theta_i = 1)$	0.338
$P(D_i = 1 \mid \theta_i = 0)$	0.347
$P(D_i = 0 \mid \theta_i = 0)$	0.662

VAM tests were conducted on a plain concrete slab specimen to perform ASR damage localization. Details of the test procedure are described next. In VAM tests, the pump and probe excitations are delivered using piezo-stack actuators. The locations of these actuators and the frequencies at which they operate (i.e., the pump and probe frequencies) are varied. The structural component of interest's response is measured using a finite number of accelerometers placed on the element's surface. The relative magnitude of a sidebands-based metric at various accelerometer locations are used to map the damage (ASR-induced cracks) in the component. The VAM test's performance depends on the values of parameters used, as well as on the methodology used for processing the data collected during a VAM test. The first fundamental frequency (920 Hz) of the specimen is used as the f_{pump}.

It has been reported in the literature that the probing frequency, f_{probe}, should be at least 10 to 20 times the pumping frequency [34]. When the f_{probe} is N times the pumping frequency, it allows for the crack to open and close N times in a pumping cycle. Thus, the ratio between the pump and probe decides the opportunity (number of times per cycle of pumping vibrations) for

modulation to occur. Following these guidelines, probing frequencies ranging from 10 to 21 kHz were used in our experiments. Different probing frequencies are associated with the different wave lengths that can detect different sized cracks. Higher frequencies can detect smaller cracks. A study by Karve [35] showed the ideal probing frequency to be about 25 times higher than the pumping frequency. The highest frequency we could use, given our acquisition system and the sensor sensitivity, was 21 kHz. The amplitude and location of the probing actuation were also varied in different experiments. Specifically, we used four amplitudes for the output voltage of the probing signal generator: 500, 250, 100, or 50 mV. The output voltage of the pumping signal generator was maintained at 500 mV. The pumping and probing signals were amplified and sent to the piezo-stack actuator, and a constant amplification factor (+28 dB) was maintained for all tests. Tests using multiple frequencies and configurations were conducted to ensure the cracks were being consistently captured and to decrease uncertainty in localization. A maximum of 21 accelerometers was placed on the concrete specimen and connected to the data acquisition system. The locations of these accelerometers were varied for each experiment. The accelerometers had a sensitivity of 100 mV/g.

2.5.2. Preliminary test to confirm the onset of ASR

The details of EXP-A used to conduct and validate the proposed VAM-based based damage localization were discussed in Section 2.4.1. Here, we provide results of preliminary ASR detection tests performed to confirm onset of ASR in the specimen. After curing the specimen in aggressive (high temperature, high humidity) environment, the first indication of ASR damage was detected using NIRAS. For NIRAS tests, EXP-A was excited using an impact hammer at four different locations near the four corners of the top surface. Five different excitation (impact)

amplitudes were used. The NIRAS technique showed amplitude-dependent shift in resonant frequency (Figure 8).

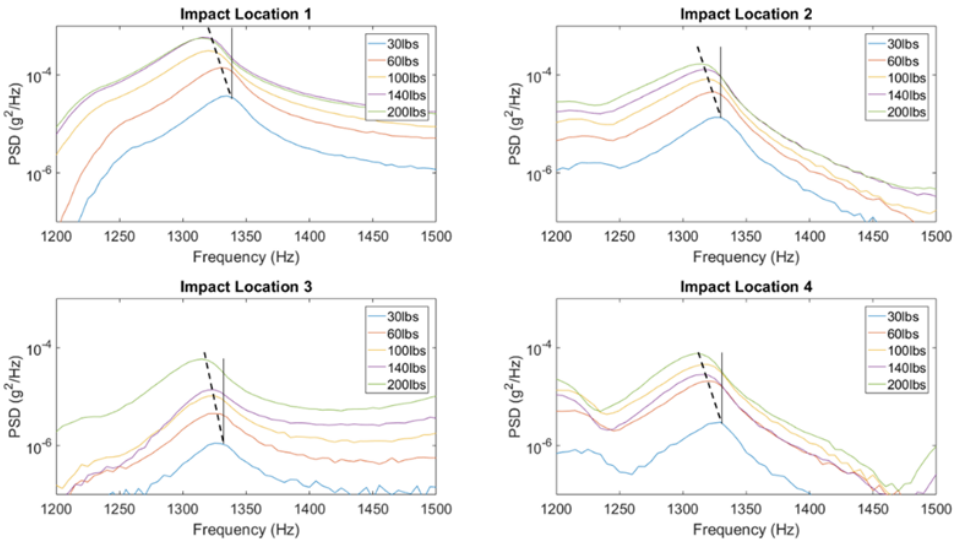


Figure 8: Results of NIRAS test confirming the excitation-amplitude-dependent shift in the resonant frequency of the specimen. The plots show PSD of time history of acceleration at a point on the surface of EXP-A.

The deformation measurements of EXP-A were not significant (as expected), since the aggregates were only in a single layer mid-thickness at four locations in the slab (significant ASR-related expansion is only expected when the aggregates are distributed throughout the slab). Hairline cracks were then observed on the surface of the specimen, and a gel effluent seeped out of the sides of the specimen at three locations. In Figure 9, the red squares identify the three locations where seepage of the gel effluent was observed. Figure 9 also shows images of the gel effluent on the sides of the specimen. The gel effluent was dissolved in 10 mL of 2% nitric acid, and concentration of various chemical elements in the solution was obtained using inductively coupled plasma-optical emission spectroscopy (Table 4). The elements typically found in the ASR gel [40], [41] were found in the gel effluent, confirming that the substance was a product of ASR (the high concentration of sodium is due to the addition of NaOH solution).

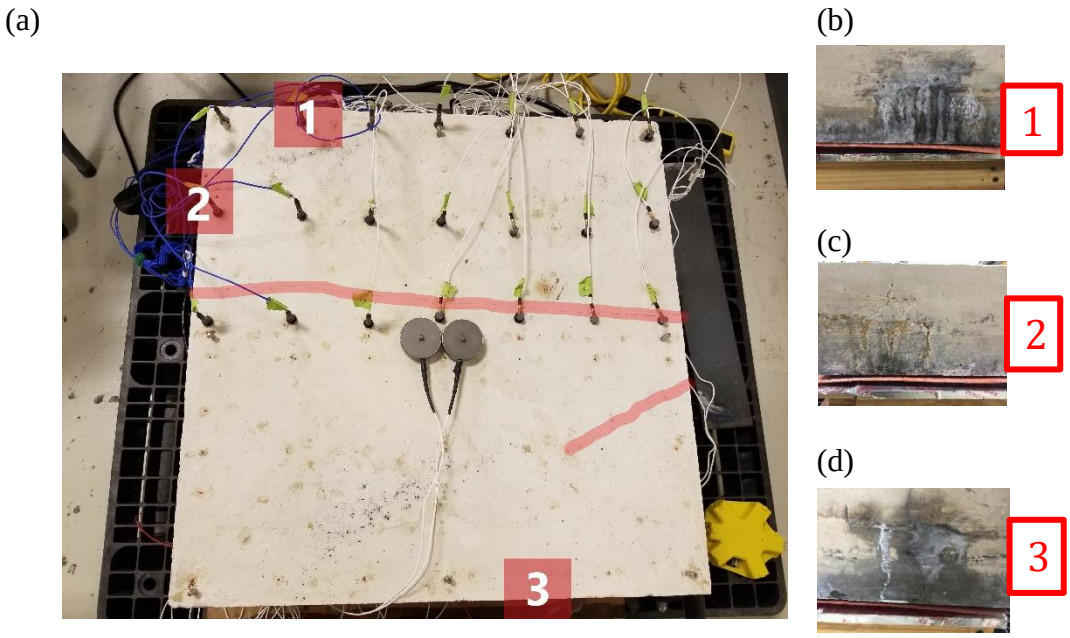


Figure 9: Preliminary indicators of ASR (a) Visually observed ASR gel and powder exudation sites 1,2, and 3, as well as surface cracks on the top surface of EXP-A (shown with red lines); (b), (c), (d) ASR gel and powder effluent observed at the locations noted in (a).

Table 4. Results of inductively coupled plasma-optical emission spectroscopy.

Element	Concentration (mg/L)
Aluminum	1.3
Calcium	3.3
Potassium	520.0
Sodium	1,100.0
Sulfur	2.0
Silicon	1,600.0

2.5.3. Petrographic study

After confirming the presence of ASR (as discussed in section 2.5.2), VAM tests were conducted on EXP-A to obtain damage maps. After collecting VAM data for multiple test parameters, the petrographic examination was performed to validate results of VAM tests. Here, the results of petrographic examination are discussed first, before the VAM test results, in order to

ease the future discussion of VAM test validation. Thus, approximately three years after casting, core samples intersecting the aggregate pockets were extracted from the cement paste slab (see Figure 10 (c), compare with Figure 4).

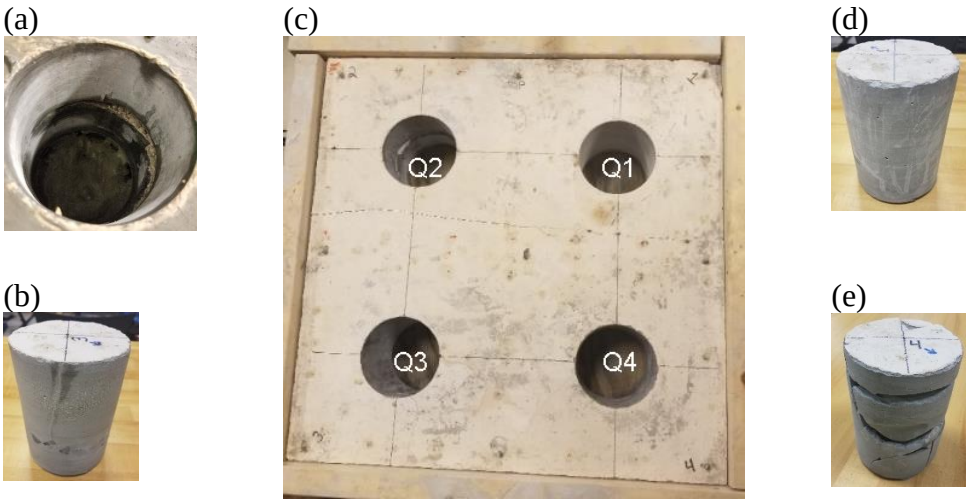


Figure 10: (a) Specimen after removing the core sample from quadrant 2. The large gap, where the silica layer was placed, is seen at mid-depth of EXP-A, (b) core sample removed from quadrant 3, (c) EXP-A after removing the core sample from four quadrants, (d) core sample removed from quadrant 1, (e) core sample removed from quadrant 4.

Cores taken through the Maine aggregate (quadrant 1) and amorphous silica (quadrant 2) pockets were unsuitable for analysis; the amorphous silica appeared to have completely reacted, leaving a large void in the sample. No aggregate particles were identified in the Maine core. However, trace evidence of ASR gel was observed lining a void on a fracture surface of the Maine core, indicating that the aggregate had moved from its initial location during casting. A petrographic examination conducted on the cores taken through the pockets containing the Ontario and New Mexico aggregates (Quadrants 3 and 4 respectively in Figure 10) confirmed that the significant damage and gel exudations observed were the result of ASR occurring in these aggregate pockets. The examination was conducted using optical microscopes with reflected

(polished section) as well as transmitted polarized light (thin section); and a scanning electron microscope with backscatter emission (BSE) as well as energy dispersive spectroscopy (SEM-EDS) (thin section).

Examples of observed ASR-induced cracking and reaction products (gel) are shown in Figure 11 (Ontario aggregate, quadrant 3) and Figure 12 (New Mexico aggregate, quadrant 4). Numerous microcracks were observed traveling from the aggregates into the paste, bridging between particles and extending to the sides of the cores parallel to the surface in both cores. Perpendicular cracks traveled from the aggregate pockets to the top and bottom of the cores as well. The abundance and width of cracking was greater in the core containing the New Mexico aggregate. ASR gel and crystalline ASR product comprised of silicon, potassium, calcium, and a trace amount of sodium were present in cracks within the aggregate, in cracks that extend into the paste, along the paste/aggregate interface, and filling air voids in the paste. The calcium content of ASR gel at the paste/aggregate interface and within the cement paste is known to increase over time as the potassium and sodium are leached out and replaced by calcium from the cement paste, a phenomenon known as alkali recycling [42], [43]. An example of gel that has experienced significant exchange of alkalis for calcium is shown in the lower right of Figure 12. ASR gel can also crystallize over time when there is a loss of access to water. Crystalline ASR products such as that shown in Figure 11 were observed in cracks within both the aggregates and the paste in both cores; this is consistent with the slab being stored in a dry laboratory environment for approximately two years prior to coring. Cool and dry conditions do not promote the development of ASR or continued swelling of gels, thus the reaction would have likely ceased after being removed from the environmental chamber for an extended period, during which continued drying of the slab would also have promoted crystallization of ASR gel.

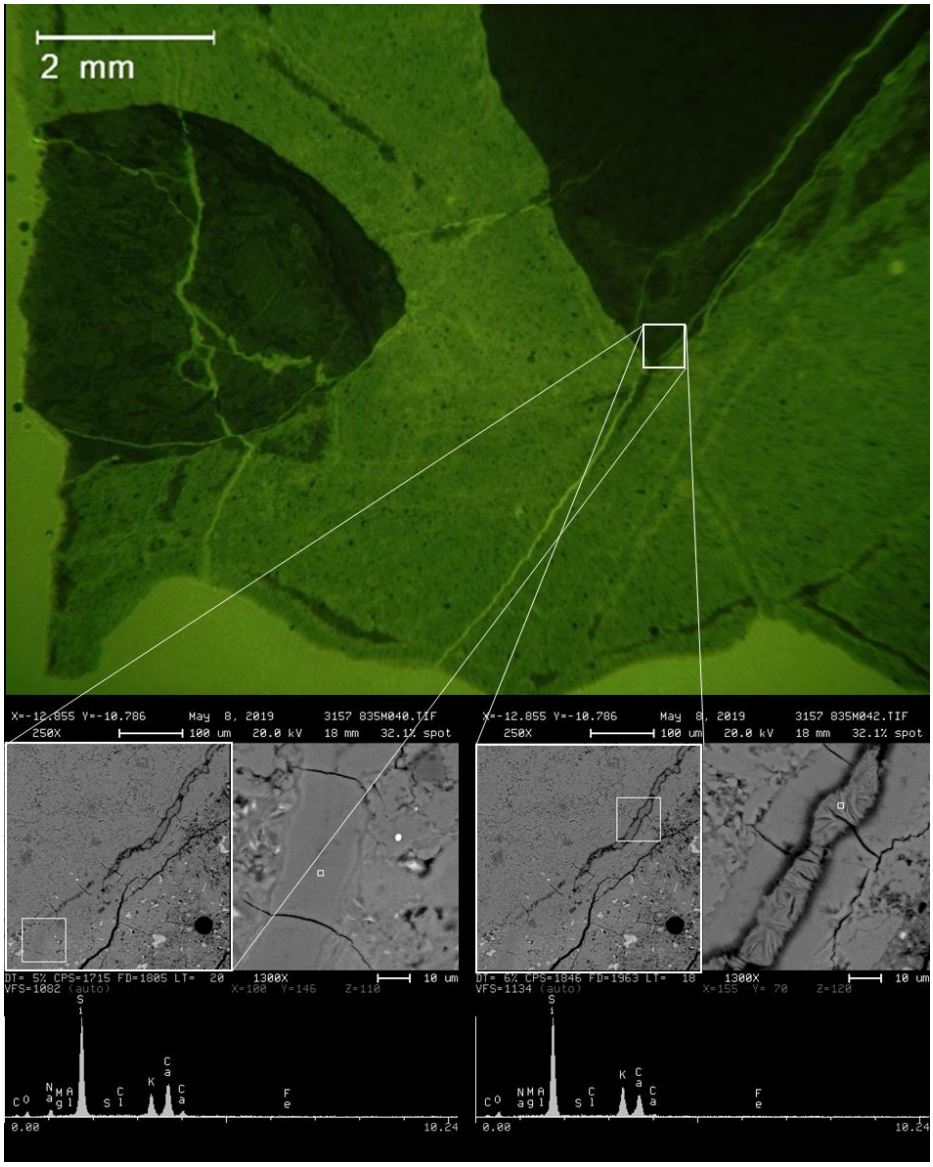


Figure 11: Top: Micrograph in fluorescent light of thin section taken from Ontario aggregate core showing extensive ASR-induced microcracking extending both between and within aggregate particles. Lower left and lower right: Backscatter electron (BSE) images of an area of interest highlighted by white box in top image, and EDS point spectra of ASR gel on the left and crystalline ASR product on the right at two locations within a crack in the limestone aggregate.

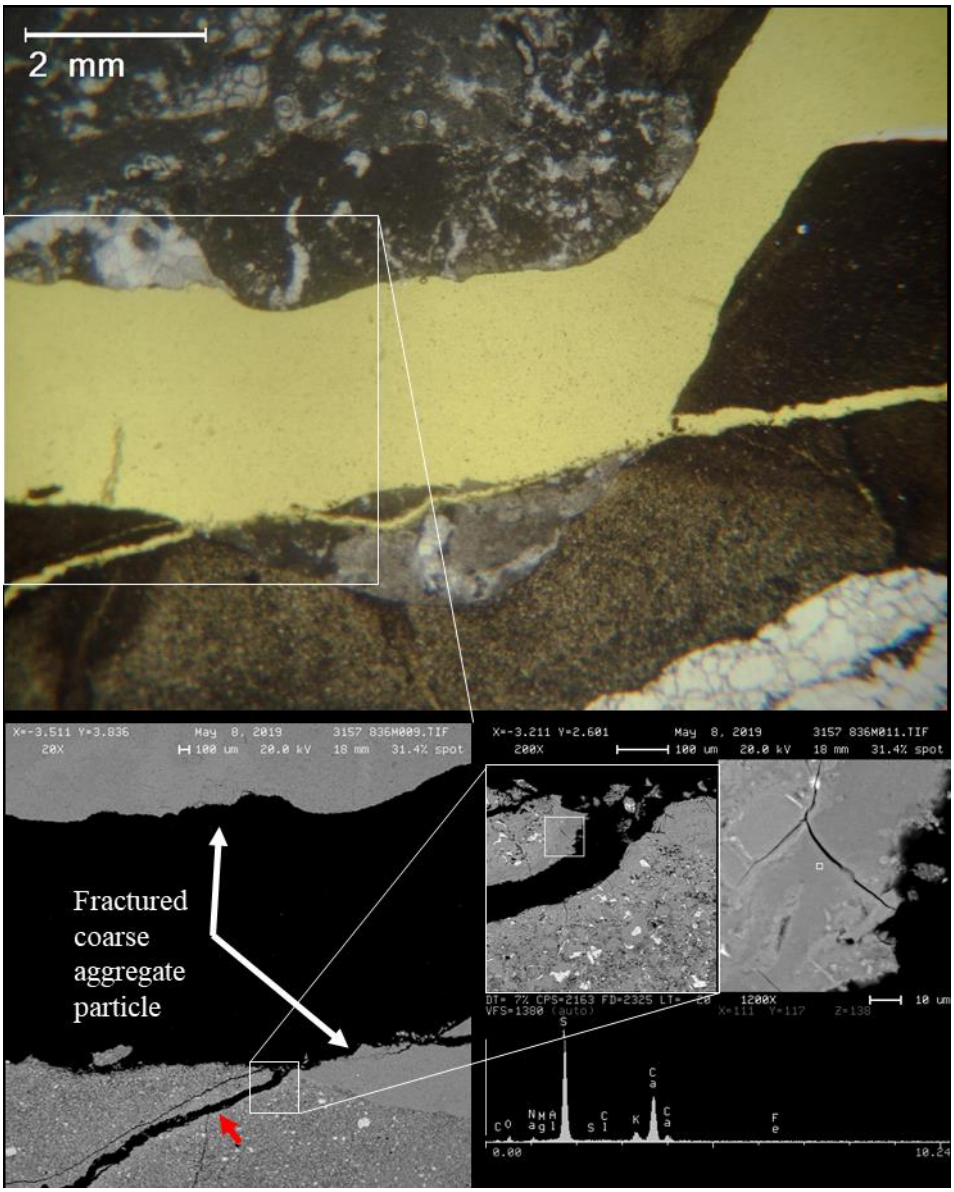


Figure 12: Top: Micrograph in plane-polarized light of thin section taken from New Mexico aggregate core showing extensive ASR-induced microcracking. The large (yellow) epoxy-filled crack extending across the full field of view shows a fracture through a limestone aggregate particle which extends into the paste on either side. Lower left and right: Backscatter electron (BSE) images showing oblique microcracking extending from the main fracture, and EDS point spectrum of ASR gel lining one side of the highlighted crack.

2.5.4. VAM Test Results

In this section, we report the results of VAM-based damage localization for EXP-A. As described in section 2.4.2, due to the limited number of accelerometers available for conducting tests, we divide the slab surface into two halves. We record acceleration time history for each half using 21 accelerometers, compute the corresponding linear spectra, and obtain the SBSum metric. We then plot the variation of the SBSum metric over each half using linear interpolation of the SBSum metric obtained at the accelerometer locations. We juxtapose the damage maps for the two halves in order to obtain a damage map for the entire specimen. Since there is no overlapping of sensors in the two halves, it is assumed that this is reasonable. This is especially true for damage mapping using the damage index defined above (equation 1), because the damage index is based on the relative SBSum values for a given test. Next, we discuss the results of VAM tests conducted on EXP-A in the context of: (a) the utility of the proposed Bayesian data fusion method, (b) selection of suitable probe frequencies for concrete (or concrete-like) specimens, and (c) Bayesian data fusion results showing damage locations in the slab.

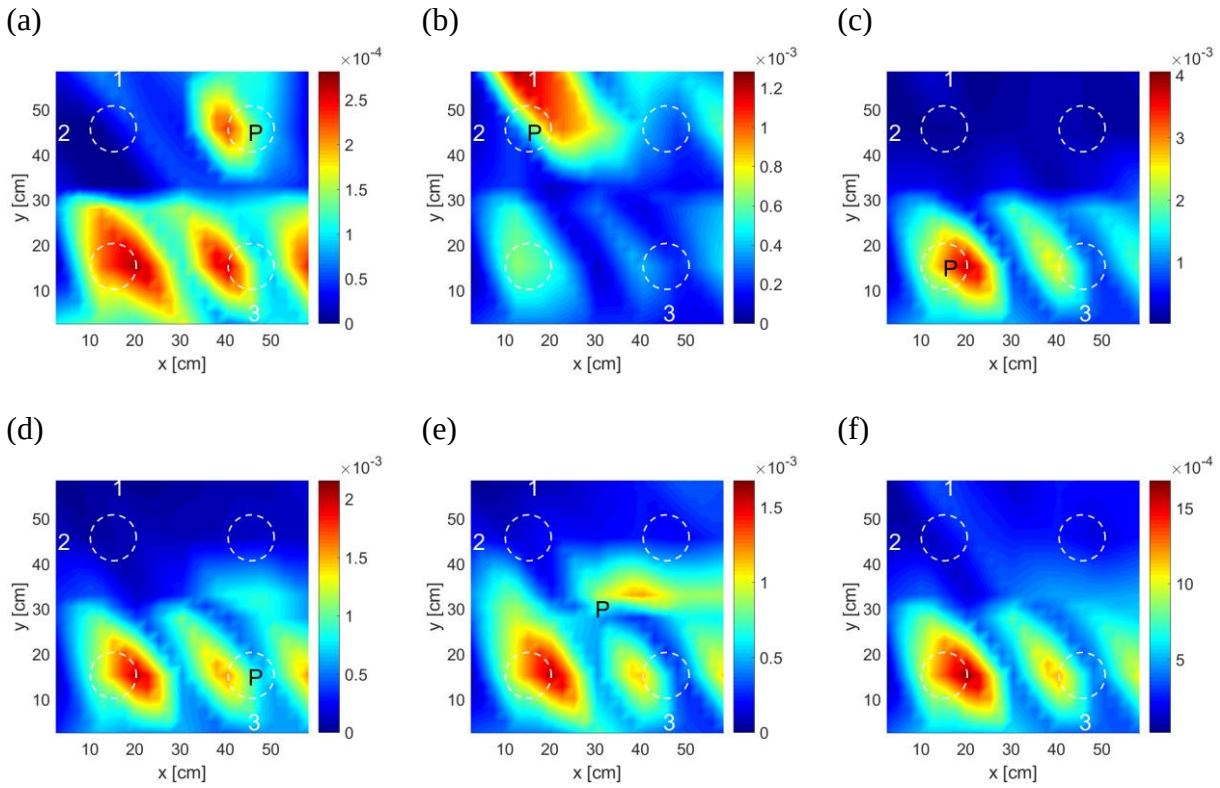


Figure 13: Variation of SBSum over the top surface of EXP-A for pump frequency of 920 Hz, probe frequency of 19 kHz, and probe amplitude of 250 mV; (a)-(e): SBSum plots for the five configurations of pump and probe, (f) average of SBSum values plotted in (a)-(e). The pump/probe locations are denoted by letter "P", the ASR gel/powder exudation locations are denoted by numbers (1, 2, and 3), and the locations of core samples taken for petrographic examination are denoted by dashed circles.

VAM test results for the probe frequency of 19 kHz, and probe amplitude of 250 mV (peak-to-peak) are shown in Figure 13. Subfigures (a) to (e) show the variation of SBSum on the surface of the slab obtained by interpolating the SBSum values at 42 sensor locations for all five configurations (i.e., pump/probe locations). The SBSum distribution changes with configuration, affirming that different steady state vibration patterns (nodes and antinodes of the standing waves) as well as attenuation are at play. Different configurations identify different possible damage locations (ASR gel/powder exudation points or surface cracks) by showing relatively higher SBSum values at/near the damage sites. Subfigure (f) shows the average SBSum plot obtained by

averaging the values of SBSum at each sensor over all the five different pump/probe locations (configurations). Figure 13 shows a simple method (mean of SBSum data) for fusing data from multiple test configurations. It can be seen that some of the test configurations show damage on the top half of the slab, however after computing the mean, damage is seen only on the bottom half of the slab.

We next compare two candidate alternatives of assimilating information from multiple VAM tests: (a) computing the mean of SBSum data obtained from VAM tests (Figure 14), (b) Bayesian data fusion described in 2.3.3 (Figure 15).

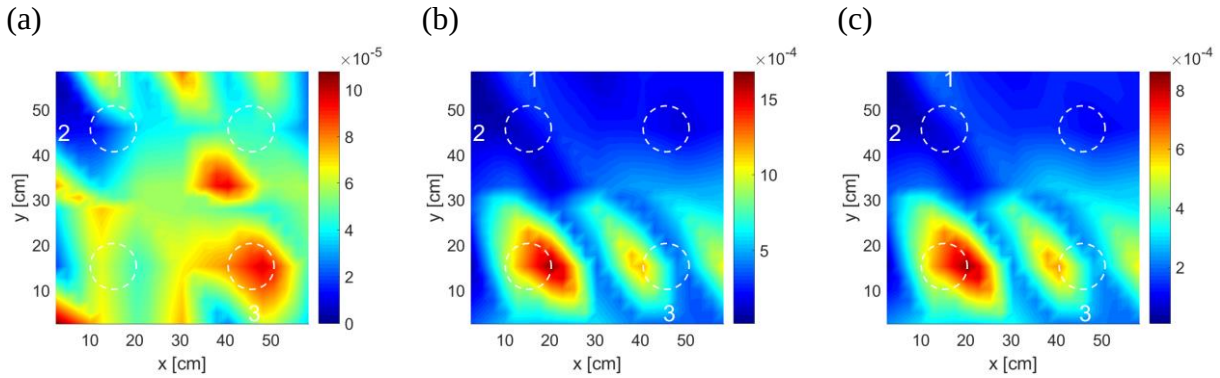


Figure 14: Variation of SBSum over the top surface of EXP-A for all five configurations (average), pump frequency of 920 Hz, and probe amplitude of 250 mV; (a) probe frequency = 16 kHz, (b) probe frequency = 19 kHz, (c) average of SBSum values from (a) and (b).

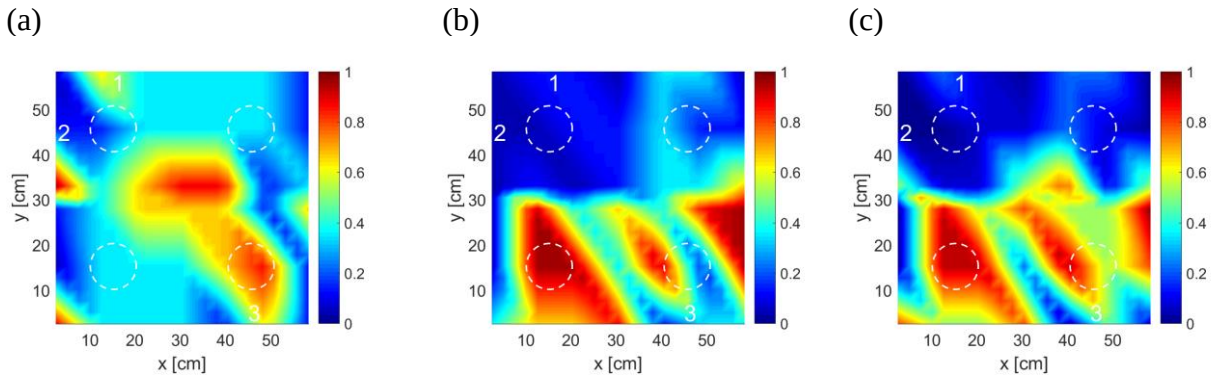


Figure 15: Variation of damage probability (computed using Bayesian information fusion method detailed in section 2.3.3) over the top surface of the slab for all five configurations, pump frequency of 920 Hz, and probe amplitude of 250 mV; (a) probe frequency = 16 kHz, (b) probe frequency = 19 kHz, (c) fusion of information from (a) and (b).

Figure 14 depicts the simple averaging scheme to combine VAM test data from two tests that used different probing frequencies. Different frequencies may illuminate flaws of different sizes and help enrich the diagnostic information obtained using a VAM test, making this approach useful for solving real-world diagnostic problems. However, it can be seen in Figure 14 (a) and (b) that the magnitudes of SBSum values for these two test cases are different – the highest SBSum value for the 19–kHz-case is about fifteen times higher than that for the 16-kHz-case. The averaged (fused) damage map depicted in Figure 14 (c) appears to lose the information from the 16 kHz damage map. Specifically, the higher SBSum values near (x,y) ~ (40, 35) cm seen in Figure 14 (a) are *drowned* by the high SBSum values in the bottom half of the slab in Figure 14 (b). The magnitudes of SBSum are different for different configurations. In an average-based data fusion procedure, the high magnitude SBSum values from one configuration dominate the low magnitude SBSum values from other configurations. This leads to loss of information. The key concept in VAM-based damage localization is, however, the relative degree of SBSum values in a given VAM test (not across different VAM tests). Therefore, it is not informative to average the SBSum data in its raw form across multiple tests. This difficulty is overcome by the proposed Bayesian

data fusion method. In this method, first, the damage map in a given test is developed using the observed damage index, then the results from different tests are fused using Bayesian updating, as described in Section 2.3.3. The result of Bayesian data fusion for two tests (at probe frequencies 16 kHz and 19 kHz) is shown in Figure 15. A careful observation of Figure 15 (c) shows that this approach indicates high damage probability around (x,y) ~ (40, 35) cm. Thus, the proposed Bayesian methodology prevents loss of diagnostic information while performing information fusion.

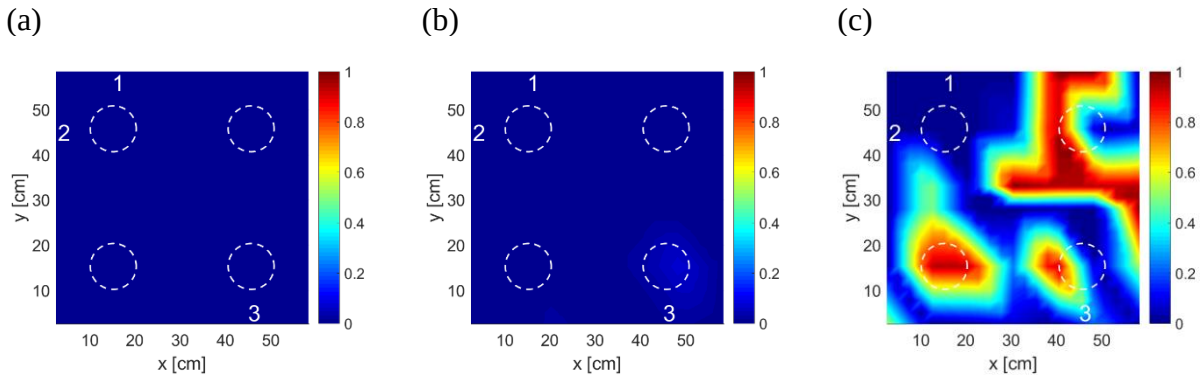


Figure 16: Damage probability maps for EXP-A (a) probe frequency = 10 kHz; (b) probe frequency = 15 kHz; (c) probe frequency = 20 kHz.

Figure 16 shows the damage probability maps obtained by fusion of data from tests with four probe amplitudes and five pump/probe locations, at three probe frequencies (10 kHz, 15 kHz, and 20 kHz). This figure represents the general trend in the damage probability plots as the probing frequency increases. It can be seen that the lower frequency is unable to detect any damage. The wavelengths corresponding to lower frequencies are longer than those corresponding to higher frequencies. Thus, the low frequency waves may be unable to detect flaws of small size. Our numerical experiments suggest that in the VAM test, a true positive rate higher than 70% is achievable if the ratio of flaw size to the shear wavelength in the material is greater than 0.7 [35].

The damage probability at a sensor location is dependent on the consistency of occurrence of a high-enough SBSum value at that sensor location. The plots showing low damage probability for lower probing frequency imply that the higher SBSum values do not occur consistently at the same sensor locations. The Bayesian updating algorithm correctly assigns lower damage probability to these test results. Based on this observation, we use the data (five configurations and four probe amplitudes) obtained for six probe frequencies (16 kHz, 17 kHz, 18 kHz, 19 kHz, 20 kHz, and 21 kHz) to arrive at the aggregated VAM -based damage maps for the slab (Figure 17). Figure 17 also shows the results of data fusion using an average of SBSum values for all the tests.

The results obtained from the two data fusion techniques are subjected to validation assessment using the ASR damage information obtained from petrographic examination (discussed in section 3.1). According to the petrographic examination, only trace amount of ASR damage was found in quadrant 1 (Maine). Diffused macro/micro cracking was not observed in the core sample taken from this quadrant. In quadrant 2 (silica, Figure 10(a)), the core sample showed a large (about 1 cm thick) gap where silica was placed while casting the slab. Cracks of this size are typically not created by ASR gel expansion around aggregates. The applied pump excitation was not strong enough to induce opening and closing of this large gap. Thus, it is not reasonable to expect the proposed VAM-based methodology to detect this large gap. Characteristic, diffused cracking was also absent in quadrant 2 as there were no aggregates in this quadrant. Core samples taken from quadrants 3 and 4 showed characteristic diffused macro cracking, as well as ample evidence of ASR. Thus, the damage localization methodology should be able to detect surface cracks (Figure 9), and diffused cracking in quadrants 3 and 4. In fact, both damage maps in Figure 17 show high damage probability at/near the visible surface cracks (Figure 9) as well as in quadrants 3 and 4. Overall, the damage localization method is able to locate damage at the expected

damage sites (quadrants 3 and 4). It can also be seen that the Bayesian information fusion technique produced the sharpest damage map. We remark that if the computational modeling required to obtain approximate likelihood values is prohibitive (due to the computational cost), the average of SBSum values can, potentially, be used for damage localization. However, this approach may be susceptible to false positive results (as seen in the top half of Figure 17 (a)).

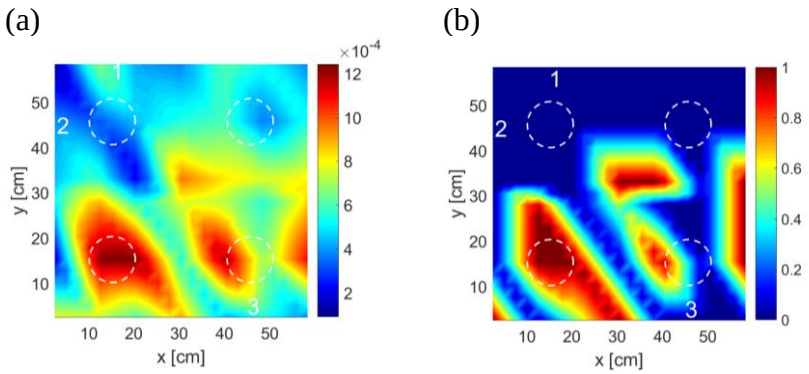


Figure 17: Damage maps for EXP-A: (a) average of SBSum; (b) Bayesian fusion of the damage index.

2.6. Conclusion

In this chapter, the VAM technique was used to detect and localize ASR-related damage in a concrete specimen exposed to accelerated aging conditions in a laboratory setting. Specifically, the VAM technique was applied to EXP-A, a cement slab cast with four pockets of reactive aggregates (known specific locations). A damage index based on sidebands present in the linear spectrum of nonlinear dynamic structural response (acceleration) was used to map the distribution of damage in the concrete specimen. We studied several practical aspects of data processing and noise reduction involved in VAM testing for real structures. We experimented with important test parameters that affect the VAM results, such as locations, frequencies, and amplitudes of the probe and pump excitations. For the slab specimen containing pockets of reactive aggregate at known

locations, VAM-based damage mapping revealed damage signatures at locations that showed visible signs of ASR-induced cracking (that could open and close due to the applied pumping excitation). We proposed a Bayesian data fusion methodology to assimilate the information gained from the VAM tests with different test parameters and compared the results of this method against simpler average-based data fusion methods. The proposed Bayesian methodology can be utilized with features of ASR-damage-induced nonlinear dynamic structural response other than the sidebands-based metric considered in this work. The VAM-based damage localization technique was able to identify most of the known damage sites for the slab, as confirmed by petrographic and chemical validation experiments.

We remark that if a damage mechanism other than ASR induces nonlinear structural behavior that creates sidebands in the spectra of the recorded response, the proposed method cannot differentiate between the sidebands induced by this other mechanism and those induced by ASR. However, it may be of interest to know the extent and growth of this other damage mechanism as well. Thus, the proposed methodology may be applicable for localization of sideband-inducing damage mechanisms other than the ASR. Work needs to focus on scaling up the technique to larger scale, field implementation of VAM-based diagnosis in realistic size concrete structures such as those found in nuclear power plants. Consideration of realistic structures also requires the study of reinforced concrete, extending from the plain cement slab considered here. One of the key challenges is the number of sensors (accelerometers) needed to localize the damage. For a large structure, it is not feasible to use a large number of accelerometers; therefore, the number of sensors needs to be minimized. Wave sources that can illuminate parts of a real-world concrete structure with sufficiently strong vibration (to cause detectable VAM) also present challenges in the field application of VAM-based damage diagnosis.

CHAPTER 3

Damage Localization in a Plain Concrete Specimen Using 2D Physics Model-Informed Machine Learning

3.1. Motivation

Both model-based and model-free methodologies have been used in SHM studies for damage diagnosis [1] [44]–[47], as discussed in Chapter 1. The model-free approach discussed in Chapter 2 for VAM-based damage localization [48] suffers from the following limitations: a) the methodology relies on relative values of damage indices measured at different sensor locations, and hence requires the analyst's judgment to define the damage index threshold for damage classification and localization; b) in Bayesian diagnosis, computation of the likelihood of damage at a sensor location given the damage index value and the threshold is expensive (the methodology above used an approximate, averaged likelihood-based method); and c) the methodology localizes damage along two (surface) dimensions of the component but does not provide the location along the third (depth) dimension. This objective investigates whether the above limitations could be overcome by using supervised machine learning techniques and nonlinear wave physics models.

3.2. Relevant Background

Machine learning (ML) models have an enormous capacity to learn complex, nonlinear phenomena [49]. Hence, they are well-suited for modeling nonlinear dynamics for complex, heterogeneous materials like concrete. Both supervised and unsupervised techniques have been

studied for concrete SHM [1]. Unsupervised learning is typically used to extract damage sensitive features from diagnostic data [50]–[56]. In supervised learning, labeled training data consisting of the chosen model inputs (damage sensitive features) at different damage levels for the real-world structure of interest is needed. As this data is typically not available, most of the past studies have used numerical simulations or scaled-down laboratory specimens for generating training as well as validation datasets to localize damage in 2-dimensions [51], [56]–[58]. Some investigations considered real-world structures and were limited to building and validating machine learning models for structures for which labeled training data is available [52].

State of the art for damage index-based SHM can be broadly categorized into: (a) simple rules (e.g., thresholds) determined by domain experts by studying physics simulations of the test procedure or laboratory test data for the test in question [34], [48], [59], (b) ML-based pattern recognition using field or laboratory experiment data [58], [60], [61], and (c) ML-based pattern recognition using data generated by simulation models for the diagnostic test process (different test parameters, damage severity levels, damage locations, etc.) [52], [62]. The first category relies on the expert's ability to assimilate information from experimental data and physics models. The second category requires labeled training data corresponding to known damage severities and test conditions, which is typically difficult, expensive, and often impossible to obtain. Here, we investigate the utility of a hybrid approach, where physics-informed machine learning models (models trained using simulation data) are used for the NDE of a laboratory specimen. An overview of various physics-informed ML (PIML) strategies that could be used for SHM is provided next.

PIML has mostly been used for building forward prediction models [60], [63]–[65]: these could be used in prognosis to predict the quantity of interest or in diagnosis to compute the

likelihood function while solving the inverse problem. This chapter, on the other hand, investigates PIML for the more complex damage index-based diagnosis problem, which is tantamount to learning the task performed by solving an *inverse problem* [63], [64] without explicitly computing the likelihood function. That is, the complex relationship between the damage index and damage estimate, in effect, captures what a solution of the related inverse problem would have achieved. Different types of approaches have been pursued in the literature for training forward prediction PIML models [66], such as: (a) impose constraints established by the physics model on the loss function while training the machine learning model [63], [65], [67], [68]; (b) use physics model estimates as additional inputs to the ML model; (c) incorporate data obtained from a computational simulation of the governing physics into the pool of training data [62], [69], [70]; and d) train an initial ML model using data generated by the physics model, and then update, retrain or correct this initial model with observation data [12]. In this work, we use the third approach for damage index-based diagnosis. We provide next a review of past PIML efforts for SHM.

Yuan et al.[63] used a PIML model for beam vibration as the forward model to be used in an (iterative) inverse problem solution for beam structure health monitoring. They used artificial neural networks (ANNs) as the ML model and used the laws of physics to guide the back-propagation process. Similarly, Raissi et al. [64] implemented an artificial neural network (ANN) training with observation data augmented by data from physics model runs, for solving inverse problems in SHM. Rai and Mitra [62] used physics (finite element) models to extract damage-sensitive features to train a feed-forward neural network for localizing damage in a simulated thin aluminum plate. Here, the knowledge gained from governing physics simulations is implicitly used to train the ML model (to decide the damage indices or the inputs to the ML model). Seventekidis et al. [70] use physics model simulation data to train a health state classification ML model. They

used a single fidelity physics model, which was calibrated with experimental data before generating the training data. The resulting classification ML model was shown to be adequately accurate in identifying simple damage states. Note that Yuan et al. [63] have provided a review of PIML methods for SHM. PIML has shown promise for applications in SHM and in vibration-based damage localization. Much of the previous work has focused on building PIML models to be used as forward models in the inverse problem solution for relatively simple physics (e.g., linear vibrations of a beam). For complex (three-dimensional) structures, with complex diagnostic tests involving nonlinear system behavior, the damage index-based approach is desirable.

There are thus two main objectives in this chapter: investigate whether the aforementioned drawbacks of VAM-based damage localization could be overcome using ML (data-driven model-based SHM), and explore the utility of physics-based ML for tackling the training data related challenges faced by data-driven model-based SHM methods. We use computational physics simulations to generate data for training the ML models. The computational cost involved in numerical simulation of the nonlinear phenomenon of interest, for heterogeneous, three-dimensional (3D) domains could also be prohibitive in real-world applications. Hence, in this research, the governing wave physics is simulated for two-dimensional (2D) domains (under plane strain conditions). The response data obtained from 2D finite element analysis (FEA) are used to compute the corresponding values of SBSum (damage sensitive feature), which are then used in the training data for supervised ML models.

One example of a popular data-driven surrogate model is an ANN. These networks use a series of connected layers with multiple nodes and neurons, intended to represent synapse connections in the brain. The models do not require previous knowledge or empirical formulas, making them desirable for many applications such as image and speech recognition, video games, consumer use

predictions, and damage estimation. Input values feed into the ANN and pass through the hidden layers before arriving at the output value(s) for that set of inputs. Each time the input travels through a layer, it is assigned a different weight. The learning method is called backpropagation, where the difference between the output of the network and the actual output is minimized [71]. The process is repeated, and the weights are updated until either the desired number of runs (epochs) are completed, or the value of a loss function converges. Figure 18 shows a shallow neural network with only one hidden layer.

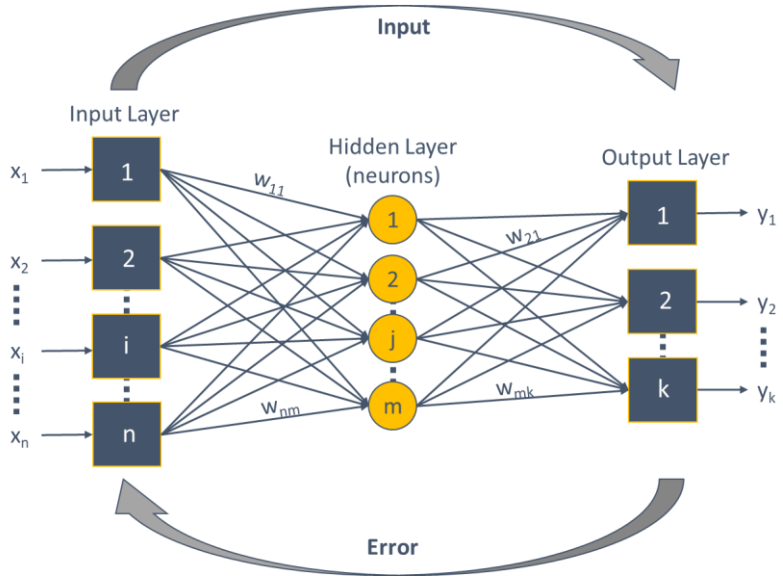


Figure 18: Example of an artificial neural network.

Machine learning models have been used to estimate damage, durability, and service life of reinforced concrete and other composite structures [57]. The models can be either supervised or unsupervised. In supervised learning, the desired output is a known target. Estimating supervised continuous target variables is known as regression, whereas the estimation of discrete target variables, such as in this experiment, is known as classification. In unsupervised, learning the

output is unknown, so the model identifies patterns and relationships within the data [49]. Unlike physics models, machine-learning models allow for additional input variables to be easily incorporated in the estimation model.

The simulation-based training data is used in this study to construct two feed-forward artificial neural networks (ANNs). The first ANN is used to predict the value of the damage index (SBSum). Since the damage index is a continuous variable, this ANN is a prediction model that estimates SBSum value at any given sensor location given test parameters and damage location as well as extent. This model is referenced as the prediction model. The second ANN classifies the region in the given sensor's neighborhood into two discrete categories, damage present or damaged absent. This ANN is a damage classification model. The performance of the trained machine learning models is evaluated using simulation test data (2D FEA models) and laboratory test data (a cement slab specimen previously investigated in [48]). The methodology aims to overcome the aforementioned drawbacks of VAM-based damage localization methodology by building physics model-informed machine learning models to perform damage classification by learning the damage index threshold from the training data, to facilitate computation of the likelihood of damage given damage index data, and to estimate damage location in the depth dimension.

3.3. Methodology

This section provides the details of the computational model used to generate training data and the proposed supervised ML models that facilitate VAM-based damage diagnosis. Chapter 2 sections 2.3 and 2.4 , discuss the details of VAM, the experiments that are being used for validation (EXP-A), and the details of the test parameters used.

Past research on damage localization using VAM in thick concrete slabs has primarily relied

on either a fully data-based (model-free) approach or a physics model-based approach [35], [48]. In a deterministic setting, an analyst-defined statistical metric (e.g., mean) of SBSum values was used in previous work as the threshold SBSum value. This threshold was used to determine which sensors indicated damage and to obtain the damage map. Here we explore whether for a given damage metric (e.g., SBSum), supervised ML could be used to perform binary classification (damage present v. absent) and thus implicitly infer a damage index threshold given multiple parameters without the analyst having to select a threshold. Further, a probabilistic (e.g., Bayesian) damage localization approach relies on physics models to obtain the likelihood function. Due to the computational expense of simulating the complex governing physics, approximate (averaged) likelihoods have been used in previous work [48]. Note that this is a common aspect of many damage diagnosis algorithms. These involve the solution of an inverse problem, which necessitates multiple evaluations of the forward model (to compute the error between estimated and measured damage indices in a deterministic approach, or to compute the likelihood of observed damage index value given a damage state in a probabilistic approach). In this work, we investigate the utility of a computationally inexpensive ANN-based surrogate (forward model) to aid the likelihood estimation. In this manner, the classification model is a ML-based generalization of damage detection for concrete SHM. In contrast, the regression model can be thought of as an application of the well-known surrogate modeling approach, subsequently used in a Bayesian estimation methodology for damage localization.

3.3.1. Physics-informed, supervised machine learning for VAM-based damage diagnosis

Two VAM test simulation data-driven models are constructed for damage diagnosis. The first approach builds a regression model that estimates the SBSum value at a sensor location. The

regression model and VAM test (SBSum) data are used in a Bayesian damage diagnosis algorithm, where a search is done for the damage locations that best-fit the observed VAM test (SBSum) data. The second approach employs the VAM test parameters as well as the measured SBSum values at a particular sensor to classify the sensors as either indicating or not indicating the presence of damage. This is the classification approach, where each sensor location is classified into one of the two categories: (1) damage present, or (0) damage absent. The details of the two models are discussed in the following sections.

3.3.1.1. Model structure and training

The choice of the ML model inputs and outputs is a critical consideration for successful damage diagnosis. In this work, the choice is made based on a phenomenological understanding of the VAM test setup and the wave physics governing the problem of interest. The two ML models built to facilitate damage diagnosis have different purposes. The regression model aims to estimate the damage index (SBSum) value for a given damage state and test configuration at a given sensor. The output of this model is thus the SBSum value at a given sensor location. The input should include all parameters that affect the SBSum value: VAM test parameters as well as damage location and size. The SBSum value at a sensor may also be affected by the distance of the sensor from the pump/probe excitation location. This effect can be accounted for in the model by providing the sensor's location to the regression model. Note that the constitutive behavior of the test specimen may also affect SBSum values. However, it is unlikely that precise knowledge regarding the material state and constitutive behavior in the structural component to be tested will be available for performing damage diagnosis. Hence, the regression model estimates the damage index value given the test parameters, crack location, crack extent, and the sensor location relative to actuator locations. This model can then be used to determine the location of damage by solving

the inverse problem, given experimental data of SBSum values at different sensors for a given dual excitation.

The classification model, on the other hand, aims to directly identify the presence or absence of damage in the vicinity of a sensor. The output of the model is thus a binary indicator (1: damage present, 0: damage absent). The input includes important test parameters, location of the sensor, and the damage index (SBSum). In addition to these quantities, the classification model needs to assess the relative SBSum value at the sensor in question compared with other sensors in the VAM test. VAM-based damage localization works by finding the sensors that have relatively higher SBSum values in a test. To provide this information, a statistical metric (e.g., mean) of the SBSum values could be provided to the classifier. Note that it is not specified that an SBSum value above the mean SBSum value implies the presence of damage. Instead, the classifier learns the (potentially test-parameter-dependent and sensor-location dependent) SBSum threshold value from the training data.

The importance of the inputs chosen for both ML models is confirmed by performing an analysis of variance (ANOVA). In this manner, the structure of the two ML models is decided based on insights obtained from the physics of the VAM test process as well as statistical analysis (ANOVA). Practical considerations in manufacturing and testing slab specimens with hidden cracks situated in predetermined locations make it difficult to develop a supervised learning model using only experimental data. Hence, in this work, a computational physics model is used to simulate VAM tests with different damage configurations and VAM test parameters. A brief summary of this model is provided next.

3.3.1.2. Computational physics models for vibro-acoustic modulation

In this section, we describe the nonlinear wave propagation model used to simulate VAM tests

in this work. Various nonlinear crack–wave interaction models, such as classical nonlinear elasticity, stiffness asymmetry (bilinear stiffness), hysteresis, nonlinear dissipation have been reported in the literature [72], [73]. Here, we employ a contact acoustic nonlinearity (bilinear stiffness) model, to simulate a brittle, frictionless, internal (hidden) crack. We model the geometric nonlinearity at the crack faces as a contact nonlinearity that does not exhibit cohesion between the contact surfaces. The equations governing wave propagation in the 2D domain of interest (Ω) are given by [34]:

$$\boldsymbol{\nabla} \cdot \boldsymbol{\sigma} = \rho \boldsymbol{u} \text{ for } \boldsymbol{x} \in \Omega, \tag{6}$$

$$\boldsymbol{\sigma} \cdot \boldsymbol{n} = \boldsymbol{T}_N \text{ for } \boldsymbol{x} \in \Gamma_{\mathrm{N}}, \tag{7}$$

$$\mathbf{u} = \boldsymbol{u}_D \text{for } \boldsymbol{x} \in \Gamma_{\mathrm{D}}, \tag{8}$$

where $\boldsymbol{\sigma}(\boldsymbol{x}, t)$ is the Cauchy stress, $\boldsymbol{u}(\boldsymbol{x}, t)$ is the displacement, ρ is the density, $\boldsymbol{T}_N(\boldsymbol{x}, t)$ is the prescribed traction at the Neumann boundary, $\boldsymbol{u}_D(\boldsymbol{x}, t)$ is the prescribed displacement at the Dirichlet boundary, and dependance on spatial and temporal coordinates is dropped for brevity in the governing equations. We use linear elastic strain-displacement as well as constitutive model for the material. For the brittle, frictionless crack, the displacement-dependent (nonlinear) boundary condition is modeled by specifying $\boldsymbol{T}_N = \mathbf{0}$, when the crack is open and $\boldsymbol{T}_N = \tau \boldsymbol{n}$, when the crack is closed, where τ is the magnitude of the compressive force. Analytically, this corresponds to a *hard* contact condition, where the two crack faces do not impart force on each other when they are not in contact. The surfaces can impart any amount of pressure (compressive force) on each other when they are in contact, and penetration of one crack face into another is not allowed. Note that the frictionless contact implies that the interaction force between two crack faces is applied only in the normal direction (and not in the tangential direction). The governing equations are cast in their weak form, and upon spatial discretization using the finite element

method, we arrive at the discretized system of equations. We solve the governing equations using a commercial finite element program (Abaqus [39]). Further details about the computational model can be found in [35]. A detailed discussion of different (finite element) modeling techniques for contact acoustic nonlinearity can be found in [18].

Even with a few simplifying assumptions (e.g., homogeneity with smeared properties), the simulation of nonlinear wave propagation in three-dimensional media is computationally expensive. The geometric nonlinearity at the crack interface necessitates the employment of an implicit time-integration scheme for the simulations, which increases the computational burden significantly. For the slab specimens considered in this work, finite element simulation of one VAM test for a 2D domain (Ω representing a vertical cross-section of the slab), which does not fully capture the true (3D) nonlinear wave propagation, takes approximately 1 hour. On the other hand, finite element simulation of a VAM test for a 3D domain (Ω representing the entire slab) takes about 45 hours. One of the goals of this study is to evaluate whether damage index variation pattern (across sensors) learned from the approximate (2D) simulation model could be used for damage localization in real-world (3D) structures, given the prohibitive computational effort of a 3D simulation. To this end, we simulate the governing nonlinear wave physics for a 2D domain. This approximate model is used for generating training data for supervised learning models used for diagnosis and discussed in the next section.

3.3.1.3. Damage localization using the regression model

In this approach, a surrogate model is built for the phenomenon (nonlinear dynamics in the present case) underlying the diagnostic test procedure. This model uses the (candidate) flaw locations and test parameters to predict the damage index (SBSum) value at different sensor locations. Thus, the regression model is a surrogate for the forward model. To perform diagnosis,

an inverse problem needs to be resolved. In this article, the regression model is coupled with a Bayesian estimation algorithm to diagnose the location and extent of damage in the domain of interest. The Bayesian approach represents the state of knowledge about the values of parameters of interest through the use of probability distribution functions. The updated knowledge (i.e., the posterior distribution) about a parameter is obtained by combining prior knowledge (based on intuition, experience, model prediction, prior data, etc.) and observed data. The observed data is included in the (Bayesian) inference algorithm by computing the likelihood of observing the data for a given value of the parameter.

In this work, the location of the centroid of the crack (X_{crack}), the crack length (L_{crack}), and the crack depth (Z_{crack}), are the quantities to be estimated. The Bayesian update equation for this estimation can be written as:

$$P(\boldsymbol{\theta}|SBSum) \propto P(SBSum|\boldsymbol{\theta}) * P(\boldsymbol{\theta}), \tag{9}$$

where $P(\boldsymbol{\theta})$ is the prior distribution of the parameters $\boldsymbol{\theta}$ to be estimated, $P(SBSum|\boldsymbol{\theta})$ is the likelihood, and $P(\boldsymbol{\theta}|SBSum)$ is the posterior distribution of $\boldsymbol{\theta}$. Here $\boldsymbol{\theta}$ is a vector, with three components, i.e., $\theta_1 = X_{crack}$, $\theta_2 = L_{crack}$, $\theta_3 = Z_{crack}$ are the three model inputs to be calibrated (Figure 19). We use uniform priors spanning the extent of the test specimen for P (θ_1), P(θ_2), and P(θ_3) since we do not know the crack location, length, and depth a priori. Using the prediction ANN model and Markov Chain Monte Carlo (MCMC) simulations, the posterior distribution of θ_1, θ_2, and θ_3 is computed. MCMC algorithms are sequential sampling methods commonly used in solving complex integration and optimization problems. Samples from a continuous random variable are used to evaluate a variable's posterior; the Metropolis-Hastings (MH) algorithm is used here for sampling[74], [75].

VAM-based damage localization fuses damage state information obtained from multiple

VAM tests with different test parameters (i.e., excitation frequencies, amplitude ratios, etc.). Bayesian estimation incorporates information from various VAM tests when evaluating a variable's posterior. Damage maps are created for the crack's location by sampling the crack's posteriors and taking an average of the sampled damage locations. This method is computationally expensive since three parameters are estimated in MCMC simulation.

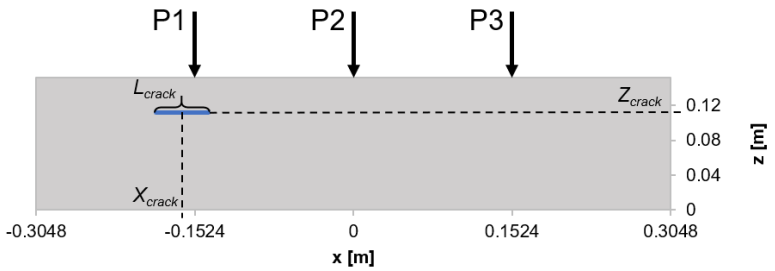


Figure 19: Computational domain with a single crack and geometric parameters that define crack location and size

3.3.1.4. Damage localization using the classification model

This model uses the damage index (SBSum) and other relevant test parameters to determine whether damage is present or absent in the vicinity of the given sensor. The model needs to learn the relationship that separates the damage index observed at a sensor in the vicinity of the damage from a sensor away from the damage while considering the context provided by test parameters and other relevant information (e.g., SBSum values in the neighborhood). In supervised learning for a classifier, the inputs need to be carefully chosen. A large input space may increase the training time, and not including important variables as inputs may lead to poor accuracy.

In a training data set, where multiple inputs and corresponding outputs are available, analysis of variance (ANOVA), a type of statistical hypothesis testing, could be used to determine the

importance of each input parameter. In selecting inputs for the model, we use a p-value threshold of 0.05 in this paper, i.e., if the p-value for any input is less than 0.05, it is included in the model. The damage identification model thus incorporates selected VAM test parameters and the damage metric (SBSum) as the input, and the damage indicator as the output. The damage indicator has a value of 0 or 1, where 1 indicates the presence of damage, and 0 indicates the absence of damage. The training data are obtained from numerical simulations. The details of training data generation are discussed in section 3.4.1.

Due to physical (geometric and material attenuation) and practical (measurement noise, sensor density, equipment limitations, etc.) considerations [35], [48], VAM-based damage localization has to rely on assimilating damage state information obtained from multiple VAM tests with different test parameters. The damage classification model provides damage state information at each sensor for all VAM tests. A methodology is thus needed to assimilate the damage diagnosis for all VAM tests performed on the specimen of interest. To this end, the classification results (damage = 0 or 1) from all VAM tests are averaged at all sensor locations. If this average is greater than 0.5, that location is identified as damaged. In creating the damage map, the region between adjacent sensors that indicate damage is also assumed to be damaged. The centroid of the cluster of adjacent sensors showing signs of damage is used to estimate the crack centroid along the x-axis, and the distance between the extremities of the cluster is used to approximate the crack length. The classification model is thus used to estimate the hidden crack location as well as size.

The classification model only provides damage location along the x-y (horizontal) plane (no estimate of damage depth), whereas the regression model also provides the damage depth (in addition to the damage location in the horizontal plane). The classification model-based diagnosis is computationally inexpensive (i.e., a few seconds to perform diagnosis), whereas the regression

model-based diagnosis is computationally expensive (i.e., about 30-32 hours to perform diagnosis due to the use of Markov chain Monte Carlo sampling for constructing the posterior distribution). A combined approach is thus utilized, which involves obtaining the damage location first on the x-y plane using the classification model and then the damage depth using the regression model (described in the previous subsection). To this end, the parameters θ_1 and θ_2 (defined in section 3.3.1.3 above) are fixed at the estimated crack location and length from the classification model, and Bayesian estimation is performed (using the regression model for likelihood computation) to estimate the third parameter, damage depth, θ_3. This technique is computationally less expensive than the procedure outlined in section 3.3.1.3 since only one parameter is estimated using MCMC simulations.

In summary, two damage diagnosis approaches are investigated (Figure 20). In the first approach, Bayesian estimation is used (along with the trained SBSum regression model) to determine the posterior probability distributions of the crack centroid, length, and depth. The second approach is a combined approach, where the damage classification model is used to determine the centroid for the crack and crack length, and then Bayesian estimation is applied using the regression model to determine the posterior for damage depth. The second approach is computationally cheaper than the first approach.

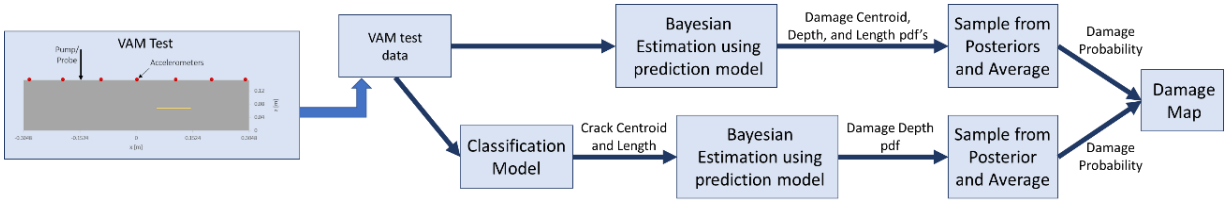


Figure 20: Two different damage diagnosis approaches to determine the damage map. The top row shows the Bayesian Estimation approach using only the prediction model, whereas the bottom row shows the approach using both classification and Bayesian estimation.

3.3.2. Validation Methodology

The proposed methodology is validated by performing VAM tests on a cement slab containing pockets of reactive aggregate at known locations (EXP-A). The slab was cast and cured in an aggressive (ASR-inducing) environment in the laboratory. The ASR gel formation and the subsequent expansion around the reactive aggregates induce internal macro/micro-cracking in the slab. The occurrence of ASR in the slab was confirmed by performing chemical tests on the gel effluent [48]. VAM tests were then carried out to collect test data, which are used in this paper for validation of the ML models. After completing the VAM tests, cylindrical cores were drilled out of the slab, and a petrographic examination of the cores was performed. The petrographic study revealed the approximate locations of the internal micro/macro cracks due to ASR. The details of slab casting, curing, chemical testing, petrographic study, as well as the VAM tests performed on this slab, can be found in section 2.4.

The VAM test data for this slab was collected using a rectangular grid of sensors (accelerometers) placed on the surface of the slab. The ground truth about damage presence/absence for the laboratory slab specimen is obtained by performing petrographic study [48] and destructive testing (breaking the slab open) as described above. As the data used for training the ML models are obtained using a 2D finite element model, the damage in the cement slab specimen is diagnosed along each vertical slice containing a row of sensors on the top edge. That is, each row of sensors installed on the surface of the slab is evaluated separately, and then the results are synthesized to create the damage map. Based on the ground truth regarding damage presence/absence, the accuracy, sensitivity and specificity of the proposed methodology, when applied to the experimental specimen, is evaluated. The mathematical definitions of these metrics

are given by:

$$Accuracy = \frac{True\ positives + True\ Negatives}{Total\ Data\ Points}\ , \tag{10}$$

$$Sensitivity = \frac{True\ positives}{True\ Positives + False\ negatives}\ , \tag{11}$$

$$Specificity = \frac{True\ negatives}{False\ Positives + True\ negatives}\ . \tag{12}$$

A practical challenge in using simulation-based training data for diagnosing damage in a laboratory specimen is that the exact magnitude of the dual-frequency force applied in laboratory experiments is unknown. Furthermore, the geometric attenuation in 2D models vs. 3D specimens is different in magnitude, and it is difficult to quantify this ratio for all sensors using a simple analysis. As a result, the damage indices (SBSum values) for training data and real-world data may differ by an order of magnitude. We overcome this challenge by leveraging the key principle underlying VAM-based damage localization: the methodology relies on *relative* values of the damage index at all sensors in a given VAM test (and not on the absolute values). Keeping this in mind, the SBSum values obtained from the 2D model runs are simply scaled up or down to match the experimentally obtained SBSum values. A series of FE analyses are performed at multiple excitation amplitudes, and a linear regression model showing the relationship between the excitation amplitude and the SBSum is built. The FEA data can then be appropriately scaled so that the experimental SBSum values and FEA SBSum values are of the same order of magnitude.

3.4. Supervised Machine Learning

3.4.1. Training data generation

As described in Section 3.3, two-dimensional (2D) finite element models simulating VAM

tests for multiple crack sizes and locations are used to generate the training data. ANN models are trained using data obtained from this 2D analysis. The domain geometry and the pump and probe excitation locations used in the numerical simulations are shown in Figure 19. The pump and probe actuators are placed at the same location for each simulation. The computational domain is 60.96 cm wide and 15.24 cm thick. It contains a single crack, placed at different locations and of varying lengths. All of the locations for the single cracks used are displayed in Figure 22. The commercial finite-element program Abaqus was used to perform numerical simulations. The domain was discretized using a structured mesh of 8-noded finite elements. An implicit scheme was employed for time integration of the governing equations and ensured that at least ten computational nodes were available over the smallest wavelength (0.068 m). The interaction at the crack interface was modeled using a hard, frictionless contact condition. The material properties used in these simulations are given in Table 5. The parameters used in conducting VAM tests on this numerical specimen can be found in

Table 6. Our past research [35] reported studies that evaluated the diagnostic performance of VAM tests for different ratios of flaw size to probing wavelength. The study showed that VAM-based hidden crack diagnosis shows acceptable values of sensitivity and specificity when the crack length is approximately equal to or longer than the smallest probing wavelength (0.068 m in this case). Hence, most of the crack lengths in

Table 6 are greater than or equal to 0.06 m. Acceleration time-history data is recorded at sensors located along the top edge of the specimen and is used to calculate SBSum values [35].

Since the proposed methodology is to be validated using data obtained from actual VAM experiments performed on a cement slab, the SBSum values for a given laboratory test configuration (probe excitation amplitude, probe excitation location, etc.) are compared with

corresponding SBSum values from the numerical simulations. It is found that the experimentally measured SBSum values are higher than the 2D FEA-based values. These higher SBSum values are due to the physics model discrepancy, as well as the difference in input excitation magnitudes. In this work, we do not quantify the physics model discrepancy. Instead, we (indirectly, through scaling of the data) adjust the excitation amplitude for simulations to ensure that SBSum values for training (simulation) data and validation (experimental test) data are similar in magnitude. Note that the damage diagnosis process depends on the relative values of the damage index (SBSum), and if all SBSum values for all tests are multiplied by a scalar, the diagnosis results will not be affected. To establish that input excitation magnitudes indeed affect SBSum values, the correlation between SBSum values in a VAM test, and the probe excitation amplitudes (the pump excitation amplitude was constant across all tests) is studied for numerical simulations (see Figure 21). These two quantities show a strong positive correlation, with a Pearson correlation coefficient of 1. This implies that SBSum values of magnitudes similar to those obtained from simulation data could be obtained using laboratory tests by reducing the probe excitation amplitude by a scalar factor (less than 1). As re-testing the specimen with different probe excitation amplitudes is a time-consuming process, in this work, we scaled the simulation-based SBSum values by a scalar factor (greater than 1). Given the range of experimental SBSum values, the approximate value of this the scalar multiplier is found to be 1000. Since the relationship between the excitation amplitude and SBSum is linear and shows a very strong positive correlation, all the SBSum values from simulation were multiplied by 1000. After the scaling, the simulation-based training data SBSum values and laboratory test SBSum values exhibit the same order of magnitude.

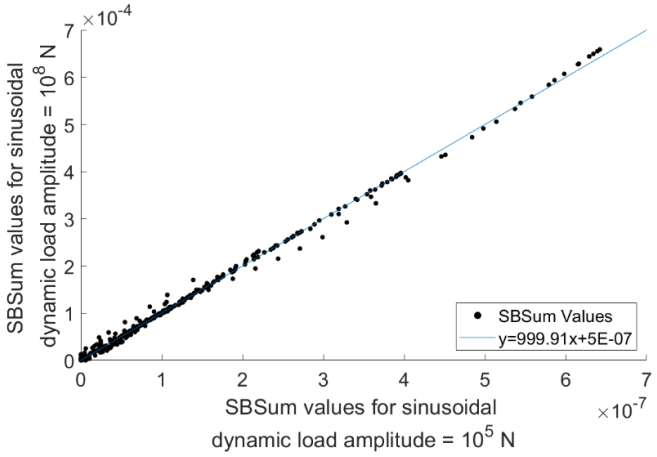


Figure 21: Scaling the 2D simulation model outputs: data and regression fit.

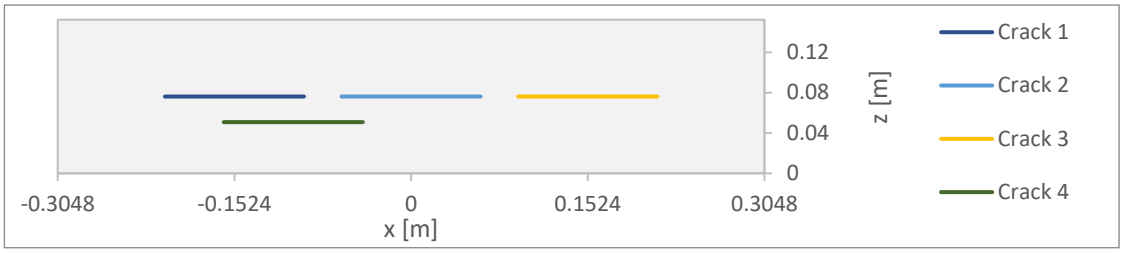
Table 5: Material properties used in the 2D FEA model.

Material property	Value
Young's modulus (E)	27 GPa
Density (ρ)	0.15
Poisson's ratio (ν)	2400 kg/m^3
Mass proportional Rayleigh damping parameter (a)	2120.04
Stiffness proportional Rayleigh damping parameter (b)	1.787× 10^{-7}

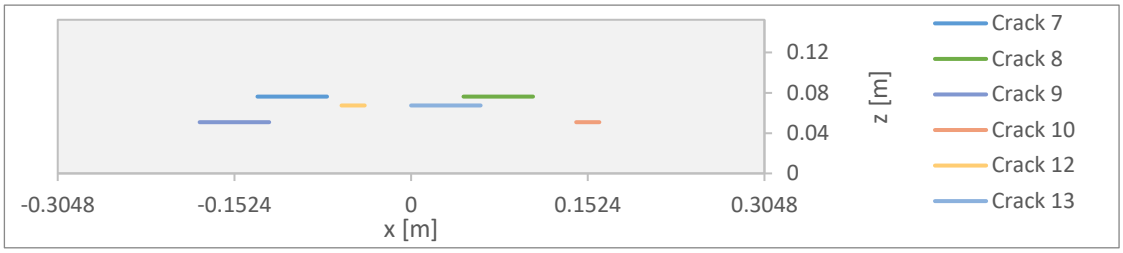
Table 6: Test parameters for the 2D FEA model.

Crack #	Centroid (x-direction) [m]	Length [m]	Depth from Surface (z-direction) [m]	Amplitude Ratio (probe/pump) R_{amp}	Frequency Ratio (probe/pump) R_{freq}	Pump/Probe Actuator Location [m]
1	-0.1524	0.12	0.0762	1, 0.5, 0.2, 0.1	10,12,14,16,18,20	-0.1524, 0.0, 0.1524
2	0	0.12	0.0762	1, 0.5, 0.2, 0.1	10,12,14,16,18,20	-0.1524, 0.0, 0.1524
3	0.1524	0.12	0.0762	1, 0.5, 0.2, 0.1	10,12,14,16,18,20	-0.1524, 0.0, 0.1524
4	-0.1016	0.12	0.1016	1, 0.5, 0.2, 0.1	10,12,14,16,18,20	-0.1524, 0.0, 0.1524
6	-0.1016	0.12	0.0762	1, 0.5, 0.2, 0.1	10,12,14,16,18,20	0.0
7	-0.1025	0.06	0.0762	1, 0.5, 0.2, 0.1	10,12,14,16,18,20	-0.1524, 0.0, 0.1524
8	0.0752	0.06	0.0762	1, 0.5, 0.2, 0.1	10,12,14,16,18,20	-0.1524, 0.0, 0.1524
9	-0.1524	0.06	0.1016	1, 0.5, 0.2, 0.1	10,12,14,16,18,20	-0.1524, 0.0, 0.1524
10	0.1524	0.02	0.1016	1, 0.5, 0.2, 0.1	16,18,20	-0.1524, 0.0, 0.1524
11	0.1016	0.09	0.0762	1, 0.5, 0.2, 0.1	16,18,20	-0.1524, 0.0, 0.1524
12	-0.05	0.02	0.085	1, 0.5, 0.2, 0.1	18,20	0.0
13	0.03	0.06	0.085	1, 0.5, 0.2, 0.1	18,20	0.0
14	0.1016	0.09	0.085	1, 0.1	18,20	0.1524
15	0.1016	0.09	0.095	1, 0.1	18,20	0.1524
16	0.1016	0.09	0.1016	1, 0.1	18,20	0.1524
17	0.1016	0.09	0.0508	1, 0.1	20	0.1524
18	0.1016	0.09	0.02	1,0.1	18,20	-0.1524, 0.0, 0.1524
19	-0.1016	0.09	0.02	1, 0.5, 0.2, 0.1	18	-0.1524

(a)



(b)



(c)

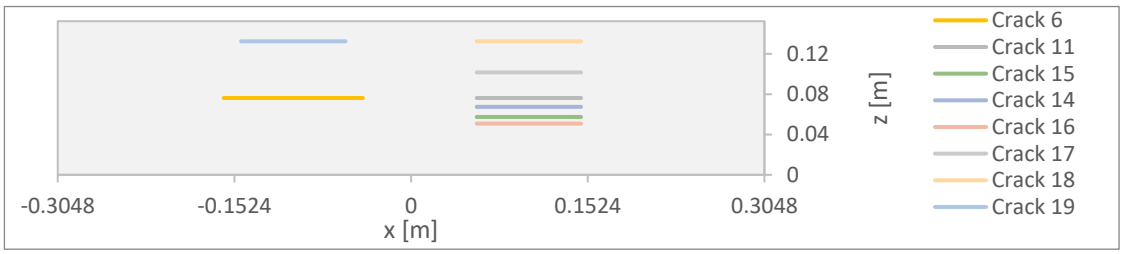


Figure 22: The domain geometry and crack locations used in the numerical simulations, only one crack is present in the specimen domain for a given simulation.

3.4.2. ANN model training evaluation

As mentioned in Section 3.3, two feed-forward ANN models are constructed. The first model estimates the SBSum value for a candidate crack location. The second model classifies the given sensor as showing the presence or absence of damage. In training both the neural networks, first, the weights were randomly initialized as values between -1 and 1 for the given inputs. Both models were trained using a learning rate of 0.1. The number of layers and the number of neurons in each layer are optimized to decrease errors in predicting the testing data. In order to train the

classification and regression models, the available (115,538) data points for cracks 1-4 and 6-19 (Figure 22) were randomly split into training (70%), testing (15%), and validation data (15%). Due to the random splitting of the data, information for any one given crack location could be present in all three data sets since the 115,538 data points were obtained using combinations of 6 excitation frequency ratios, 4 excitation amplitude ratios, 3 pump/ probe locations, and SBSum values at varying numbers of sensor locations (149 to 197 sensors depending on the number of nodes on the top edge (mesh-dependent)) in each VAM test simulation for 18 different crack locations.

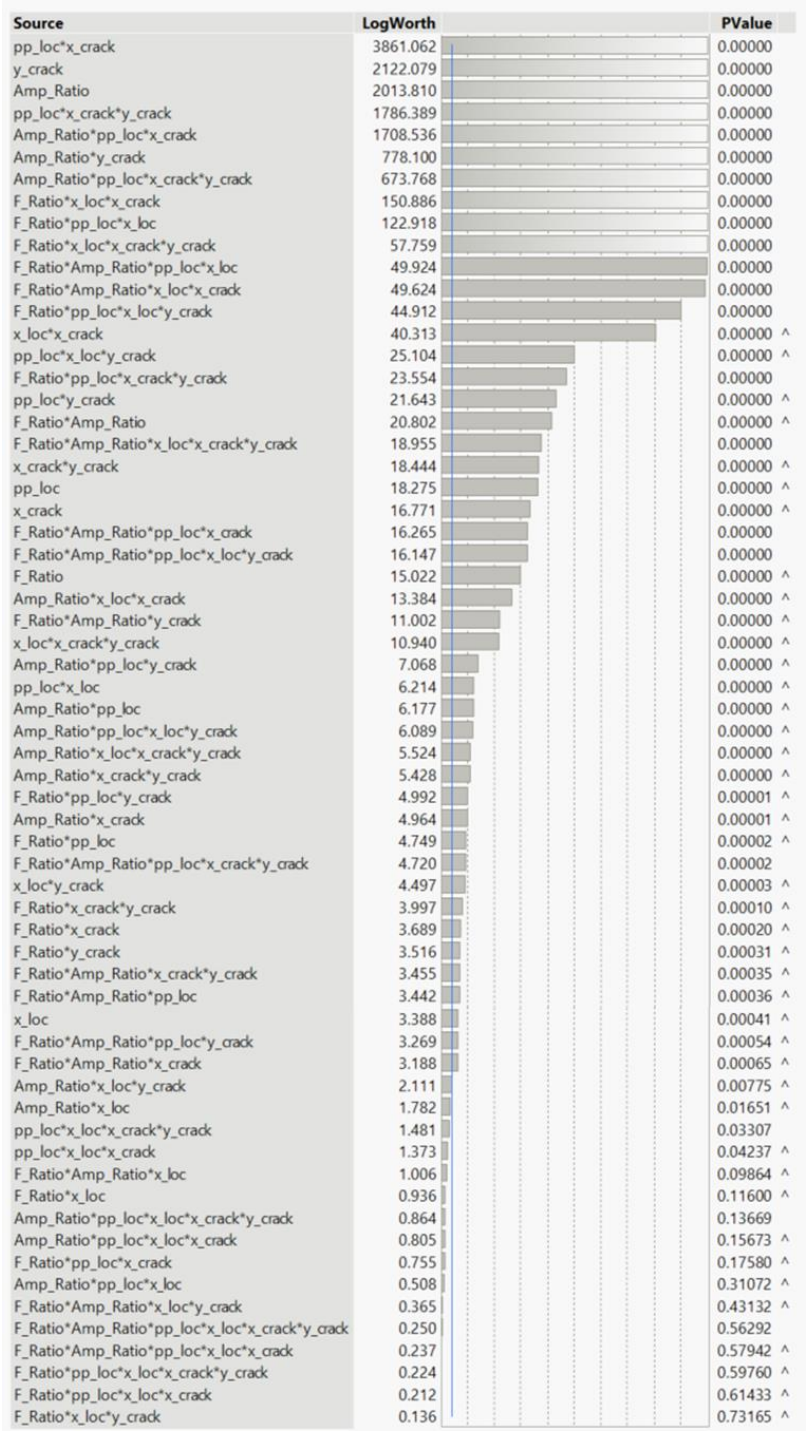
For the regression neural network, the accuracy is quantified using the mean squared error (MSE). MSE is the average squared difference between the ANN-predicted values of SBSum and the FEA-generated values. For the damage classification neural network, the accuracy is calculated using the metrics defined in equations (10), (11) and (12). A grid search is used in optimizing the hyperparameters of the ANN models; the set of hyperparameters that yield the highest cross-validation accuracy are selected.

3.4.2.1. Regression Model

The inputs for this model include the frequency ratio (R_{freq}) of the pump and probe excitations, the amplitude ratio (R_{amp}) of the pump and probe excitations, the location ($X_{crack,}$ Figure 19) of the centroid of the crack, the depth of the crack from the surface (Z_{crack}, Figure 19), the crack length, the location of the pump/probe excitations (X_{pp}) and the location of the sensor (X_{sensor}). See

Table 6 for various combinations of the inputs. The output is the SBSum value. In the data pre-processing phase, all inputs and outputs are normalized to zero mean and unit standard deviation. ANOVA was conducted on this data to determine the importance of each input parameter. Figure 23 displays the p-values for each input and their interactions. All the inputs had p-values less than $4x10^{-4}$, confirming that all 7 inputs are important and need to be used in the

regression model. The prediction feed-forward neural network to estimate SBSum (Figure 24) consists of two layers. The first layer has 16 nodes, and the second has 8. The Levenberg-Marquardt method was used as the training algorithm with a least squares loss function [76], [77]. This network converged after 488 epochs. Early stopping [78], using information about the error of the cross-validation data, was used to prevent the ANN model from overfitting to the training data. Using the 15% randomly selected cross-validation data, the cross-validation error was found to be approximately 25%.



Source	LogWorth	PValue
pp_loc*x_crack	3861.062	0.00000
y_crack	2122.079	0.00000
Amp_Ratio	2013.810	0.00000
pp_loc*x_crack*y_crack	1786.389	0.00000
Amp_Ratio*pp_loc*x_crack	1708.536	0.00000
Amp_Ratio*y_crack	778.100	0.00000
Amp_Ratio*pp_loc*x_crack*y_crack	673.768	0.00000
F_Ratio*x_loc*x_crack	150.886	0.00000
F_Ratio*pp_loc*x_loc	122.918	0.00000
F_Ratio*x_loc*x_crack*y_crack	57.759	0.00000
F_Ratio*Amp_Ratio*pp_loc*x_loc	49.924	0.00000
F_Ratio*Amp_Ratio*x_loc*x_crack	49.624	0.00000
F_Ratio*pp_loc*x_loc*y_crack	44.912	0.00000
x_loc*x_crack	40.313	0.00000 ^
pp_loc*x_loc*y_crack	25.104	0.00000 ^
F_Ratio*pp_loc*x_crack*y_crack	23.554	0.00000
pp_loc*y_crack	21.643	0.00000 ^
F_Ratio*Amp_Ratio	20.802	0.00000 ^
F_Ratio*Amp_Ratio*x_loc*x_crack*y_crack	18.955	0.00000
x_crack*y_crack	18.444	0.00000 ^
pp_loc	18.275	0.00000 ^
x_crack	16.771	0.00000 ^
F_Ratio*Amp_Ratio*pp_loc*x_crack	16.265	0.00000
F_Ratio*Amp_Ratio*pp_loc*x_loc*y_crack	16.147	0.00000
F_Ratio	15.022	0.00000 ^
Amp_Ratio*x_loc*x_crack	13.384	0.00000 ^
F_Ratio*Amp_Ratio*y_crack	11.002	0.00000 ^
x_loc*x_crack*y_crack	10.940	0.00000 ^
Amp_Ratio*pp_loc*y_crack	7.068	0.00000 ^
pp_loc*x_loc	6.214	0.00000 ^
Amp_Ratio*pp_loc	6.177	0.00000 ^
Amp_Ratio*pp_loc*x_loc*y_crack	6.089	0.00000 ^
Amp_Ratio*x_loc*x_crack*y_crack	5.524	0.00000 ^
Amp_Ratio*x_crack*y_crack	5.428	0.00000 ^
F_Ratio*pp_loc*y_crack	4.992	0.00001 ^
Amp_Ratio*x_crack	4.964	0.00001 ^
F_Ratio*pp_loc	4.749	0.00002 ^
F_Ratio*Amp_Ratio*pp_loc*x_crack*y_crack	4.720	0.00002
x_loc*y_crack	4.497	0.00003 ^
F_Ratio*x_crack*y_crack	3.997	0.00010 ^
F_Ratio*x_crack	3.689	0.00020 ^
F_Ratio*y_crack	3.516	0.00031 ^
F_Ratio*Amp_Ratio*x_crack*y_crack	3.455	0.00035 ^
F_Ratio*Amp_Ratio*pp_loc	3.442	0.00036 ^
x_loc	3.388	0.00041 ^
F_Ratio*Amp_Ratio*pp_loc*y_crack	3.269	0.00054 ^
F_Ratio*Amp_Ratio*x_crack	3.188	0.00065 ^
Amp_Ratio*x_loc*y_crack	2.111	0.00775 ^
Amp_Ratio*x_loc	1.782	0.01651 ^
pp_loc*x_loc*x_crack*y_crack	1.481	0.03307
pp_loc*x_loc*x_crack	1.373	0.04237 ^
F_Ratio*Amp_Ratio*x_loc	1.006	0.09864 ^
F_Ratio*x_loc	0.936	0.11600 ^
Amp_Ratio*pp_loc*x_loc*x_crack*y_crack	0.864	0.13669
Amp_Ratio*pp_loc*x_loc*x_crack	0.805	0.15673 ^
F_Ratio*pp_loc*x_crack	0.755	0.17580 ^
Amp_Ratio*pp_loc*x_loc	0.508	0.31072 ^
F_Ratio*Amp_Ratio*x_loc*y_crack	0.365	0.43132 ^
F_Ratio*Amp_Ratio*pp_loc*x_loc*x_crack*y_crack	0.250	0.56292
F_Ratio*Amp_Ratio*pp_loc*x_loc*x_crack	0.237	0.57942 ^
F_Ratio*pp_loc*x_loc*x_crack*y_crack	0.224	0.59760 ^
F_Ratio*pp_loc*x_loc*x_crack	0.212	0.61433 ^
F_Ratio*x_loc*y_crack	0.136	0.73165 ^

Figure 23: p-values results for each input and their interactions using ANOVA.

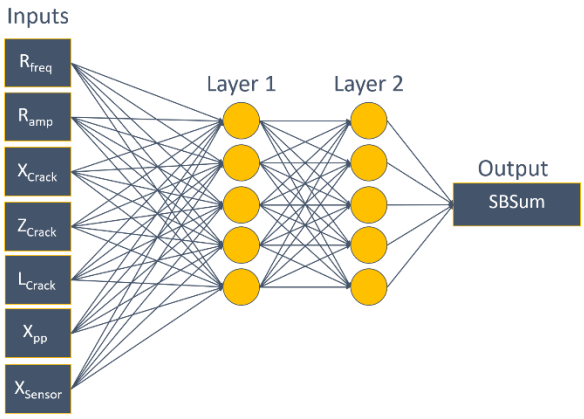
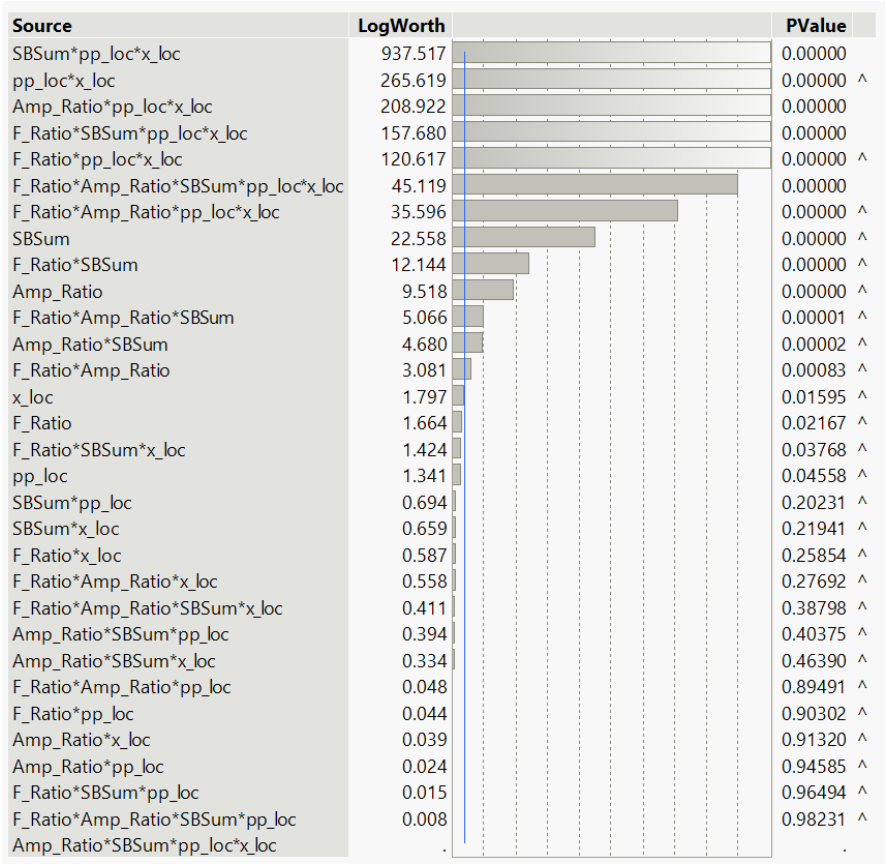


Figure 24: Regression model with seven inputs (VAM test and damage parameters), one output (SBSum), and two hidden layers of five neurons each.

3.4.2.2. Classification Model

ANOVA was conducted on many possible model inputs, including SBSum, excitation frequency ratio (R_{freq}), excitation amplitude ratio (R_{amp}), pump and probe excitation actuator locations (X_{pp}), mean SBSum for each test (μ_{SBSum}), and sensor location (X_{sensor}), with the output being the damage indicator {0,1}. Figure 25 displays the p-values for each input and their interactions. Although the X location of the sensor, pump and probe location, and frequency ratio are not significant for damage prediction on their own, their combined effects are very significant in the damage prediction model. All of these inputs had p-values smaller than 0.05, indicating that all were significant in damage estimation. The input parameters with the smallest p-values were SBSum and R_{amp}, which had values smaller than $1x10^{-5}$. The damage classification model thus incorporates selected VAM test parameters, the damage metric (SBSum), and the mean SBSum of each sensor for every VAM test, as inputs to the model. The use of mean SBSum allows the model to learn if a sensor had a relatively high SBSum compared to other sensors in the same VAM test.



Source	LogWorth	PValue
SBSum*pp_loc*x_loc	937.517	0.00000
pp_loc*x_loc	265.619	0.00000 ^
Amp_Ratio*pp_loc*x_loc	208.922	0.00000
F_Ratio*SBSum*pp_loc*x_loc	157.680	0.00000
F_Ratio*pp_loc*x_loc	120.617	0.00000 ^
F_Ratio*Amp_Ratio*SBSum*pp_loc*x_loc	45.119	0.00000
F_Ratio*Amp_Ratio*pp_loc*x_loc	35.596	0.00000 ^
SBSum	22.558	0.00000 ^
F_Ratio*SBSum	12.144	0.00000 ^
Amp_Ratio	9.518	0.00000 ^
F_Ratio*Amp_Ratio*SBSum	5.066	0.00001 ^
Amp_Ratio*SBSum	4.680	0.00002 ^
F_Ratio*Amp_Ratio	3.081	0.00083 ^
x_loc	1.797	0.01595 ^
F_Ratio	1.664	0.02167 ^
F_Ratio*SBSum*x_loc	1.424	0.03768 ^
pp_loc	1.341	0.04558 ^
SBSum*pp_loc	0.694	0.20231 ^
SBSum*x_loc	0.659	0.21941 ^
F_Ratio*x_loc	0.587	0.25854 ^
F_Ratio*Amp_Ratio*x_loc	0.558	0.27692 ^
F_Ratio*Amp_Ratio*SBSum*x_loc	0.411	0.38798 ^
Amp_Ratio*SBSum*pp_loc	0.394	0.40375 ^
Amp_Ratio*SBSum*x_loc	0.334	0.46390 ^
F_Ratio*Amp_Ratio*pp_loc	0.048	0.89491 ^
F_Ratio*pp_loc	0.044	0.90302 ^
Amp_Ratio*x_loc	0.039	0.91320 ^
Amp_Ratio*pp_loc	0.024	0.94585 ^
F_Ratio*SBSum*pp_loc	0.015	0.96494 ^
F_Ratio*Amp_Ratio*SBSum*pp_loc	0.008	0.98231 ^
Amp_Ratio*SBSum*pp_loc*x_loc	.	.

Figure 25: p-values results for each input and their interactions using ANOVA.

All crack sizes used for generating the training data are smaller than about a quarter (0.1524 m) of the computational domain length (0.6096 m). This results in a fewer number of nodes lying directly above a crack in the training data, which implies that there is a class imbalance in training data. Specifically, about 84 % of nodes in the training data belong to class 0 (damage absent), whereas only 16 % of nodes belong to the other class 1 (damage present). To alleviate the effects of class imbalance on model training, we downsample the training data. If N_D denotes the number of nodes (sensors) directly above a crack (showing the presence of damage (Figure 26)), we sampled the same number of sensors from the areas in which no damage was present (A and B in Figure 26). The number of sensors located in zone D is approximately equal to the number of

sensors selected (downsampled) from zones A and B combined ($N_D \approx N_A + N_B$). The total number of data points after downsampling from the 18 crack locations is 38,370.

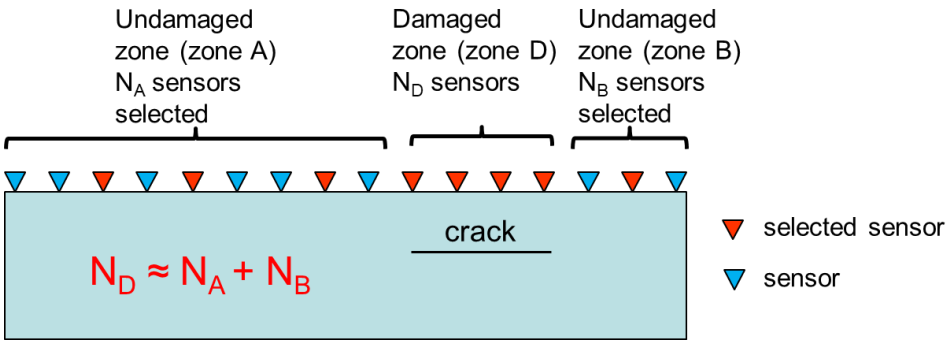


Figure 26: Downsampling scheme used to alleviate class imbalance in the training data.

The ANN used for classification was a supervised pattern recognition network. The network classified the outputs as a probability of damage. In training the neural network, first the weights were initialized for the inputs. Next the signals were activated for each input—this step is referred to as the forward propagation of the model. The final step of the network is back propagation. In this step, the gradients in the model were computed and stochastic gradient descent was used to calibrate the weights over each iteration. The model had a learning rate of 10%, meaning the weights were updated for every 10% of the data. This entire process of forward and backward propagation was repeated until convergence [79].

The damage classification feed-forward neural network (Figure 27) consists of two layers with 32 nodes in the first layer and 16 in the second layer. The conjugate gradient method with Powell/Beale Restart algorithm [80] was used as the training algorithm and cross-entropy loss function was used. Cross entropy loss measures the performance of a classification model with probabilities between 0 and 1. It is a logarithmic function that measures the distance between the

estimated probability and the true value. Since our true values for the outputs were 0 and 1, the cross entropy can be treated like a percent error between the estimated and actual outputs. The SoftMax [81] activation function was applied to the output. This model converged at 306 epochs and had a training accuracy of 84%, sensitivity of 90%, specificity of 78%, as well as a cross-validation accuracy of 84% (using the 15% randomly selected cross-validation data).

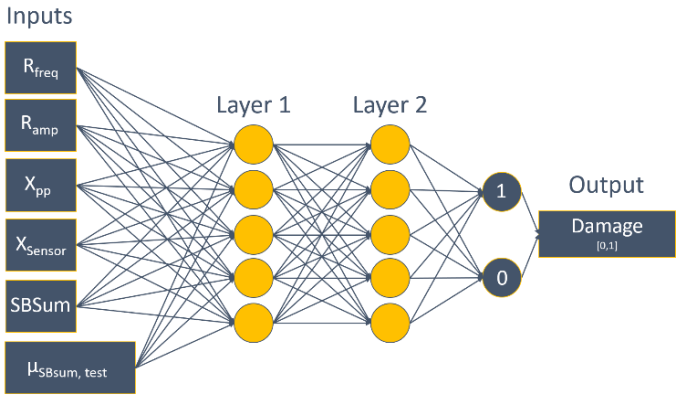


Figure 27: Classification model with six inputs (VAM test parameters and data), an output (damage present (1) or absent (0)), and two hidden layers.

3.5. Validation Results

The prediction and classification models developed in Section 3.4 are first verified using numerical experiments (with separately generated comparison data using two-dimensional FE models). The models are then used to diagnose ASR-induced cracks in a cement slab specimen and validated with petrographic analysis of cores taken from the slab specimen.

3.5.1. Verification of diagnostic methodology using simulation experiments

The proposed physics-informed ML models were first validated by using simulation data for a crack of size and location that was not used in the models' training. The location of this crack and testing parameters are described in Figure 19 and Table 7. For both of the diagnosis methodologies (Figure 20), data from 15 sensor locations on the top edge of the computational domain (15 equally spaced finite element nodes) is used. Note that the wavelengths of shear waves corresponding to the probing frequencies (18 kHz – 22 kHz) are in the range of 0.083 m – 0.068 m. According to the analysis performed in [35], the model-free method is able to provide a sensitivity of around 0.45 – 0.60 for this shear wavelength to crack length ratio.

Table 7: Test parameters used in the 2D FE model for validation test data generation for crack 22.

Parameter	Values
Frequency Ratio (probe/pump) (R_{freq})	18,20,22
Amplitude Ratio (probe/pump) (R_{amp})	1, 0.5, 0.2, 0.1
Pump/Probe Location (x_{pp}) (m)	-0.1524, 0, 0.1524
Crack Centroid (X_{crack}) (m)	-0.1448
Crack Depth from Surface (Z_{crack}) (m)	0.0508
Crack Size (L_{crack}) (m)	0.03

3.5.1.1. Damage localization using the SBSum regression model

Bayesian estimation, with the SBSum regression model, was used to determine the posterior probability distributions of the crack centroid, length, and depth of Crack 22. The prior distribution for the centroid was uniformly distributed over the entire length of the specimen (-.03048, 0.3048); the prior for crack size was assumed uniform for values between 0.01 to 0.3048 m (half of the specimen length), and the prior for the depth was uniformly distributed over the thickness of the

entire slab. 3000 MH samples were used in MCMC simulations with a 20% burn-in. Gaussian measurement noise was assumed with zero mean and a standard deviation of 0.1. Simulations were conducted using 15 sensor locations along the surface (Figure 28).

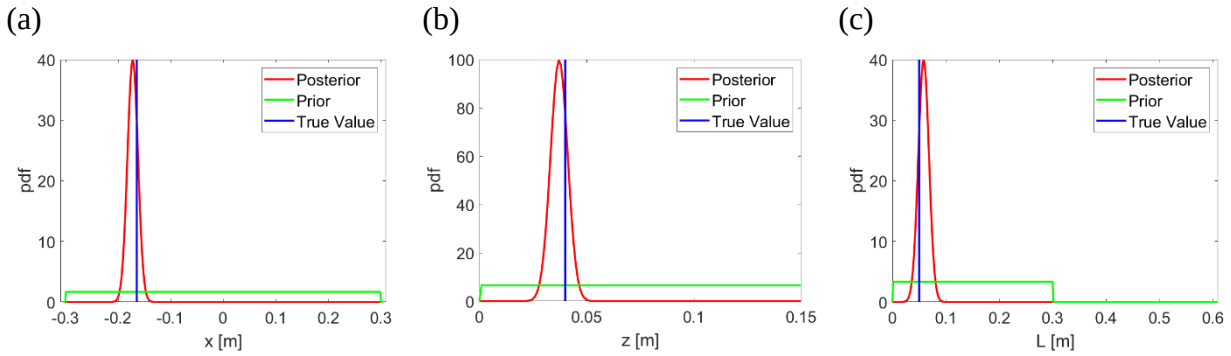


Figure 28: Bayesian estimation of damage for validation crack 22 at 15 sensors; (a) Damage centroid probability density functions (pdfs) (prior and posterior); (b) Damage depth pdfs (prior and posterior); (c) Damage length pdfs (prior and posterior).

With the results of the above Bayesian damage diagnosis, the posteriors of the centroid and crack length are sampled 1000 times to create 1000 different damage maps. Every sensor located within half the length of the crack on each side of the centroid is considered to be damaged (1), and all other locations are assumed to be undamaged (0). These damage maps are averaged to obtain an overall damage probability (Figure 29). The resulting damage map shows remarkably accurate crack localization and crack length estimation results for crack 22.

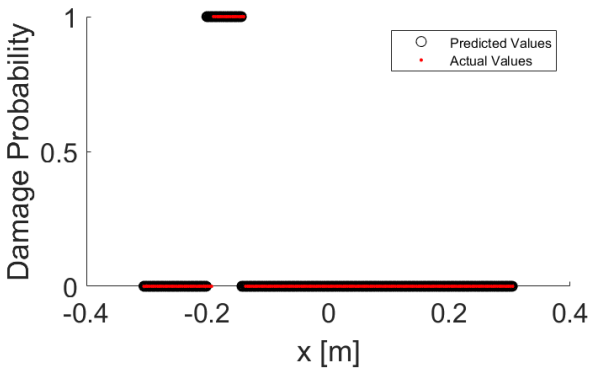


Figure 29: Damage probability estimation using Monte Carlo samples of the posterior distributions for crack centroid and crack length (validation test, crack 22).

3.5.1.2. Damage localization using the classification model

The classification model is constructed to determine crack centroid and crack length, as described in Sections 2 and 3. This model used all R_{freq}, all R_{amp}, and all x_{pp} for crack 22. Two models were built and validated using crack 22. One model was trained using all available data, the other was trained only with the downsampled data. The model trained with the downsampled data was selected as the classification model given the validation accuracy and high sensitivity (Table 8).

Table 8: Model structure and validation test results of the two damage classification models for crack 22.

Model Label	Inputs						Downsampling	Validation test results		
	$f_{probe,in}/f_{pump,in}$	$A_{probe,in}/A_{pump,in}$	X_{pp}	X_{Sensor}	SBSum	μ_{SBSum}		Accuracy	Sensitivity	Specificity
A	X	X	X	X	X	X		0.89	0.61	0.92
B	X	X	X	X	X	X	X	0.76	1.00	0.73

Estimation of the crack centroid and length were conducted using the downsampled classification model described in section 3.4.2.2 and Model B in Table 8, and validated on crack 22. The damage identification results using the deterministic classification output (i.e., damage absent (0) or present (1)) were averaged for all 36 different VAM tests (Table 7) at the 15 sensors (Figure 30 (a)). Additionally, diagnosis was performed by averaging the probability values provided by the SoftMax layer in the classification neural network (results shown in Figure 30 (a)). It can be seen in Figure 30 that considering the probability obtained from the SoftMax layer, the overall diagnosis results do not change significantly compare with Figure 30(a), where deterministic classification results are averaged. In either case (Figure 30 (a)) using a probability threshold (e.g., 0.5) provides the same damage location/extent result. The estimated crack centroid is -0.147 m (-0.1448 m), and the length is 0.123 m (0.03 m).

Next, per the combined approach described in section 3.3.1.4, Bayesian estimation of damage depth used the SBSum values for the tests using R_{freq} of 20, all amplitudes, and all x_{pp} for the validation crack. The estimated crack centroid and length from the classification model were also used to calculate the posterior for the crack depth. The crack depth prior was assumed to be uniformly distributed over the entire thickness of the specimen. The model used 1000 MCMC (Metropolis-Hastings) samples and a 20% burn-in. Figure 30(b) shows the results for the pdf of the damage depth using R_{freq} 20, all R_{amp}, and all x_{pp} at 15 sensors. Even for a relatively small (3 cm long) crack, the classification model was able to localize the crack centroid and estimate its size accurately using 15 sensors. The computational cost for the combined approach is about 4 hours, thus providing good savings in computational effort.

(a) (b)

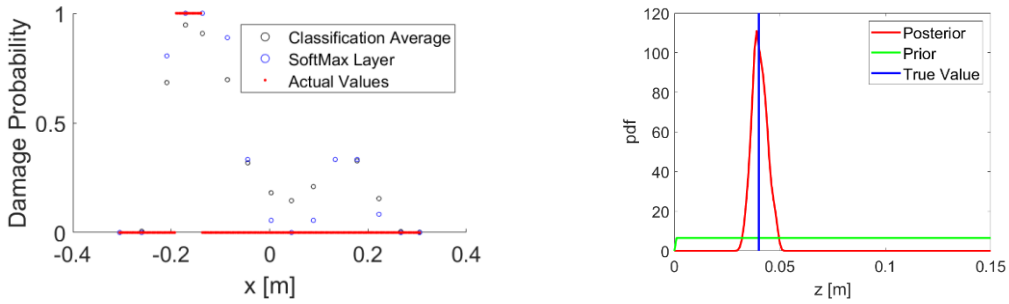


Figure 30: (a) Average of deterministic classification from 36 tests at 15 sensor locations for validation crack 22 and damage probability obtained by the SoftMax layer (average of 36 tests at 15 sensor locations); (b) Damage depth pdfs (prior and posterior) using Bayesian estimation for crack 22. The true depth is 0.0508 m below the top surface.

The crack length reported by the probabilistic diagnostic method depends on a) the accuracy of the method, b) the sensor spacing, and c) the probability threshold chosen to classify the area around a given sensor as damaged. For example, in Figure 31, if two sensors lie directly above a 0.03 m long crack (labeled B and C in Figure 31), when the sensor spacing is 0.04 m, and if two additional neighboring sensors (labeled A and D in Figure 31) show probability of damage slightly higher than the threshold, then a region of length 0.12 m (three times the sensor spacing) would be identified as damaged. In VAM tests, sensors in the neighborhood of the damage (but not directly above the damage) typically show higher SBSum values. Thus, the false positive results for these sensors have a relatively higher likelihood. The regression model, which is used to determine the crack depth, has learned that false positive at a sensor in the neighborhood of the crack is a high likelihood event. Hence it gives less importance to the (potentially erroneous, longer) crack length provided as input, and yet arrives at the correct depth estimate.

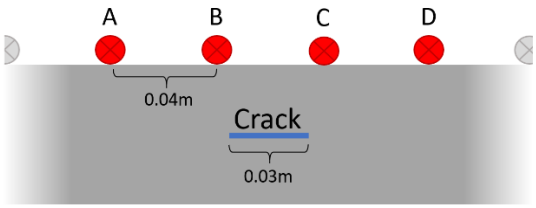


Figure 31: Sensors spaced every 0.04m on the surface of a specimen containing a 0.03m crack.

3.5.2. Validation of diagnostic methodology using EXP-A

The proposed methodology is validated by performing physical VAM tests on a cement slab (dimensions 60.96 cm X 60.96 cm X 15.24 cm) containing pockets of reactive (ASR-inducing) aggregate or silica (EXP-A). The details of the concrete test specimen and VAM tests performed on this specimen can be found in section 2.4. PCB Piezotronics accelerometers (model numbers PCB352C68 and PCB352C65) were used to conduct VAM tests. The pump and the probe actuators (PCB Piezotronics piezoelectric disk actuators, model number PCB712A02) were placed at the center of each quadrant and the center of the specimen. The VAM tests used five different pump and probe actuator locations, four R_{amps} (1, 0.5, 0.2, 0.1), and six different probing frequencies (16, 17, 18, 19, 20, 21 kHz), with a pumping frequency of 920 Hz. For each test at each accelerometer, acceleration response is measured five times, the duration of each (acceleration response) record is 0.2 s, and the sampling frequency is 128 kHz. Four cores were taken in the center of each of the four quadrants of the specimen. The locations of these cores, as well as the approximate locations of the aggregates after curing are shown in Figure 32 (a). (The approximate locations of the aggregates were determined by breaking the specimen after taking the cores and visually observing the approximate locations of the aggregates in each area). Alkali-silica reaction damage was found in cores Q3 and Q4. Q1 failed to contain any aggregates at their original locations (because of the shifting of the aggregates away from the center during casting), so there was no ASR damage

present in the core taken from Q1. Core Q2 did show signs of ASR; however, the area occupied by silica showed a large void that would not be diagnosed using VAM (since VAM is only suitable for detecting breathing cracks, not voids [48]). Damage in this quadrant could be localized along the edges of the aggregate area, where the crack is small enough to breathe open and close due to the applied excitation.

Note that the ML models in section 3.4 were trained using 2D FEA analyses. To apply these models to the 3D cement slab specimen, six slices were taken along sensor locations (parallel to the x-axis) in the specimen (Figure 7 and Figure 32 (a)). There are 7 sensors located on the top surface of each vertical slice. The damage probability using model-free (averaged likelihood) approach discussed in [48] is provided in Figure 32 (c). Damage location identification using only the regression model to estimate all three parameters and damage localization using both the classification model (crack centroid and length) and the regression model (damage depth) is performed for each slice.

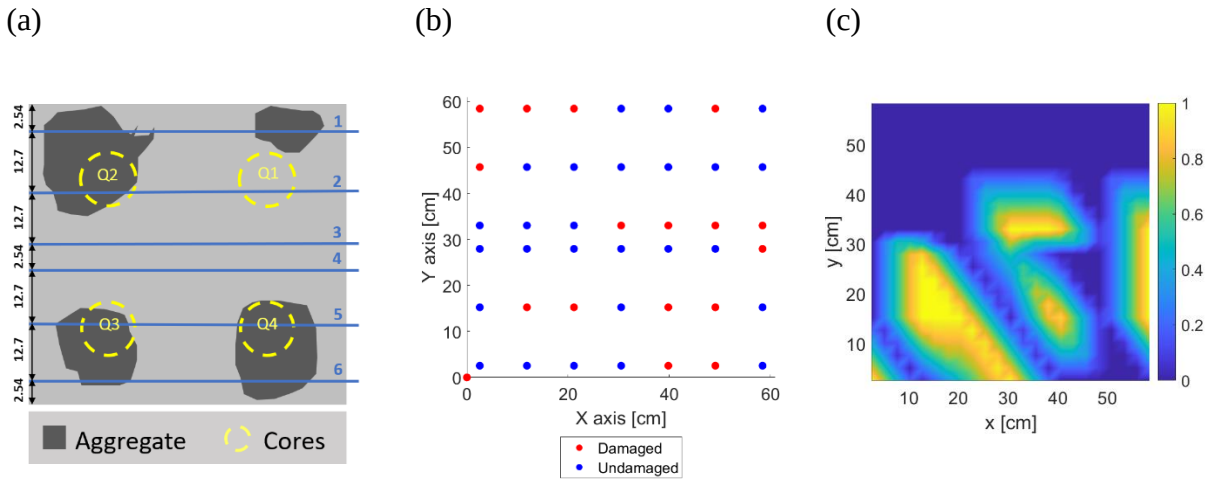


Figure 32: (a) Vertical slice locations; (b) Damage indicators at sensors; (c) Damage probability estimated using model-free (averaged likelihood) approach [48]).

In classification models, the assessment of model performance is given by metrics such as accuracy, sensitivity etc., based on the confusion matrix. Note that the proposed damage diagnosis methodology provides a probabilistic estimate of damage presence at/near a given sensor. The ground truth about damage presence in our numerical or laboratory experiments is deterministic (a fixed value at each sensor, i.e., either damaged or undamaged), as shown for the laboratory (slab) specimen (Figure 32 (b)).

Bayesian estimation using the regression model was performed for each slice to determine the posteriors for estimating the crack centroid, length, and depth. This analysis does not use the damage classification model. The model used the experimentally observed SBSum values for the crack given VAM test parameters of the R_{freq} of 20, all R_{amp}, and all x_{pp} (x-direction). The prior for the centroid was assumed to be uniformly distributed over the entire length of the specimen [-.03048, 0.3048], the prior for crack size was assumed as uniform from 0.01 to 0.3048 m (half of the specimen length), and the prior for crack depth was uniform over the entire depth. The standard deviation for error was assumed to be 0.1. 1000 MH samples were used in MCMC simulations with a 20% burn-in.

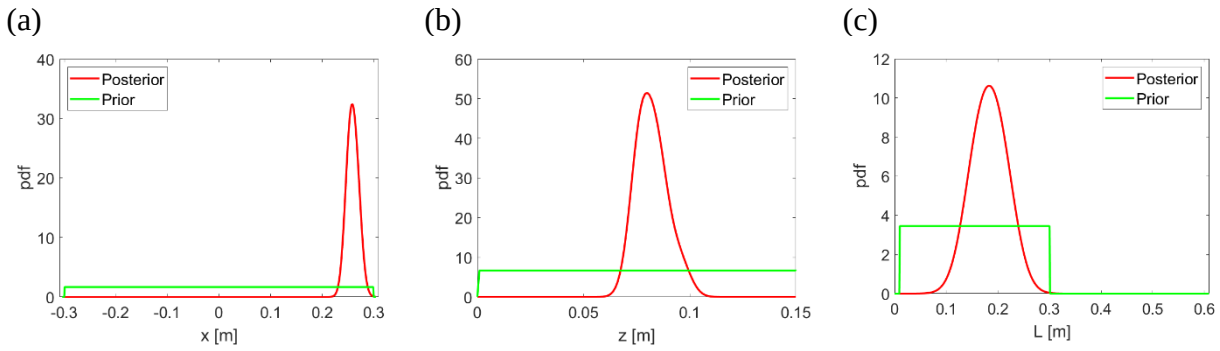


Figure 33: Bayesian estimation of damage for slice 6 using R_{freq} of 20, all R_{amp}, and all x_{pp} at 7 sensors; (a) Damage centroid pdfs (prior and posterior); (b) Damage depth pdfs (prior and posterior); (c) Damage length pdfs (prior and posterior).

With the results of Bayesian estimation (Figure 33), the same methodology used in section 0 is applied to create 1000 damage maps along each slice. These damage maps are averaged and damage probability for points between the slices is interpolated to obtain the overall damage probability map (Figure 34(a)).

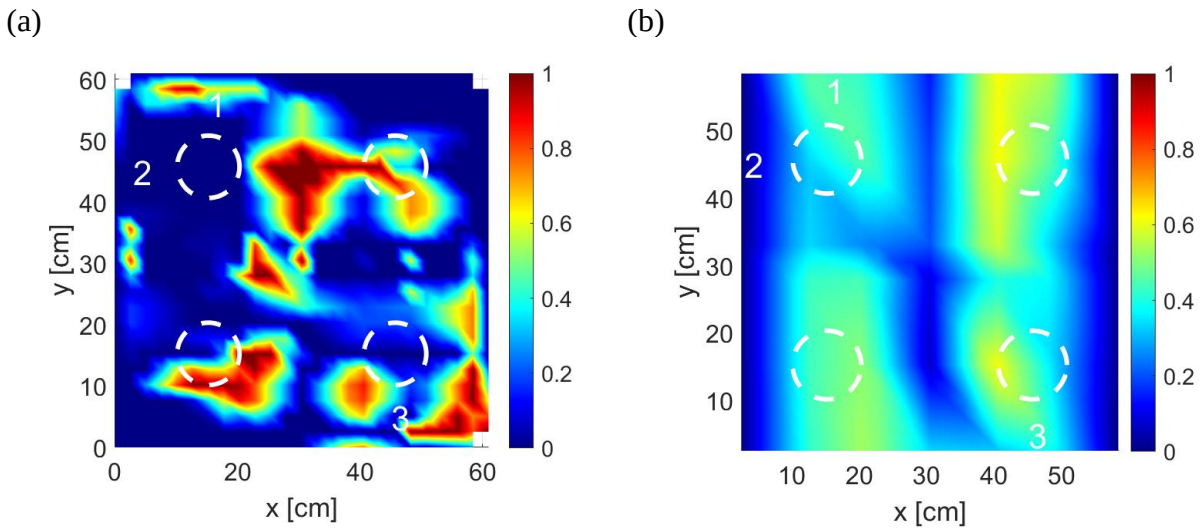


Figure 34: (a)Damage probability map using Bayesian estimation and the (SBSum) regression model;(b) Damage map based on all the slices of the slab specimen, using the classification model.

In Figure 34(a), the results for quadrant 3 and 4 (bottom half of the specimen) are consistent with damage probability using model-free (averaged likelihood) approach (Figure 32 (c)), petrography results for the specimen (Figure 10, Figure 11, Figure 12) and visual inspection after breaking the slab. The damage predicted in the bottom right corner of quadrant 1, at the location of the core in quadrant 1, and in between quadrants 1 and 2, indicate a damage probability higher than that indicated by the model-free data-driven approach (Figure 32 (c)). The low sensor density (7 sensors along the length of the slice), as well as the use of (vertical) 2D slices for damage diagnosis, may be possible reasons for the observed difference. The model did successfully predict

damage around the edges of the void in quadrant 2, as visually observed after breaking the slab.

Next, the classification model is analyzed for sensors located along each slice. The μ_{SBSum} used for each slice is the μ_{SBSum} for all 42 sensors for a given VAM test, not just for the sensors along that slice. This ensured that the diagnosis process for each slice has average SBSum information for the entire test surface. Damage location predictions for all the slices are combined to create an overall damage map (Figure 34(b)).

Overall, the classification model estimate shows a lower probability of damage as compared to the regression model results. The model successfully localized damage in quadrants 3 and 4. However, the area of damage in quadrant 2 is larger than we expect to detect with VAM, and the high-probability damaged area appears to be in quadrant 1. The core of quadrant 1 showed no ASR damage or cracking, and there was no noticeable surface cracking in this area. The errors in the 2D-model-based ML models could be due to the reduced dimensionality of the physics models, lower sensor density, and the approximate process of stitching together the results of multiple 2D cross-sections of a slab to obtain a damage map.

In order to facilitate comparison with discrete indicators of true damage (damage present (1) or absent (0)) at each sensor (Figure 32 (b)), the damage probabilities provided by the diagnostic method are converted into a binary result at each sensor (damage present or absent). For this purpose, we have chosen a threshold probability of 0.5. That is, after fusing information from multiple VAM tests, if the method yields a damage probability greater than or equal to 0.5 at a sensor, we label the (region near the) sensor as “damaged,” otherwise the (region near the) sensor is said to be undamaged. As a result, since the damage states for both the actual case and diagnosis method output are now binary variables, sensitivity and specificity metrics can now be calculated to provide a quantitative assessment of the methodology (Table 9). It can be seen in Table 9 that

both models exhibit good specificity and accuracy, but suffer from low sensitivity.

Table 9: Validation results for the experimental slab specimen.

	Classification Model	**Regression Model**
Accuracy	0.64	0.59
Sensitivity	0.25	0.38
Specificity	0.88	0.73

In this chapter, we discussed the validation of VAM-based damage diagnosis using ML models. The regression model is a forward model that estimates the damage index for given values of test parameters, damage location, and damage extent. Thus, the regression model needs to be used in conjunction with a parameter estimation methodology (such as Bayesian estimation), which can estimate the damage location and extent given the VAM test data. It is computationally expensive to use this model for damage diagnosis. The classification model is a diagnostic model, which provides the state of damage at a sensor location given the damage index value recorded at the sensor and other relevant VAM test parameters. It provides damage location in the plane of the (2D) testing surface; however, it does not provide damage location information in the direction perpendicular to the testing plane. The damage depth thus needs to be estimated using the regression model and a parameter estimation algorithm. If only the damage presence and damage extent are to be ascertained, the classification model can quickly provide the estimates.

The performance of the proposed diagnosis methods was evaluated using numerical experiments and previously reported data-driven model-free damage localization approach. Both methods performed well when validated against numerical simulation data. Note that the data used for validation was not used for training the model. Even relatively small cracks were successfully

diagnosed in the numerical experiments. For the plain concrete slab used in validation tests, the machine-learning model results for the bottom half of the slab matched better than the top half of the previously reported results (section 2.5)[48]. Of the two models studied, the regression model performs slightly better than the classification model with respect to the bottom half of the specimen. Both prediction and classification model-based damage diagnosis indicate high damage probability regions in the top half of the slab. These regions were not highlighted as damaged zones by the previously investigated model-free approach. Note that the damage in quadrants 1 and 2 is difficult to detect using VAM, since the aggregates in quadrant 1 shifted, leaving pure cement paste with no visible cracks in the core sample, and since silica in quadrant 2 left a large void that would be unable to open and close to produce frequency modulation. However, VAM should be able to detect the damage along the edge of the silica aggregate in quadrant 2 (Figure 31(a)), specifically location 1, where ASR gel seeped out from a small crack on the edge of the specimen. Thus, the damage locations that could be diagnosed using VAM (quadrant 3, quadrant 4, and the top left corner of quadrant 1) have been highlighted in both model-based techniques.

3.6. Conclusion

This chapter presented a physics-informed ML approach to facilitate nonlinear dynamics-based damage localization of internal, breathing cracks in concrete structures. The key difficulty of sparse training data availability for ML applications in SHM was overcome by using (approximate) physics simulation-based data for training the ML models. Two different neural network models for estimating damage, namely regression model and classification model, were trained solely with 2D FEA data. The 2D FEA simulation of one VAM test takes approximately 1 hour and data from about 250 VAM (simulation) tests was used for training. The computational

effort for training data generation could be significant, depending on the dimensionality of the simulation domain and the amount of data. Numerical experiments showed that both techniques successfully localize damage. The trained models were then used to localize damage in a cement slab (a 3D specimen) by considering multiple 2D slices and interpolating damage estimation between analyzed sections. Both methods for damage localization successfully estimated damage in the validation specimen (accuracy of 60 – 64%). The classification model eliminates the need for analyst intervention to define the damage index threshold for diagnosing damage. Both the regression, as well as the classification model, use computationally inexpensive neural networks to accelerate the damage diagnosis process. Additionally, the regression model enables hidden crack depth estimation, which has not been reported by previous VAM-based damage localization studies. This work has also automated the process of collecting and analyzing the SBSum data from multiple VAM tests for each pump/probe actuator location. A real-time damage probability map could be developed for a tested specimen's surface using the classification model.

CHAPTER 4

Multi-Fidelity Physics-Informed Machine Learning for Damage Localization using 2D and 3D Finite Element Models

4.1. Motivation

In the previous chapter, two different neural network models for estimating damage, namely prediction model and classification model, were trained solely with 2D FEA data since generating 3D FEA data is computationally expensive. The 2D FEA simulation of one VAM test takes approximately 1 hour on a desktop computer, and data from about 250 simulated VAM tests were used for training. 3D FEA simulation of one VAM test takes approximately 45 hours on a desktop computer. The computational effort for training data generation could be significant with 3D FEA simulation, depending on the simulation domain's dimensionality and the amount of data. This chapter will augment the low fidelity data from 2D FEA simulations with a few higher fidelity 3D FEA simulations for generating a multi-fidelity training data set. The accuracy will be compared to the computation cost to determine which FE models we should allocate time to running. We look to explore whether a higher fidelity physics model simulations for generating training data will help build a more accurate machine learning-based SHM framework. Additionally, we create a classification model that mimics the methodology in Chapter 2, where the damage index is the only value being used for prediction.

4.2. Relevant Background

The selected PIML approach (Section 2.2), requires expertise to build simulation models of the test procedure (at the required level of fidelity) and the computational resources to execute these models for different damage severities and locations. Model fidelity, i.e., a model's ability to simulate reality faithfully, is an important consideration in this process. Low-fidelity models typically require low computational resources and may be able to provide a large amount of simulation data for training the ML models, but with potentially large model errors. High-fidelity models are computationally expensive but may be better at capturing important phenomena for accurate damage diagnosis. Therefore, this chapter investigates the ability of PIML models, trained using physics models of different fidelity, to learn the relationships between true damage and the damage index for complex (nonlinear) physics-based diagnostic tests. The focus of the multi-fidelity modeling in this work is on building diagnostic multi-fidelity models. Most previous literature on multi-fidelity modeling [82]–[88] is regarding prognostic modeling. We also analyze the tradeoff between the computational cost of higher-fidelity simulation and the potential improvement in the accuracy of PIML-based diagnosis.

In the numerical implementation of this chapter's methodology, we consider the problem of localizing hidden cracks in concrete slabs for our investigation. In Chapter 3 [89], we investigated the performance of two PIML models for learning damage index patterns for vibro-acoustic modulation (VAM) tests of concrete structures in different test settings. The PIML models were trained using data from nonlinear dynamics simulation of the VAM test procedure for two-dimensional (2D) domains containing an internal crack. The physics model includes simulation of a geometric nonlinearity at the crack interface and hence requires the solution of a nonlinear system of equations. The 2D model does not capture the real-world three-dimensional wave propagation

and is termed the *low-fidelity* simulation. The low-fidelity physics-informed ML models showed moderate success in localizing damage in laboratory tests conducted on concrete specimens containing hidden cracks. It is reasonable to expect that nonlinear wave physics simulations for 3D domains, which capture geometric attenuation in the domain of interest and are a more faithful representation of the real-world VAM test set-up, would provide a more realistic and richer training data set. In the current paper, we investigate whether incorporating information from *high-fidelity* (nonlinear wave physics in 3D domains) simulations improves the diagnosis accuracy. However, the simulation of nonlinear wave physics for 3D domains using the finite element method is computationally demanding. Therefore, a multi-fidelity approach, where the training data is obtained from both low- and high-fidelity simulations while adhering to a computational resource budget, may be desirable. We investigate whether incorporating a small number of high-fidelity training runs provides a more accurate diagnostic model compared to the training data set generated from a large number of low-fidelity physics runs.

We investigate the performance of PIML models for the diagnosis of an internal crack in a concrete slab using VAM-based diagnostic tests when computational physics models of two different levels of fidelity are used to generate the data for training the PIML models. The diagnosis model is a classification model that uses damage index values from a VAM test and separates the sensor locations into two classes (damaged and undamaged). The threshold for separating damage index values is not specified by an expert but is learned from the physics simulation-generated training data provided to train the machine learning model. The training data are generated by finite element (physics) simulations of the test procedure for different damage locations, severities (crack lengths), and test parameters. The diagnostic performance of three types of models is investigated: one trained with data from only low-fidelity models, one trained with data from only

high-fidelity models, and different multi-fidelity models trained with different combinations of data from low- and high-fidelity models. The computational resource budget required for each diagnostic model is reported to enable informed decisions for practical application. The main methodological contributions of the work in this chapter include: a) development of multi-fidelity physics-informed machine learning models for structural damage diagnosis, using the transfer learning approach; and b) analysis of the trade-off between computational cost (to perform the computational physics simulations used to train the diagnostic model) and model performance in damage diagnosis.

4.3. Methodology

In this section, we discuss the methodology for supervised training of PIML models that facilitate damage localization. We also provide a brief overview of the computational physics models used to build the physics-informed ML models and the method of evaluating the performance of PIML models built using physics models of different fidelity.

4.3.1. Training data for diagnostic PIML

Damage identification methodologies for SHM vary based on the application; however, each methodology is composed of three important steps; (a) signal acquisition, (b) signal processing, and (c) signal interpretation [90]. Signal acquisition refers to the means by which the structure's response to active or passive excitations is monitored during a diagnostic test. Different excitations such as static/dynamic mechanical loading, heat exchange, optical illumination, etc., could be used; and different quantities such as strain, acceleration, temperature, thermal/optical images could be measured or acquired. The signal processing step refers to processing the signal for noise

reduction. This includes filtering and/or averaging techniques that attempt to remove measurement noise. After noise reduction, we enter the signal interpretation phase, where the large (temporal) dimension of the measured signal is typically reduced by extracting a few important damage-sensitive features or damage indices from the signal [48], [59], [91]. A damage index is a statistical indicator that is sensitive to the subtle changes in the measured signal from the original (undamaged) state to the damaged state [92]. An anomalous damage index may not always be interpreted as high damage probability because of measurement noise and the complexities of the diagnostic test process. These include variability in the set-up of the diagnostic test apparatus, the effect of environmental factors (not related to the damage) on signal acquisition, human errors, etc. A carefully trained supervised PIML model is thus needed for ascertaining damage presence/location/severity given the damage index.

The main difficulty in training the diagnostic ML models is obtaining a sufficient amount of labeled training data, which requires all three steps listed above (signal acquisition, processing, and interpretation) for known damage conditions. In the context of SHM, signal acquisition is traditionally conducted on real-life structural specimens. Given cost limitations, a large number of realistic experimental specimens to train the models used in signal interpretation are not always available. Some researchers have used computer-simulated diagnostic test data as training data [52], [93], [94]. Simulated data, however, poses its own challenges. High-fidelity simulations of the physics governing the diagnostic test (linear/nonlinear dynamics, heat conduction) have a high computational resource demand. Note that in this chapter, the computational cost or demand refers to the time needed for generating the training data, i.e., the time required for performing computational physics simulations using available computing resources. The time needed for performing a single (low- or high-fidelity) computational physics simulation is recorded, and as

more training data is acquired by performing more simulations, the computational cost increases linearly. Some previous studies have reduced the computational cost by using low-fidelity physics models. A low-fidelity model can be created by neglecting some of the physics involved, by reducing the dimensionality of the computational domain (use one- or two-dimensional domains instead of three-dimensional domains), by neglecting the heterogeneity and anisotropy of materials [18], [70], [89]. This chapter examines the effect of using physics simulations for computational domains of different dimensionality for training a PIML model for damage diagnosis.

The main issue investigated in this work with respect to fidelity is related to the dimensionality of the simulation domain used for generating training data for the PIML models. Nonlinear wave physics simulations in two-dimensional (2D) domains are termed low-fidelity simulations as they neglect the important geometric attenuation phenomenon that occurs in real-world three-dimensional (3D) structures. Nonlinear wave physics simulations in 3D domains, on the other hand, are referred to as high-fidelity simulations. The validity of using the low- and high-fidelity models for damage localization using relative damage index values has been established in our previous publications [48], [89]. In this chapter, we seek to answer this question: given computational physics models of two fidelities and the associated computational cost, what is the best strategy to generate training data for effective PIML-based diagnostic models? There are three options: a) use numerous low-fidelity model runs to generate a large amount of training data, b) use many low-fidelity model runs, and a few high-fidelity model runs to generate a moderate amount of, potentially richer, training data, and c) use a few high-fidelity model runs to generate a small amount of, potentially richest, training data. We build diagnostic PIML models (Figure 35) using the three alternatives described above and study their performance using both numerical and laboratory experiments.

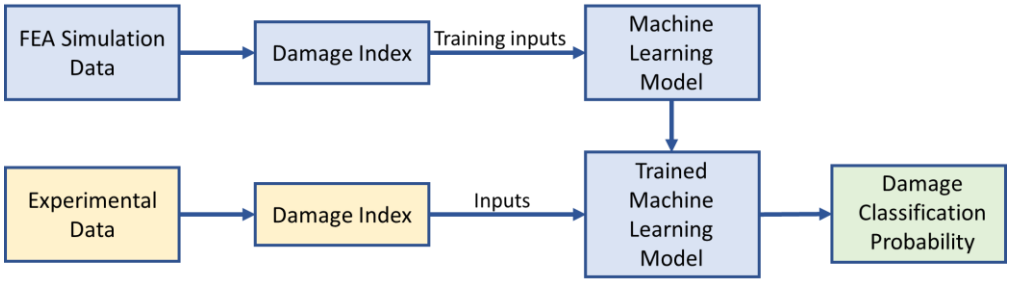


Figure 35: Proposed approach to build PIML model for damage diagnosis.

4.3.2. Training of PIML model using multi-fidelity data

There are multiple ways to use data from two different sources, such as data from 2D and 3D computational physics models, for training the ML models. Note that in most cases, when combining data from different sources for training ML models, the literature refers to different sensors or specimens and not to data from physics models of different fidelities [95]. In some cases, multi-fidelity models are built for the forward prediction of the system response to input excitation and are composed of a low- and a high-fidelity model, with an additive model correction term [96]–[98]. These models estimate a numerical response quantity, and such an additive approach cannot be used in a straightforward manner to build multi-fidelity diagnosis models that provide a damage classification result that is categorical.

Many different ML model architectures (artificial neural networks (ANNs), support vector machines (SVM), convolutional neural networks (CNNs), recurrent neural networks (RNNs)) are available to build classification models. Artificial neural network (ANN) models are the most suitable for transfer learning because they allow for the addition of new layers of neurons to be trained with newly acquired data. SVM and decision trees would only allow us to use a single data set, containing information from both high- and low-fidelity models. CNN and RNN are very

popular and powerful machine learning model architectures; however, CNN is best suited for classification of an entire image, not the sensor-by-sensor approach being explored in this study. The RNN architecture is typically employed for sequence-to-sequence (e.g., time series, text, etc.) modeling and is not suitable for the scalar inputs here. Hence, we choose feed-forward ANN model architecture in this work. The ANN is to be trained using the simulation-generated values of the damage index (input) at the sensors for a given diagnostic test, the mean value of damage index for all sensors (input), and the true damage state (i.e., presence/absence; output) as the training data. The model learns the relationship between the damage index observed at the sensor in the vicinity of damage in the context of the mean value of the damage index for the diagnostic test. For a future test, the model can then be used to classify each sensor location as either showing or not showing signs of damage.

In this work, a method similar to the transfer learning approach [56], [94], [99] (used in image processing) is employed to train a model using multi-fidelity training data. This approach is especially well-suited for SHM applications, where the initial diagnostic model needs to be updated using the newly acquired data. To build the multi-fidelity PIML model, the ANN model is first trained using only the low-fidelity data (Figure 36 (a)). A layer of neurons is then added to the trained model, and this additional layer is trained using only the high-fidelity data (while keeping the previously trained layer parameters unchanged) (Figure 36 (b)). The transferred layer(s) extract feature information from the first model and implement it in the training of the second model [100], [101].

(a) (b)

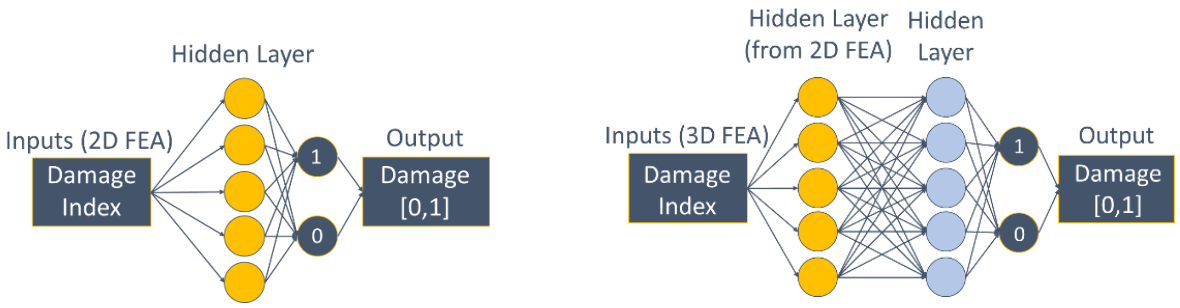


Figure 36: ML model architecture: (a) model trained using training data from low-fidelity physics model; (b) model trained using training data from low- as well as high-fidelity physics models.

In training the neural networks with low- and high-fidelity physics simulation data, the weights for each layer of the neural network need to be initialized; a common choice in the literature is to use random values between -1 and 1. A SoftMax layer is used as the activation function for the final layer of the model to normalize the output layer to a probability distribution [81]. In the next section, we discuss methodological details of signal acquisition, processing, and interpretation involved in VAM-based localization of hidden cracks in concrete.

4.3.3. Damage diagnosis with the PIML model

In Chapter 3, we developed machine learning models for hidden crack localization in concrete slabs using VAM. The PIML models used a damage index (SBSum value, see Eq. 1) and other relevant VAM test parameters as inputs and helped determine whether the damage is present or absent in the neighborhood of a given sensor. The focus of the previous work was on minimizing analyst intervention by learning damage index patterns from VAM test simulations of a single fidelity (all ML models were trained using 2D simulations of VAM tests). The current work investigates whether additional simulation data from a few 3D simulations to build the enhanced

PIML model shown in Figure 36 (b) improves the diagnostic performance. Note that this classification model only localizes the damage in the plane of the sensors (x-y plane); it does not localize the damage in the depth dimension (z direction) within the specimen. A different PIML model was developed earlier (section 3.3.1.3)[89] for depth quantification; the model uses the crack location and size identified by the classification model and the SBSum value at that location to determine the depth of the crack.

There are two approaches for obtaining damage probability at each sensor location. The first approach is to calculate the average of the classification results (values 1 or 0) from a significant number of VAM tests. The second approach is to directly use the probability provided by the SoftMax layer. Our previous work, (Figure 30) showed that the probability obtained from the SoftMax layer model provided similar results as those obtained by averaging the classification results from a large number of VAM tests. Hence, if only a small number of VAM tests are available to perform diagnosis, the probability given by the SoftMax layer can be used to as the damage probability. We next discuss the methodology used for the performance evaluation of the various diagnostic PIML models built.

4.3.4. Evaluation of diagnostic PIML model performance

The multi-fidelity diagnostic PIML methodology is evaluated using an experimental test specimen. The ground truth regarding damage presence/absence at each sensor in the validation specimen is needed to evaluate the performance of the proposed models. This may be obtained by performing destructive tests on the specimen. The damage presence/absence at each sensor is ascertained using the machine learning models and compared with the ground truth using four metrics: sensitivity (eq. 10), specificity (eq. 11), accuracy (eq. 12), and F1 score. The values of

these metrics range from 0 to 1, with one being the best. The mathematical definitions of F1 score:

$$F1\ Score = \frac{True\ positives}{True\ positives + \frac{1}{2}(False\ positives + False\ negatives)}, \qquad (13)$$

The above metrics are used to evaluate each model's performance for damage diagnosis using numerical tests and laboratory test specimens (EXP-A). Accuracy represents the fraction of the times the model correctly classifies the presence or absence of damage at a sensor; it is a combined metric that assimilates the information in the sensitivity and specificity metrics. However, a high accuracy value does not always mean the model has desirable performance because, for example, the accuracy metric could be easily biased by a large number of correctly identified (true) negative results in a structure containing damage localized in a small region. The F1 score is another combined metric that helps to alleviate the problem caused by the disproportionately high number of true negative values by taking the harmonic mean of sensitivity and precision. The F1 score is the harmonic mean of precision and sensitivity, where precision, also known as the positive prediction rate, is defined as the number of true positives divided by the total number of true and false positive values. The F1 score does not account for true negatives; however, it is a metric that ensures that high specificity does not lead to a disproportionately better model performance metric.

4.4. Diagnostic PIML model for concrete slab specimen

The aforementioned methodology is demonstrated using the dynamics-based SHM technique, VAM. This section examines the generation of the numerical training data for PIML and the details of training of the supervised machine learning models.

4.4.1. Training data generation

The details of generating training data using 2D and 3D finite element analysis (FEA) representing a concrete slab specimen are discussed below. The computational domains used for simulating the VAM tests are shown in Figure 37.

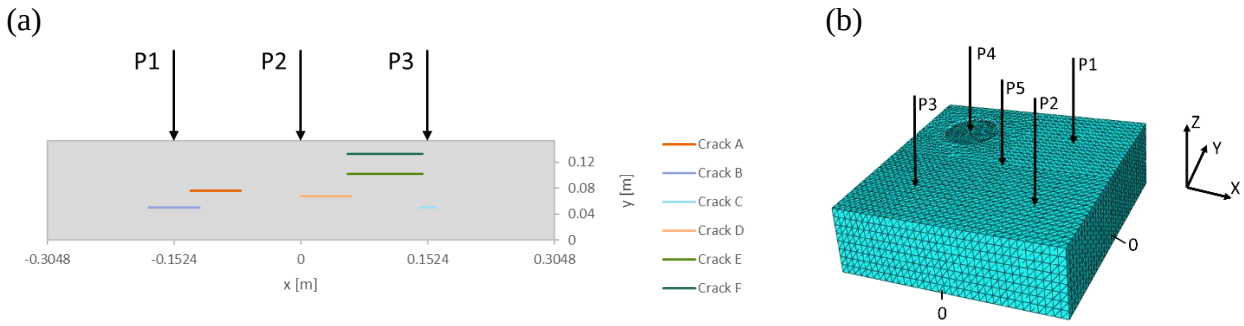


Figure 37: (a) The domain, the three pump/probe excitation locations, and crack locations used in 2D FE simulations; (b) The domain and the pump/probe excitation location used in the 3D FE simulations.

4.4.1.1. 2D FEA

2D FE models simulating VAM tests for multiple crack sizes and locations are used to generate the low-fidelity training data. The computational domain of the 2D specimen is 60.96 cm wide and 15.24 cm thick. The domain geometry and the pump and probe excitation locations used in the numerical simulations are shown in Figure 37 (a). The pump and probe actuators are placed at one of the three locations for each simulation x = {-0.1524, 0.0, 0.1524} m, y = 0.1524 m. Each 2D simulation domain contains a single crack, placed at different locations and of varying lengths. All of the locations for the simulated cracks and the parameters used in conducting the VAM tests on this numerical specimen can be found in Table 10. Various combinations of these parameter values are used to obtain 50 simulated VAM tests. The commercial finite-element program Abaqus

was used to perform the numerical simulations. Material properties used in these simulations are provided in Table 11. The domain is discretized using a structured mesh of 8-noded finite elements, and additional details of the finite element analysis can be found in section 3.3.1.2. Acceleration time-history data are recorded at sensors located on the top edge of the 2D domain. These data are used in calculating the SBSum values [35].

Table 10: VAM test parameters for the 2D FE simulations.

Crack Case	Centroid (x-direction) [m]	Total Training Points for Crack	Length [m]	Depth from Surface (z-direction) [m]	R_{amp}	R_{freq}	Pump/Probe Actuator Location [m]
A	-0.1025	13,896	0.06	0.0762	1, 0.5, 0.2, 0.1	20	-0.1524, 0.0, 0.1524
B	-0.1524	13,896	0.06	0.1016	1, 0.5, 0.2, 0.1	20	-0.1524, 0.0, 0.1524
C	0.1524	6,369	0.02	0.1016	1, 0.5, 0.2, 0.1	20	-0.1524, 0.0, 0.1524
D	0.03	1,560	0.06	0.085	1, 0.5, 0.2, 0.1	20	0.0
E	0.1016	780	0.09	0.0508	1, 0.1	20	0.1524
F	0.1016	2,340	0.09	0.02	1,0.1	20	0.0, 0.1524

Table 11. Material properties used in the 2D FEA model.

Material property	Value
Young's modulus (E)	27 GPa
Density (ρ)	0.15
Poisson's ratio (ν)	2400 kg/m^3
Mass proportional Rayleigh damping parameter (a)	2120.04
Stiffness proportional Rayleigh damping parameter (b)	1.787×10^{-7}

4.4.1.2. 3D FEA

A three-dimensional (3D) finite element model simulating VAM tests for a single crack location is used to generate the training data. The 3D computational domain is 60.96 cm wide, 60.96 cm long, and 15.24 cm thick. The domain geometry and one of the pump and probe excitation locations used in the numerical simulation are shown in Figure 37 (b). These are chosen to match the cement slab specimen used for validation experiments. Pump and probe actuators are

placed at the same locations as those used in validation experiments [48]. The parameters used for simulating VAM tests can be found in Table 12. Similar to the 2D simulations, Abaqus was used to perform numerical simulations. The material properties used in these simulations are given in Table 11. Acceleration time-history data are recorded at the sensors on the top surface of the 3D specimen in the X-Y plane and used in calculating the damage index values [35].

Table 12: VAM test parameters for 3D FE simulations.

Case	Diameter [m]	Centroid [m]	Total Training Points for Crack	Depth from Surface (z-direction) [m]	R_{freq}	R_{amp}	Pump/Probe Actuator Location [m]
1	0.127	[-.1524,.1524]	4,471	0.0762	20	0.1	5
2	0.127	[-.1524,.1524]	4.471	0.0762	20	1.0	5
3	0.127	[-.1524,.1524]	4,463	0.0762	20	0.1	1
4	0.127	[-.1524,.1524]	4,463	0.0762	16	0.1	1
5	0.127	[-.1524,.1524]	4,471	0.0762	20	1.0	2

4.4.2. ANN Model Training Evaluation

As mentioned in section 4.3.1, three feed-forward ANN classification models are trained and evaluated (LF, HF and MF models). All models determine whether a given sensor is showing signs of damage or not. The SoftMax [81] activation function was applied to the output with two output classes, damage present (1) and damage absent (0). As discussed in section 4.3.2, the models have two inputs: the damage index (SBSum) at each sensor for a given diagnostic test and the mean value of damage index for all sensors (μ_{SBSum}). The output of the model is the classification of the damage state (i.e., presence/absence). The relevance of these inputs has been studied in [89] based on physics simulations. The model should be able to learn the relationship between the damage index observed at the sensor and the mean value of the damage index in the test when the

damage is present or absent in the vicinity of the sensor location.

4.4.2.1. Classification model trained using low-fidelity (2D) simulation data

The model trained using the 2D FEA low-fidelity training data is named Model LF. Model LF (Figure 38) is a feed-forward neural network consisting of 2 layers. The first layer has 16 nodes and the second layer has 8. Model LF was found to have a training accuracy of 0.70, F1 score of 0.68, sensitivity of 0.67, and specificity of 0.74.

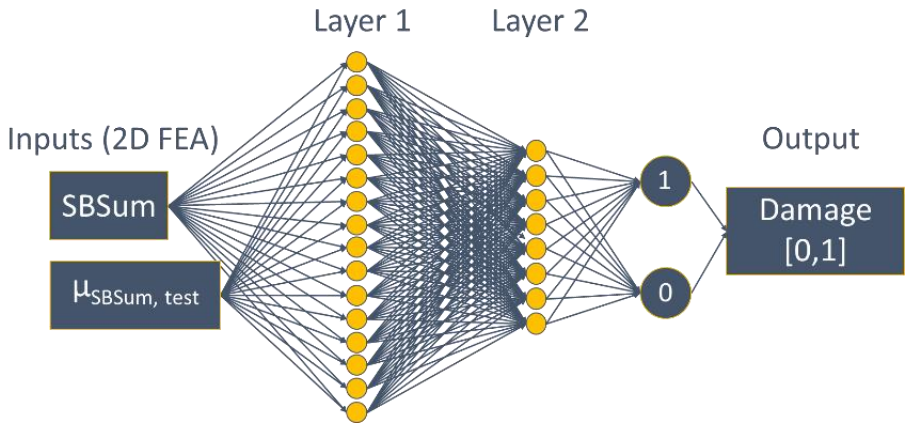


Figure 38: LF ANN model with 2 inputs (damage index) and two outputs (damage present (1) or absent (0)), with two hidden layers the first with 16 nodes, the second with 8. (The HF ANN model also has 2 layers, but with 20 and 8 nodes, respectively).

4.4.2.2. Classification model trained using high-fidelity (3D) simulation data

The model trained using the 3D FEA high-fidelity training data is named Model HF. Model HF (Similar to Figure 38) is a feed-forward neural network consisting of 2 layers. The first layer has 20 nodes and the second layer has 8. Model HF was found to have a training accuracy of 0.93, F1 score of 0.30, sensitivity of 0.18, and specificity of 0.997.

4.4.2.3. Classification model trained using multi-fidelity simulation data

The multi-fidelity data-based classification model (Model MF) is built by augmenting Model LF, trained on low-fidelity (2D FEA) data, with an additional layer of neurons. The additional (third) layer of model MF is trained using 3D FEA data; however, the weights and biases for the first two layers are kept the same as the trained weights and biases from previously trained Model LF. The MF model uses the same inputs as Model LF: the SBSum and mean SBSum for each test ($\mu_{SBSum,\ test}$). The model is trained using one 3D FEA VAM test. Multiple MF models were trained using 3D FE simulations with different settings representing different VAM tests. Based on the VAM diagnosis studies reported in [35], the VAM test shown as Case 1 in Table 12 should give a higher damage index and high sensitivity. This study showed that a probing excitation frequency 20-25 times larger than the pumping excitation frequency and probing excitation amplitude ratio of $1/10^{th}$ of the pumping excitation gave the best damage localization results. Additionally, the pumping and probing excitation actuators are located in the center of the specimen, allowing for the wave to uniformly illuminate the entire specimen.

The MF model (Figure 39) thus consists of 3 layers; the first layer has 16 nodes, the second layer has 8 nodes, and the third has 6 nodes. The weights and biases for the first two layers are set to the converged values of the two layers of Model LF. Model MF was found to have a training accuracy of 0.93, F1 score of 0.279, sensitivity of 0.65, and specificity of 0.93.

(a) (b)

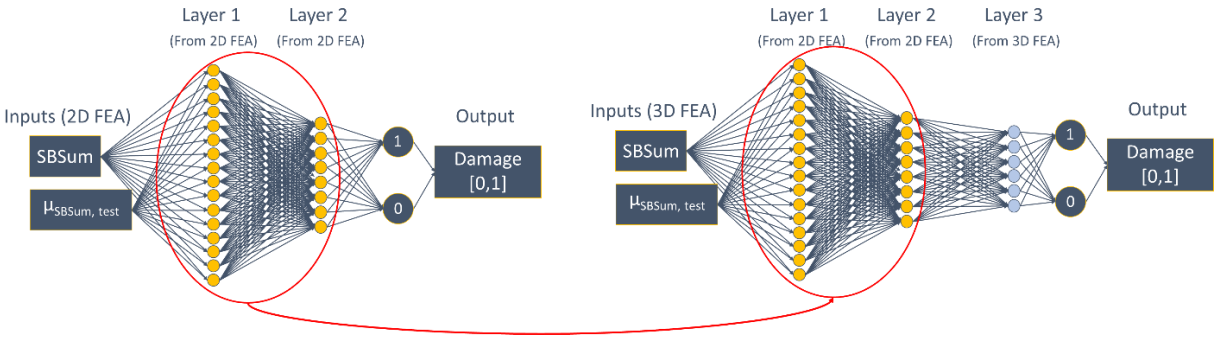


Figure 39: (a) Model LF with 2 inputs (damage index), and two outputs (damage present (1) or absent (0)) trained using 2D FEA; (b) Model MF with 2 inputs (damage index), and two outputs (damage present (1) or absent (0)). For the first two layers, weights and biases are transferred from the previously trained LF model. Only the third layer's weights and biases are trained using the 3D FE simulation data.

4.5. Performance evaluation of diagnostic PIML models

In this section, the three models trained and discussed in section 4.4.2 are evaluated using laboratory test specimens. First, laboratory experiments are discussed, and then results for the diagnostic PIML models are presented.

4.5.1. Numerical and Laboratory experiments

The PIML model performance is first assessed using a 3D FEA numerical simulation of a VAM test with a different crack location than that used to train the models. The simulation used for training contained a circular crack with a centroid at (45.72, 45.72) cm, a diameter of 0.127 cm, and a depth located at 7.62 cm from the top surface. The probing excitation frequency is 20 kHz, the pumping excitation frequency is 1 kHz, the probing excitation amplitude ratio is 1/10th the pumping excitation amplitude, and the pump and probe actuators are located at (30.48, 30.48) cm, the center of the specimen. The numerical specimen (FEA) has a few hundred computational

nodes on the top surface, and acceleration can be recorded at all these locations. However, it is not feasible to have these many sensors on a structure in practice. A grid of 100 (10 by 10) sensors is thus created on the top surface, and it is assumed that measurement data is available only at these 100 sensor locations. Figure 40 (a) is the true damage location of the crack in the 3D FEA simulation. Figure 40 (b) is the raw SBSum map for the model with a 10 x10 sensor grid.

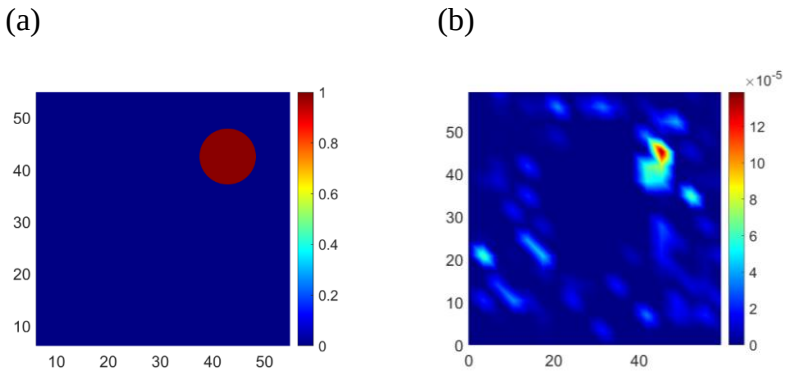


Figure 40. (a) True damage for the numerical experiment; (b) SBSum values with a 10x10 sensor grid.

The PIML models are evaluated by performing physical VAM tests on a cement slab specimen containing pockets of reactive aggregates at known locations [89]. A cement paste slab with reactive aggregate pockets at known locations and dimensions of 60.69 × 60.69 × 15.24 cm^3 was cast and cured at Vanderbilt University. The details of this specimen's casting and curing process are discussed in detail in section 2.4.1. This slab was cast and cured in an aggressive environment to produce alkali-silica reaction (ASR) induced cracking in the slab. Various VAM (non-destructive) as well as destructive (petrographic and visual inspection of broken sample) tests were performed on this specimen to develop and validate the VAM-based ASR damage localization methods. The approximate aggregate locations, core locations for petrographic testing and other

destructive testing results are shown in Figure 41 (a). Figure 41 (a) also shows the sensor locations for VAM tests conducted on the specimen and the presence/absence of damage at each sensor, based on the known aggregate locations, chemical and physical tests, and petrographic studies. In laboratory experiments, the pump and probe excitations were delivered using piezo-stack actuators. The locations of these actuators and the frequencies at which they operate (i.e., the pump and probe frequencies) were varied. This resulted in 120 VAM tests discussed in section 2.4.2. Figure 41 (b) is the average SBSum across all 120 VAM tests.

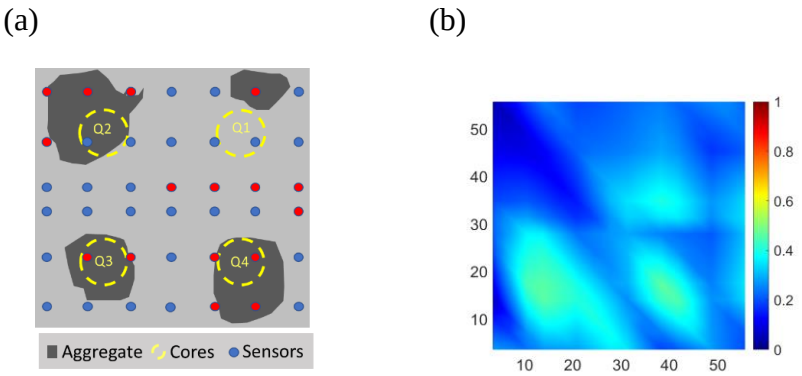


Figure 41. (a) Experimental specimen aggregate, core, and sensor locations. Red sensors are assumed damage presence at these sensor locations, blue sensors are assumed no damage present; (b) Average of scaled SBSum for validation specimen across all test parameters.

4.5.2. PIML model performance results

Damage localization on the surface of the numerical and laboratory test specimens was conducted using the three (LF, HF and MF) PIML models, and the results are discussed below.

4.5.2.1. Numerical experiment results

The damage probability results for the numerical experiment using Model LF, Model HF, and Model MF are displayed in Figure 42. The damage maps in Figure 42 are based on the SoftMax

layer output for each PIML model. Only one test is being evaluated, and the SoftMax layer output is used to obtain the damage probability (discussed in section 4.3.2). The metrics quantifying the performance of each model can be found in Table 13.

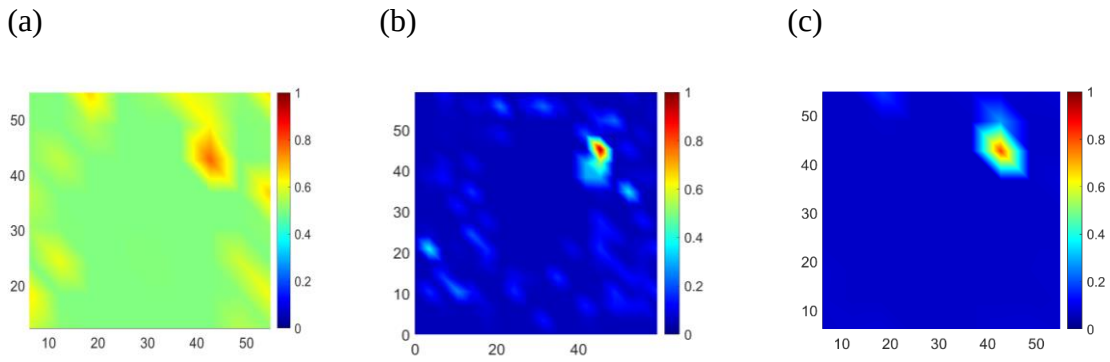


Figure 42. Damage probability results: (a) Model LF (b) Model HF; (c) Model MF.

Table 13: PIML model performance for the numerical experiment.

Model	Sensitivity	Specificity	Accuracy	F1 Score
LF	0.343	0.859	0.808	0.261
HF	0.166	0.977	0.928	0.278
MF	0.177	0.996	0.927	0.290

The damage probability results for all three models show a high probability of damage at the true damage location. For the numerical experiment, all three models show high accuracy (greater than 0.80), as shown in Table 13. The LF model shows the highest sensitivity; thus, it is more likely to find damaged areas; however, that comes at the cost of reduced specificity, which means it is also more likely to give false positives (i.e., identify undamaged areas as damaged). The MF model shows the highest accuracy as well as the highest F1 score; however, it is more likely to miss some of the damaged areas due to its low sensitivity score.

4.5.2.2. Laboratory experiment results

The three PIML models are evaluated using the true damage locations from Figure 41 (a). To this end, experimental VAM test data from 120 VAM tests conducted on the laboratory specimen is used as diagnostic data. The damage index (SBSum) is extracted from the measured signal at each sensor for each VAM test. These SBSum values are provided as inputs to the PIML models. The damage probability is calculated as the average of 120 classification results, and is displayed in Figure 43 for all three models. The performance metrics for each model for the laboratory test specimen are displayed in Table 14. This table also contains performance results for two additional models, trained without using the transfer learning approach, to better demonstrate the value of transfer learning. Specifically, for these models, (named '50 LF + 1 HF' and '196 LF + 5 HF') the data from the indicated number of LF and HF simulations is pooled together, and the model is trained in a single shot.

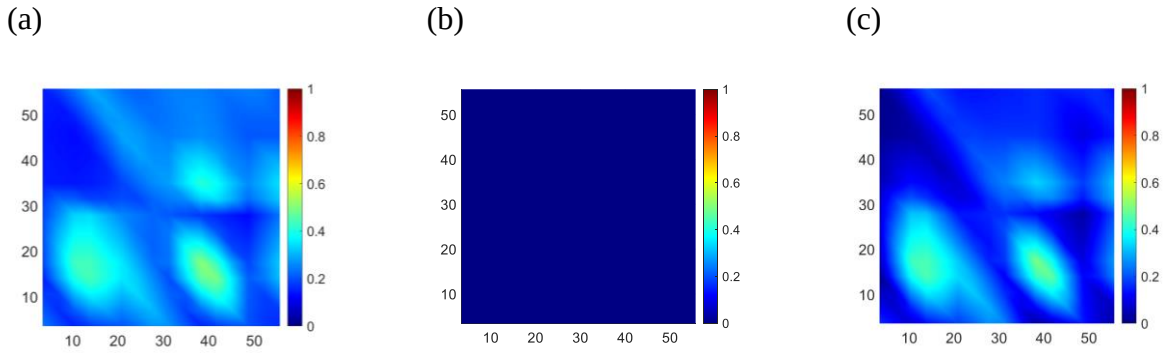


Figure 43. Damage probability results: (a) Model LF; (b) Model HF; (c) Model MF.

Table 14: PIML model performance for the laboratory test experiment over 42 sensors for 120 tests (i.e., 5040 validation points).

Model	Sensitivity	Specificity	Accuracy	F1 Score
LF	0.293	0.788	0.599	0.357
HF	0.000	1.000	0.619	0.000
MF (transfer learning)	0.228	0.864	0.622	0.315
50 LF+1 HF	0.009	0.991	0.617	0.017
196 LF + 5 HF	0	0.999	0.619	0

Model HF (Figure 43 (b)) is unable to correctly localize damage. Both Model LF (Figure 43 (a)) and Model MF (Figure 43 (c)) successfully capture the damage locations in the laboratory test specimen. This is a promising result for performing actionable, probabilistic diagnoses. Both the LF and MF models have high F1 scores and provide good damage diagnoses. In this case, although the HF model has a higher accuracy score, it is only because of the high specificity. It is clear from its F1 score (zero) that Model HF performs poorly overall. This is likely because the model is trained using data from only a single simulated VAM test (i.e., one set of damage and test parameters). A single VAM test 3D simulation test is not able to provide sufficiently diverse information regarding the damage index values in different test/damage conditions. It may be necessary to perform multiple, computationally expensive 3D VAM tests to improve the diversity of the training data. It is also apparent that the transfer learning approach provides the best performance (the highest F1 score) of the three MF models considered here. In the next section, we investigate whether adding more training data improves the performance of PIML models.

4.5.3. Analysis of computational effort for training data generation and diagnostic PIML model performance

We further evaluated the PIML model performance by training multiple PIML models for

each fidelity (LF, HF, MF) using training data from increasing number of (2D and 3D) simulations in order to analyze the tradeoff between computational effort and diagnostic model performance. For model LF 50 to 196 simulated VAM tests with 1,888 to 15,114 training data points were considered and for Model HF 1 to 5 simulated VAM tests with 4,471 to 22,339 training data points were considered. For Model MF, LF model training data (50 to 196 simulated VAM tests with 1,888 to 15,114 training data points) is used to train the first two layers, and HF model training data (1 to 5 tests with 4,471 to 22,339 training points) is used to train the third layer. Recall that the 2D FE simulation of one VAM test of a concrete slab specimen (see Section 4.4.1.1 for details) takes approximately 1 hour on a desktop computer, while the 3D FE simulation of one VAM test (see Section 4.4.1.2 for details) takes about 45 hours on a desktop computer [89]. The computational effort related to training data generation for models trained using different mount of LF and/or HF training data points can, therefore, be measured in terms of hours of computer run time. The accuracy and F1 scores for the different models (using different mounts of training data from models of different fidelities) are plotted against the computational effort (in hours) needed for generating training data for these models for the laboratory experiment (Figure 45).

Degradation in performance after adding cases 3 and 4 can also be seen in the F1 scores of the MF models for the laboratory experiment (Figure 45 (b)). The MF model's F1 score decreases dramatically with the addition of Cases 3 and 4 (Table 12). This is likely because in Case 3, SBSum values are high in regions without damage, and in Case 4, SBSum values are really low in the damaged region. This is not unexpected for VAM tests. In our previous research [48], we have encountered a few of these misleading VAM test results, and discussed the importance of using information obtained from many VAM tests to improve the diagnosis accuracy. We have shown [19] that the damage localization accuracy is dependent upon the testing parameters used in VAM.

For example, the distance of the pump and probe actuator from the location of the crack can affect diagnostic performance. The previous research demonstrated how some test settings help diagnosis more than others. This is also true for generating simulation data for training ML models: some test/damage set ups are better than the others for a given validation test case. Ideally, the additional training data could be chosen by solving an optimization problem aimed at maximizing the information gain. However, solution of this optimization problem is computationally unaffordable. Here, we studied the model performance under realistic (sub-optimal), computational resource-constrained conditions. In this setting, it is expected to have model performance degradation by addition of data that is not relevant for the testing case.

For the numerical experiment, the HF models showed consistently higher accuracy (for low as well as high computational expense for generating the training data) (Figure 44 (a)). This is likely because the numerical experiment is fairly similar (in terms of VAM test parameters) to the VAM test simulation used for training data generation. Additionally, the damage size and specimen geometry are identical for training and numerical experiment (but the damage location is not the same). Accuracy is high for these cases because of the high number of true negatives and a low number of false positives. The F1 score, which does not consider the effect of true negative results, is, however, low. For the numerical experiment (Figure 44 (b)), it is evident that model LF has the best F1 score, indicating it is the most sensitive damage diagnosis model. The MF model's F1 score decreases dramatically with the addition of Cases 3 and 4 (Table 12). This is likely because in Case 3, SBSum values are high in regions without damage, and in Case 4, SBSum values are really low in the damaged region. This is not unexpected for VAM tests. In our previous research [102], we have encountered a few of these misleading VAM test results, and discussed the importance of using information obtained from many VAM tests to improve the diagnosis

accuracy. This degradation in performance after adding cases 3 and 4 can also be seen in the F1 scores of the MF models for the laboratory experiment (Figure 45 (b)). From the F1 scores for the laboratory tests, it is clear that the LF models provide the best damage diagnosis results. The addition of one HF simulation exhibiting a damage map with high SBSum values in damaged areas and low SBSum values in undamaged areas (case 1 and 2) can also increase the F1 scores to improve diagnostic performance; however, using that computational effort to provide 45 additional LF tests leads to significant additional improvement in the diagnostic performed. This is because the additional LF simulations are able to provide a better coverage of the VAM test parameter space as well as damage configuration space, leading to more generalizable PIML models.

(a) (b)

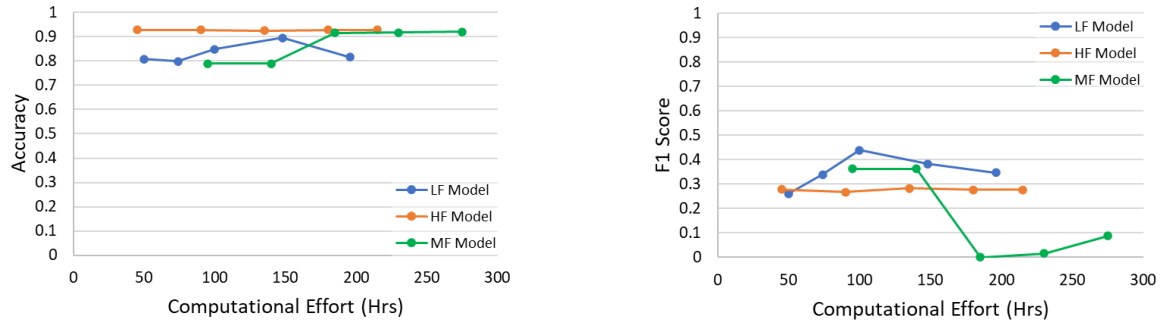


Figure 44. Numerical experiment to evaluate diagnostic PIML models: (a) Accuracy vs computational effort to generate training data; (b) F1 score vs computational effort to generate training data.

(a) (b)

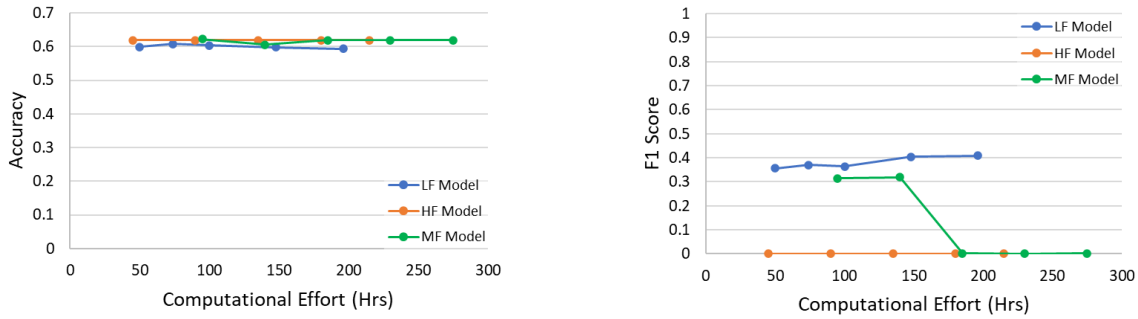


Figure 45. Laboratory experiment to evaluate diagnostic PIML models: (a) Accuracy vs computational effort to generate training data; (b) F1 score vs computational effort to generate training data.

4.6. Conclusion

This chapter investigated the performance of damage index-based (direct) diagnostic PIML models trained using data obtained from physics models of different fidelity. The term fidelity here refers to the computational domain's dimensionality used to simulate the physical phenomenon (nonlinear wave propagation) underlying the diagnostic test procedure (diagnostic signal acquisition phase). We built and evaluated three PIML models (LF, HF, and MF) for VAM-based diagnosis of hidden cracks in concrete slabs. In general, the diagnostic PIML models showed good performance in terms of identifying the damage locations. We found that information gained from additional low-fidelity simulations increases accuracy and F1 score more than adding a high-fidelity simulation to the training data. The addition of a single HF simulation-based training data to the training data from 50 LF simulations improved the diagnostic performance in numerical experiments. However, using the equivalent number of computational resources to obtain additional training data from 45 LF simulations improved the diagnostic performance even more. The transfer learning-based MF PIML approach is found to perform better than the simple HF and

LF training data pooling-based approach. The transfer learning-based PIML approach can be improved in the future by using an adaptive learning approach to select the test/damage parameters for additional HF simulations. Additionally, this research showed that using the F1 score for damage diagnosis model performance provides better information about the model's performance than standard accuracy calculations when the damage is localized in a small region (which is typically the case). Overall, the diagnostic performance evaluation results showed that, for this particular application, about 50 LF physics simulations with varying damage and test parameters could capture the information (or patterns) needed for damage index-based diagnosis using PIML.

Chapter 5

Damage Localization in Plain and Reinforced Concrete Specimens with Distributed Damage Inducement

5.1. Motivation

In this chapter, we evaluate the performance of the diagnostic PIML models developed in previous chapters for concrete specimens containing steel reinforcement and ASR-inducing reactive aggregates dispersed evenly throughout (which implies unknown damage locations). Several new issues need to be addressed for VAM-based hidden crack diagnosis in reinforced concrete specimens. Firstly, the suitable testing parameters (actuator locations, actuation frequencies) to accurately detect and localize hidden cracks in reinforced concrete specimen have not been identified. Secondly, the utility of the nonlinear diagnostic technique in the presence of an additional source of geometric nonlinearity (at the interface of steel reinforcement and concrete) needs to be investigated. Furthermore, we also evaluate the performance of the PIML-based diagnosis process for concrete specimens whose geometry is different from the computational model geometry used for generating the training data. Similar to the procedure followed in preceding chapters, we cast concrete specimens (control specimen as well as specimens with alkali load boosting in the cement and reactive aggregates), conduct VAM tests in the laboratory, diagnose damage using the PIML-based method, and take cores from multiple locations. Petrographic examination is performed for each core, and a quantitative ASR damage metric known as the damage rating index (DRI) is used to ascertain the ground truth regarding ASR

damage at the core location and validate the PIML-based diagnostic result.

5.2. Relevant Background

In this research, PIML models were trained using damage and SBSum values obtained from computational (finite element) simulations of the nonlinear dynamics-based, VAM test procedure. The computational effort needed for generating the simulated data can vary based on the fidelity of the simulation (FE) model. With the computational effort being a constraint for this research, given the information learned from Chapter 4, this chapter will focus on training ML models using only 2D FE simulations. Although the FE models used for generating the training data are 2D FE (low-fidelity) models, in this chapter, we investigate the performance of PIML models built using the multi-fidelity, transfer learning approach discussed in Section 4.3.2, but by using experimental data. In addition to simulation data, we include the data from VAM tests and true damage state for the plain concrete slab (EXP-A Section 2.4.1). The transfer learning approach for training models from both simulated and experimental data has been previously investigated for image processing applications [93]. The models in this chapter also combine simulation-based training data from simulated test specimen of different geometry. This is important since it is not feasible to simulate a computational domain for every specimen geometry to be tested in real-world application. The PIML extracts information about the relative damage index and models the complex relationship between the damage index and the presence/absence of damage at a sensor location for a few geometries incorporated in the training data generation. Here, we investigate the accuracy of PIML models for test specimens having geometries different than those used for performing data generating simulations.

Chapter 4 combined VAM with PIML models to provide a probabilistic damage localization

result for a thick concrete specimen containing pockets of reactive aggregate at a few known locations [48]. In this chapter, we examine the application of this PIML to concrete specimens with the ASR-inducing (reactive) aggregates dispersed throughout the specimen instead of in known pockets). The DRI is used to help quantify the severity of ASR-induced damage [103] at select core locations and these values are compared to the PIML performance. As mentioned in Chapter 2, most ASR techniques have been used for damage detection, not localization [104]. Frequency-banded synthetic aperture focusing techniques (FB-SAFT) have been used in ASR localization of thick reinforced specimens with dispersed ASR-induced damage, however; these approaches fail to provide a probabilistic value and difficulty in the presence of rebar (i.e., they show the location of rebar as a possible damage location) [105], [106].

In summary, the focus of this chapter is to study the performance of PIML models using the VAM damage index (SBSum) to diagnose damage in concrete specimens that resemble the conditions in real-world, reinforced concrete structural members. This chapter considers specimens with uniformly dispersed reactive aggregates to induce ASR, and considers reinforced specimens in addition to plain specimens.

5.3. Methodology

This section discusses transfer learning approaches used in building the PIML classification models for damage diagnosis. Specifically, we examine combining the physics simulation data with experimental test data to train a model.

5.3.1. Classification model

We continue to pursue the ANN architecture similar to previous chapters. As discussed in

Chapters 3 and 4, the inputs to the classification model are the damage index value at a particular sensor and the average damage index value for the VAM test. The output of the classification model is a binary variable indicating the presence or absence of damage in the neighborhood of the sensor: an output of 1 means that there is damage in the vicinity of the sensor, and an output of zero means that there is no damage in the sensor neighborhood. The data used for training such classification model consists of model inputs and the corresponding output values obtained from physics simulations or experimental tests. In this chapter, the training data is obtained from physics simulations for three types of 2D domains (Section 5.4.1: Figure 48, Figure 49, and Figure 50). The data can be combined using two main approaches. The first is by combining the data directly into a single training pool, regardless of the simulation domain from which the information comes (Figure 46). The second approach uses the multi-layered approach discussed in Section 4.3.2. In this approach, the three simulated training data sets are considered separately, with the lowest fidelity being i and the highest, iii. The lowest fidelity data set (i) is used to train a model with M layers of L nodes each (Figure 47 (a)). The parameter values from these M layers are then imported as the parameters for the first M layers in a second model, which will be trained using the medium-fidelity dataset (ii). The parameters (weights and biases) for the first M layers are kept unchanged as additional N layers are added and trained (Figure 47 (b)). The parameters for these N layers are trained using the medium fidelity dataset (ii). Lastly, a third model is trained using the highest fidelity data (iii). The parameters for the first M and N layers are fixed at the values obtained by training the first three models and the parameters are kept unchanged. The last O layers are added to the model and trained using the highest fidelity data set (iii) (Figure 47 (c)). The output of this model is the damage prediction at each sensor using inputs of multiple fidelities. Additional models are built by combining the data directly into a single training pool, the transfer learning FEA

approach, and a combination of the two, where one layer may contain multiple simulation domains. Note that for all models, the inputs for the classification model are normalized with a mean of zero and standard deviation of 1. The weights are then initialized as random values between -1 and 1. The SoftMax layer, as the final layer, normalizes the output to a probability distribution [81]. Note that the classification model only diagnoses damage along the surface of the specimen, where sensors are installed. The diagnosis provides damage location information in terms of the coordinates on this surface (e.g., the X-Y plane).

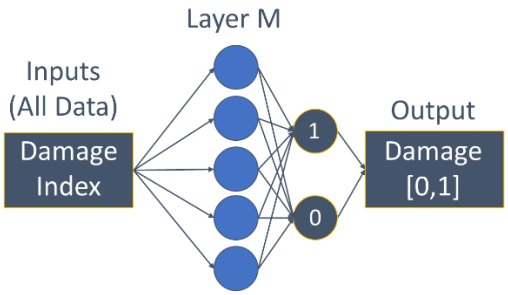


Figure 46: Model architecture for classification model using inputs from all simulated training data. M layers (M=1) and L=5 nodes.

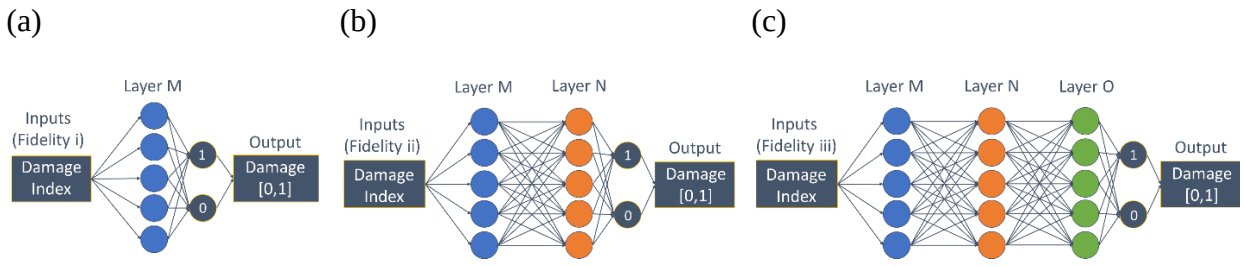


Figure 47: Multi-fidelity classification model architecture: (a) model trained with M layers of L=5 nodes using fidelity i simulation input data; (b) ANN classification model with M (M=1) layers set with fixed weights and biases from the previous model and N (N=1) layers trained with fidelity ii data; (C) ANN classification model with M (M=1) and N (N=1) layers set with fixed weights and biases from the previous model and O (O=1) layers trained with fidelity iii data.

5.3.2. Multi-fidelity model: combining experimental and simulated data

As discussed previously and in Chapter 4, the transfer learning framework combines data from different training sources (fidelities) to diagnose the damage. This research aims to build a ML model for dispersed ASR-induced damage localization using training input data from both the 2D FEA simulated VAM data and VAM data from an experimental specimen with localized ASR-induced damage (EXP-A Section 2.4.1). Damage locations in the experimental specimen were confirmed with destructive testing. This approach combines the 2D simulated data to train M layers with L nodes (Figure 47(a)). The information from these M layers is then set as the first M layers in a next model, and it is trained using the experimental training dataset (EXP-A). The M layers parameters (weights and biases) are kept unchanged as additional N layers are added and trained (Figure 47(b)).

5.3.3. Depth prediction using a regression model

This methodology uses a regression model from all the VAM parameters to predict the damage index SBSum (Figure 24) (Section 3.4.2.1). The input parameters for this model include testing parameters such as the R_{freq} and x_{sensor}, and damage parameters including damage location, crack size, and the depth of damage. Using the classification models detailed in Section 5.3.1, the approximate damage location and crack size can be calculated, leaving only one unknown parameter, damage depth (θ).

This regression model is a surrogate for the forward problem; we explore the inverse problem to perform depth diagnosis. Similar to Chapter 3, we use the Bayesian approach [89] to update the posterior by combining prior knowledge with the observed data. Theta, θ, represents the unknown parameter, damage depth. The likelihood is estimated as $P(SBSum|\theta)$. Our prior

knowledge distribution, $P(\theta)$, is the entire depth of the specimen being analyzed. (For the validation specimens discussed in Section 5.6 the prior is a uniform distribution U[0, 0.3048] m.) Using the regression surrogate model and Markov Chain Monte Carlo (MCMC) simulations the posterior distributions for the depth ($P(\theta|SBSum)$) are calculated. These simulations use a "random walk" technique, the Metropolis Hastings (MH) algorithm, as the sampling method to evaluate the posterior. This sampling method is computationally expensive since the surrogate model is run from 1000 to 5000 times for proper convergence, causing the depth model to have a significantly higher computation cost than the classification model.

5.4. Training Data for Physics-Informed Machine Learning for Localization

This section discusses the two types of training data used for the machine learning models in this chapter. The first data set is obtained using 2D FE simulations of VAM test. The damage and VAM test parameters are varied to obtain different VAM tests for different damage and test configurations. The second data set is obtained by conducting VAM experiments on EXP-A the plain cement slab with induced ASR damage (at a few locations) (Chapter 2, Sections 2.4 and 2.5.3).

5.4.1. Numerical Simulations of VAM test

As discussed in previous research, the computational cost for simulating 3-dimensional (3D) numerical training data is nearly 50 times higher than simulating mechanical vibrations for a 2-dimensional specimen (2D). Additionally, the information gained from the 50 2D models was more than one 3D model when added to the training data of the machine learning model. For this reason, we focus on training the machine learning model with 2D FEA data.

Three types of 2D FEA models are built to generate training data for the ML models. The first is a homogeneous elastic computation domain (simulating the behavior of a plain concrete specimen) 60.96 cm wide and 15.24 cm thick (2D-A) (Figure 48). The material properties for the homogeneous elastic material 9plain concrete) are given in Table 15. This is the same domain geometry and composition as the 2D FE simulations for Chapter 3 and 4 (Sections 3.4.1 and 4.4.1.1) [107]. The second domain (2D-B) is also a plain concrete specimen with a width of 60.96 cm and 30.48 cm thick (Figure 49). The third and last domain (2D-BR) is the same as the second in dimensions, a width of 60.96 cm and a depth of 30.48 cm, but contains six circular inclusions (diameter 2.2 cm) at x= {17.46, 30.48, 43.5} cm for both y= {7.32, 23.16} cm, to simulate the effect of rebar reinforcement in the specimen (Figure 50). The material properties of steel rebar are given in Table 15. Delamination was not modeled between the steel inclusion and plain concrete contact surface. The pumping and probing excitation actuators were placed at one of the three locations for each computational domain (2D-A, 2D-B, 2D-BR); x= {-0.1524, 0.0, 0.1524} m, y=0.1524 m. Only a single hidden crack was model for each simulation at different locations and lengths. Table 16 shows the locations for each of the simulated cracks and the parameters used to conduct the VAM tests in these numerical specimens. Fifty different VAM test simulations were made for each computational domain (2D-A, 2D-B, 2D-BR) for varying cracks and test parameters. The commercial software Abaqus was used to conduct the numerical simulations. The domain for all models was discretized using a structured mesh of 8-noded finite elements. Additional details on the simulation of domain 2D-A can be found in [108]. Simulations for 2D-B and 2D-BR were conducted in the same manner, with the only differences being the size of the domains and the addition of rebar.

Table 15:Material properties used in the 2D FEA model.

Material property	Value for Concrete	Value for Steel Rebar
Young's modulus (E)	27 GPa	200 GPa
Density (ρ)	2400 kg/m^3	7850 kg/m^3
Poisson's ratio (ν)	0.15	0.44
Mass proportional Rayleigh damping parameter (a)	2120.04	40.06
Stiffness proportional Rayleigh damping parameter (b)	1.787× 10^{-7}	-1.434× 10^{-8}

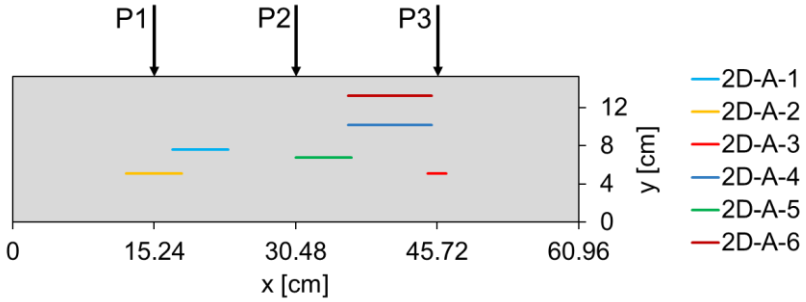


Figure 48: The domain, the three pump/probe excitation locations, and crack locations used in FE simulation 2D-A.

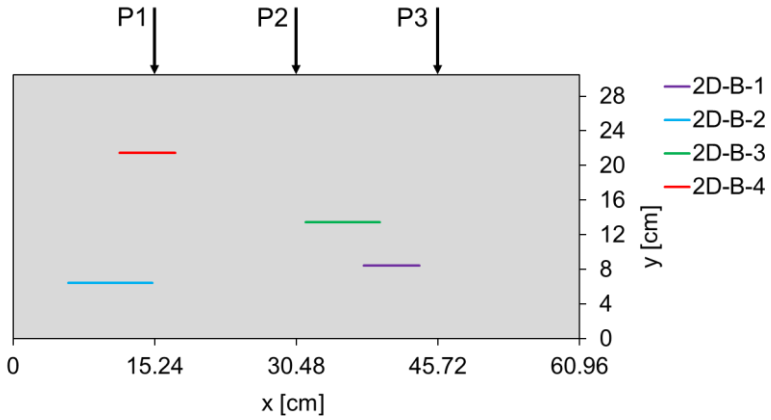


Figure 49: The domain, the three pump/probe excitation locations, and crack locations used in FE simulation 2D-B.

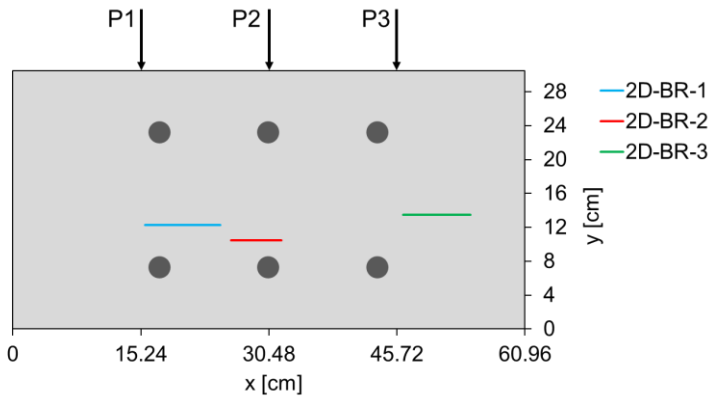


Figure 50: The domain, the three pump/probe excitation locations, and crack locations used in FE simulation 2D-BR

Table 16: VAM test parameters for the FEM simulations.

Crack Case	Centroid (x-dir) [m]	Total Training Points for Crack	Length [m]	Depth from Surface (z-dir) [m]	R_{amp}	R_{freq}	Pump/Probe Actuator Location [m]
2D-A-1	-0.1025	13,896	0.06	0.0762	1, 0.5, 0.2, 0.1	20	-0.1524, 0.0, 0.1524
2D-A-2	-0.1524	13,896	0.06	0.1016	1, 0.5, 0.2, 0.1	20	-0.1524, 0.0, 0.1524
2D-A-3	0.1524	6,369	0.02	0.1016	1, 0.5, 0.2, 0.1	20	-0.1524, 0.0, 0.1524
2D-A-4	0.0300	1,560	0.06	0.0850	1, 0.5, 0.2, 0.1	20	0.0
2D-A-5	0.1016	780	0.09	0.0508	1, 0.1	20	0.1524
2D-A-6	0.1016	2,340	0.09	0.0200	1, 0.1	20	0.0, 0.1524
2D-B-1	0.1024	5,256	0.06	0.2200	1, 0.1	18, 20, 22	-0.1524, 0.0, 0.1524
2D-B-2	-0.2000	2,316	0.09	0.2400	1, 0.1	18, 20, 22	-0.1524, 0.0, 0.1524
2D-B-3	0.0500	1,576	0.08	0.1700	1, 0.5	20, 22	-0.1524, 0.0, 0.1524
2D-B-4	-0.1600	2,352	0.06	0.0900	1, 0.1	18, 20	-0.1524, 0.0, 0.1524
2D-BR1	-0.1025	4,680	0.09	0.1824	1, 0.1	16, 18, 20, 22	-0.1524, 0.0, 0.1524
2D-BR-	-0.0150	2,340	0.06	0.0200	1, 0.1	18, 20, 22	-0.1524, 0.0, 0.1524
2D-BR-3	0.2000	2,730	0.08	0.1700	0.5, 0.1	18, 20, 22	-0.1524, 0.0, 0.1524

5.4.2. Laboratory specimen

The laboratory specimen (EXP-A) that was used for validation in Chapter 2, 3 and 4 is used in this study as a source of training data. Since this is a laboratory specimen and not simulated, the data is treated as the highest fidelity data source for training the ML models. This 60.96 x 60.96 x 15.24 cm concrete specimen was cast and cured in ASR-inducing environment at Vanderbilt University and contained pockets of reactive specimen at four known locations. Numerous VAM tests were performed on this specimen by installing accelerometers (sensors) on its top surface and by varying VAM test parameters (see Chapter 2, Section 2.4.2). To use EXP-A for training, each sensor used in VAM tests needs to be identified as a sensor in the vicinity of damage or not in the vicinity of damage {1,0}. Using the known information about the specimen, the cores taken, and observations of internal slab surface after breaking open the plain concrete slab [107], the *true* damage indicator value at each sensor is ascertained. Figure 32 (b) shows the true damage classification results to be used in training the ML models in this Chapter. Experimental test parameters used for this specimen can be found in Table 17. Additional information about this specimen and the tests conducted in it can be found in Section 2.4.

Table 17: The 240 VAM test configurations run on EXP-A.

Amplitude Ratio (probe/pump) R_{amp}	**Frequency Ratio (probe/pump) R_{freq}**	**Pump/Probe Actuator Location [m]**
1, 0.5, 0.2, 0.1	10.87, 11.96, 13.04, 14.13, 15.21, 16.30, 17.39, 18.48, 19.57, 20.65, 21.74	(0.4572, 0.4572), (0.1524, 0.4572), (0.1524, 0.1524), (0.4572, 0.1524), (0.3048, 0.3048)

5.4.3. Numerical simulation data for Verification of the Classification models

A 3D FE model simulating VAM tests for a single hidden crack is used to generate data for model verification. The size of 3D computational domain is 60.96 X 30.48 cm X 30.48 cm. The domain geometry and one of the pump and probe excitation locations used in the numerical simulation are shown in Figure 51. These are chosen to match the cement slab specimen used for validation experiments. The VAM parameters used for verification test were: an R_{freq} of 20, an R_{amp} of 1, and pumping and probing excitation actuators are placed at the center of the specimen (0, 0, 0) cm. The location of true damage had a centroid of (-15.24, 0) cm and a depth of 15.24 cm from the surface. Abaqus was used to perform numerical simulations, and the material properties can be found in Table 15.

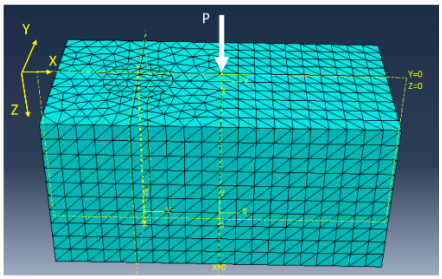


Figure 51: FE simulation for model verification, domain geometry and the location of pumping and probing excitation actuators.

5.4.4. Damage diagnosis and Evaluation of PIML Models

The presence/absence of damage at each sensor is established by comparing the results using the machine learning models to the ground truth. Four metrics were used: sensitivity, specificity, accuracy, and F1 score. Descriptions and formulas for these values can be found in Sections 3.3.2 and 4.3.4.

5.5. PIMLs for Damage Diagnosis

5.5.1. Classification model for crack diagnosis on X-Y plane

The classification models were built using a combination of methodologies, described in Section 5.3.1. All of the classification models in this section have 3 hidden layers or less. 20% of this data was then randomly selected to be used for cross-validation and not included within the training data. For all models, the conjugate gradient method with Powell/Beale Restart algorithm [80] was used as the training algorithm and the cross-entropy loss function was used. The SoftMax [81] activation function was applied to the output. Model I consists of 2 layers with 16 nodes in the first and 8 in the second and is trained with only 2D-A data. Model II is trained by combining 2D-A and 2D-B. It has 2 layers with 16 nodes in the first and 10 in the second. Model III is trained using the transfer learning approach, with the first two layers (16 and 8 nodes) being trained with 2D-A (Model I) and the last layer being trained by 2D-B with 7 nodes. Model IV is a multi-fidelity model with the first two layers being Model II and the last layer being trained by EXP-A with 8 nodes. Model V trains by combing the data sets using the multi-fidelity approach. Layer 1, trained used 2D-A has 20 nodes, layer 2 trained using 2D-B has 14 nodes, and layer 3 trained using 2D-BR has 8 nodes. Model VI is combined using the multi-fidelity approach with the first two layers being Model II and the last layer being trained by 2D-BR with 8 nodes. Model VII combines all 2D FE simulations with 2 layers of 18 and 11 nodes each. Model VIII is a multi-fidelity model with the first two layers being Model VII and the last layer being trained by EXP-A with 8 nodes. The training and verification evaluation metrics for all models can be found in Table 18. Models I-III are similar to the models trained in Chapter 3, where the input is a non-reinforced 2D concrete domain with a single crack. These models perform well for both training and verification. The multi-fidelity approach explored in training of Models III and IV have a lower F1 score for the

training performance, but a higher F1 score for verification. The decrease in the training F1 score for all models after Model I, is due to the addition of different computational domains and data sources. As we add varying training information to the input data, we expect the training performance to decrease because the relationship between the input and output has become more complex. In contrast, the verification model performance increases with the addition of the new training data because these inputs come from domains that are more similar to the domain being verified (depth, dimensionality). The addition of simulation data from the 2D-BR model in the training process for Models V, VI and VII does not significantly improve the diagnostic performance of the PIML models. Although the models had a high verification accuracy and F1 score in some cases, the training values were very poor. It is important that we examine how these models perform for the validation of the reinforced specimens (reinforcement is not considered in the verification model). Since the excitation frequencies are propagating through both concrete and reinforcement in 2D-BR FE simulation, the signal dissipates differently than if there was no reinforcement at all (2D-A and 2D-B).

Table 18: Model training and verification performance. The top value (red) is for the cross-validation values calculated when training the model, the values below (green) are the verification results.

Machine Learning Model				Model Evaluation			
Label	**Hidden Layer 1**	**Hidden Layer 2**	**Hidden Layer 3**	**Sensitivity**	**Specificity**	**Accuracy**	**F1 Score**
I	2D-A		-	0.648 0.489	0.811 0.868	0.727 0.798	0.710 0.472
II	2D-A + 2D-B		-	0.289 0.438	0.972 0.913	0.849 0.825	0.408 0.480
III	2D-A		2D-B	0.209 0.512	0.977 0.918	0.880 0.853	0.306 0.526
IV	2D-A + 2D-B		Exp-A	0.145 0.557	0.939 0.899	0.636 0.845	0.234 0.533
V	2D-A	2D-B	2D-BR	0.038 0.305	0.999 0.983	0.869 0.875	0.073 0.437
VI	2D-A + 2D-B		2D-BR	0.045 0.458	0.996 0.938	0.868 0.862	0.084 0.513
VII	2D-A + 2D-B + 2D-BR		-	0.163 0.350	0.981 0.941	0.849 0.832	0.259 0.434
VIII	2D-A + 2D-B + 2D-BR		EXP-A	0.088 0.305	0.969 0.981	0.633 0.874	0.154 0.435

5.5.2. Regression model for depth diagnosis

The regression model uses the VAM input parameters as the inputs to predict the VAM damage index SBSum. The inputs for the model are determined by their importance in predicting the SBSum using analysis of variance (ANOVA). ANOVA results indicated that all of the input parameters examined were statistically significant, having a p-value less than 0.05 (Section 3.4.2). Thus a surrogate model for the VAM test is built using 7 inputs: the ratio of f_{probe} to f_{pump} (R_{freq}), the ratio of the amplitude of the probing excitation frequency (Amp_{probe}) to the amplitude

of the pumping excitation frequency (Amp_{pump}) (R_{Amp}), the location of the crack (x_{crack}), the length of the crack (L_{crack}), the depth of the crack (z_{crack}), the location of the pumping and probing excitation frequency (x_{pp}), and the location of the sensor (x_{sensor}). Given these inputs defining the damage and test parameters as well as the sensor location, the output is the damage index (*SBSum)* at the sensor on interest. The architecture for the model can be seen in Figure 52. Note that this technique provides crack depth prediction (Z-direction) for a (2D) cross section of a 3D specimen. Therefore, for a 3D test specimen multiple regression model-based diagnoses need to be performed for multiple vertical cross-sections.

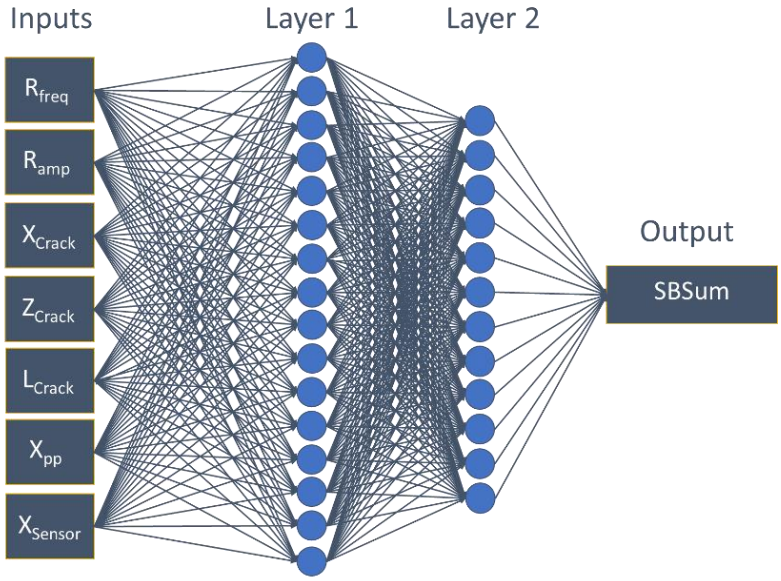


Figure 52: Regression model with seven inputs (VAM test and damage parameters), two hidden layers, the first with 16 nodes and the second with 12, and 1 output of the SBSum.

During training data pre-processing, all input and output values were normalized with a mean of zero and a standard deviation of 1. Multiple training algorithms, the number of layers, and nodes in each layer were examined to determine the best model. This model was trained by combining

all of the data for the 2-D training data (2D-A, 2D-B, 2D-BR) into one data set. 20% of this data was then randomly selected to be used for cross-validation and not included within the training data. There were 18,486 data points for training and 4621 used for cross-validation. The best training algorithm was the Levenberg-Marquardt method, with a learning rate of 0.1 and the least squares loss function [76], [77]. There were two layers trained, with 16 nodes in the first layer and 12 in the second. The model converged after 230 epochs. Early stopping [78], using information about the error of the cross-validation, was used to prevent overfitting. The best value of mean square error was found to be approximately 0.326.

5.6. Validation

This section examines four different experimental specimens used in validating the machine learning models. The first specimen is referred to as the control specimen. The control specimen is a plain concrete specimen with reactive aggregates evenly distributed throughout, but no boosting agent (NaOH) added to induce ASR. The other three specimens contained reactive aggregates distributed throughout and sodium hydroxide (NaOH) to boost the alkali loading to 5.25 kg/m3 to accelerate the production of ASR. For the casting of all the specimens, conventional concrete mixtures prepared in a rotating drum mixer was used. The aggregates were distributed throughout the blocks. All four specimens were of size $60.96 \times 30.48 \times 30.48$ cm and contained 25 mm coarse aggregates known to be susceptible to ASR (reactive) combined with a non-reactive crushed dolomite/calcite fine aggregate. All four specimens contained a reactive crushed greenschist from North Carolina. All four specimens contained low-alkali cement (0.46% Na2Oeq), while three contained added sodium hydroxide (NaOH) to boost the alkali loading to 5.25 kg/m3 to accelerate ASR and the control specimen did not. Two of the boosted specimens

were reinforced with rebar, the first unidirectionally and the second bidirectionally. The specimens are labeled as such; control (reactive aggregates, not boosted), ASR non-reinforced (reactive aggregates, boosted), ASR unidirectionally reinforced (reactive aggregates, boosted), and ASR bidirectionally reinforced (reactive aggregates, boosted), and are discussed below.

In conducting VAM testing on these specimens, compared to the slab specimen tested in [48], the crack locations are less likely to be concentrated in known quadrants. To account for this, and given the key conclusions from our previous work [108] where the SBSum was best used for damage localization when the pumping and probing actuators were located above a damage location, 10 pumping and probing configurations were used for each specimen. The locations of the actuators and sensors are displayed in Figure 53. We measure the response of the structural component of interest using 21 accelerometers (100 mV/g) placed on the component's surface. The relative magnitudes of sidebands at various accelerometer locations are calculated. The sensors were calibrated between different testing specimens to ensure that they provide accurate measurements. In these experiments, we used pumping frequency, f_{pump} of 1 kHz. For the given specimens, the fundamental frequency was much higher than the specimen testing in [48], meaning the ideal probing excitation frequency, f_{probe}, 20 times larger than f_{pump} would be too large for our actuators to read accurately. We conducted the experiments using f_{probe} of 16, 17, 18, 19, 20, and 21 kHz. We kept the amplitude of the pumping excitation constant and varied the probing excitation amplitude (1/2, 1/5, and 1/10 times the pumping excitation amplitude). The presence and severity of ASR-induced damage in each of the specimens was validated using the petrographic method, damage rating index, (DRI)

a. b.

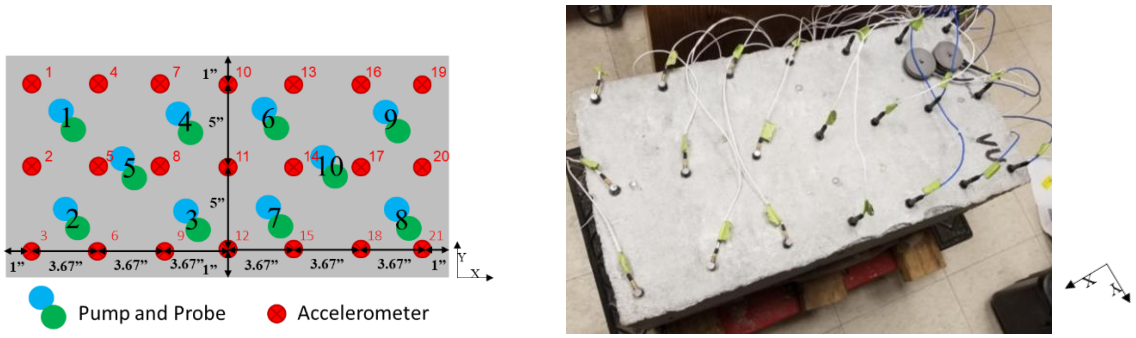


Figure 53: Locations of accelerometers and pumping and probing excitation actuators on the surface of the validation specimens: (a) diagram of the surface of the specimens with all 21 sensors tables, and all 10 pumping and probing excitation actuator locations; (b)experimental setup on the surface of the specimen for pumping and probing excitation actuator location 8.

5.6.1. Damage Rating Index

In order to establish the ground truth for each validation specimen, we need to establish damage presence or absence in the vicinity of the sensor. We use a well-known ASR damage metric, namely, damage rating index (DRI), to determine the severity of ASR-induced damage in select core locations for each specimen.

The DRI metric is obtained by a petrographic analysis method to quantify the extent of damage in ASR-affected concrete. The core samples were photographed and examined visually and using stereo-optical microscopy. A cross-sectioned core was cut, and the cut face was coated with fluorescent-dyed epoxy and then polished for optical microscopy examination. DRI analysis was performed on the bottom 6” portion of the polished cross-section slabs following the procedure outlined by Grattan-Bellew[109]. Using a stereo binocular microscope at 15x, features indicative of ASR reaction were identified and weighted according to values listed in Table 19. The results were normalized for an area of 100 cm^2. The features in this table are based on typical characteristics of ASR reaction products, and procedures for the petrographic examination are

outlined in ASTM C 856 [110]. A reaction rim is when the exterior edge of the aggregate that is in contact with the paste reacts with the alkalis in the pore solution. The ASR forms along the outer edge or rim of the aggregate. Additional definitions can be found in [111].

In general, DRI values above 50 suggest significant damage due to ASR. DRI can have variable results as different petrographic examiners examine samples. These specimens were all evaluated by the same petrographer, which helps reduce this variability and allows us to compare DRI values across all the cores. A study by Sanchez 2020 [103] showed a correlation between the DRI number and expansion levels attained by the specimens. Using this information, we assume that a core with a higher index contains more ASR-induced damage, with damage over 50 being associated with severely damaged areas.

Table 19: Weighting factors for DRI.

Feature	Weight Factor
Cracks in coarse aggregate	0.75
Cracks in coarse aggregate + gel	2.00
Open cracks in coarse aggregate	4.00
Reaction rims	0.50
Coarse aggregate debonded	3.00
Paste with cracks	2.00
Paste with cracks + gel	4.00
Gel in air voids	0.50

5.6.2. Control Specimen

The control specimen was cast using conventional concrete mixtures prepared in a rotating drum mixer, contained no additional boosting induce ASR and cured under ASR-inducing conditions (high heat and humidity) at The University of Nebraska-Lincoln, Omaha, Nebraska. Four steel ties remained in the specimen from the formwork. This specimen did not show

significant expansion during the curing process (-0.004%). Since specimen expansion is a key characteristic of ASR gel production in a concrete specimen, it is assumed that ASR damage wasn't present in the specimen.

5.6.3. Alkali-boosted Non-reinforced Specimen

This specimen was cast using conventional concrete mixtures prepared in a rotating drum mixer with boosting to induce ASR, and cured under ASR-inducing conditions (high heat and humidity) at the University of Alabama, Tuscaloosa, Alabama. The specimen showed significant mechanical expansion over the curing process (0.424%). After conducting VAM tests on the specimen, two cores (Figure 54) were taken out and sent for petrographic analysis at the R.J. Lee group in Monroeville, Pennsylvania. The core locations for this specimen are noted in Table 20. The DRI for core 1 (EXP-B-C1) was 73, and for core 2 (EXP-B-C2) was 58. The petrographic analysis showed severe ASR-induced damage at both core locations.

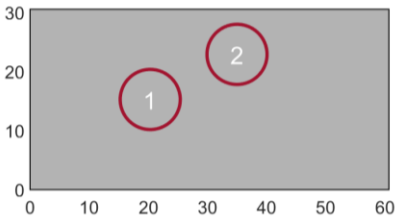


Figure 54: Core locations for the alkali-boosted non-reinforced concrete specimen (cm).

5.6.4. Alkali-boosted Unidirectionally Reinforced Specimen

This specimen was cast using conventional concrete mixtures prepared in a rotating drum mixer with boosting to induce ASR, and cured under ASR-inducing conditions (high heat and

humidity) at The University of Nebraska-Lincoln, Omaha, Nebraska. Four steel ties remained in the specimen from the formwork. The locations of the rebar reinforcement can be seen in Figure 55. The specimen showed significant expansion over the curing process (0.349%). After conducting VAM tests on the specimen, two cores (Figure 56) were sent for petrographic. The core locations for this specimen are noted in Table 20. The DRI for core 3 (EXP-BR1-C3) was 29, and for core 4 (EXP-BR1-C4) was 34. The petrographic analysis showed a small amount of ASR-induced damage at each core location.

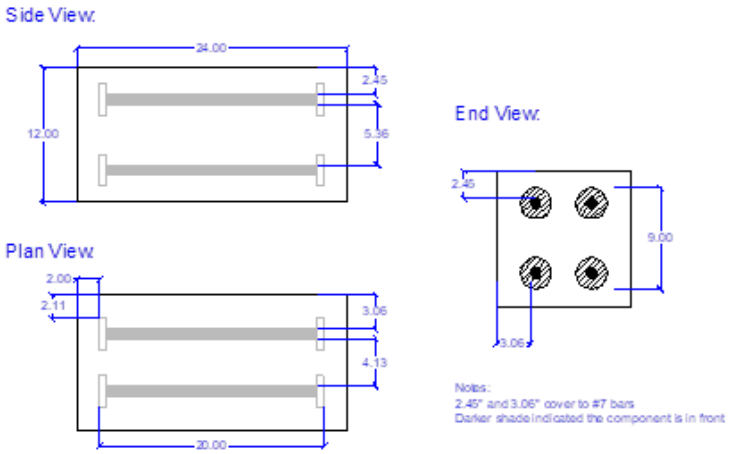


Figure 55: Rebar locations for the alkali-boosted unidirectionally reinforced concrete specimen.

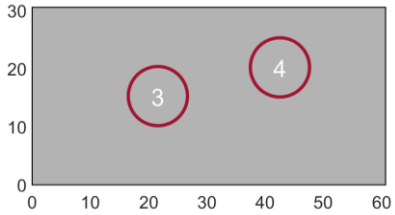


Figure 56: Core locations for the alkali-boosted unidirectionally reinforced concrete specimen (cm).

5.6.5. Alkali-boosted Bidirectionally Reinforced Specimen

This specimen was cast using conventional concrete mixtures prepared in a rotating drum mixer with boosting to induce ASR, and cured under ASR-inducing conditions (high heat and humidity) at The University of Nebraska-Lincoln, Omaha, Nebraska. Four steel ties remained in the specimen from the formwork. The locations of the rebar reinforcement can be seen in Figure 57 and Figure 58. The specimen showed significant expansion over the curing process (0.452%). After conducting VAM tests on the specimen, two cores (Figure 59) were sent for petrographic analysis. The core locations for this specimen are noted in Table 20. The DRI for core 5 (EXP-BR2-C5) was 70, and for core 6 (EXP-BR2-C6) was 51. The petrographic analysis showed both core locations were severely damaged by ASR.

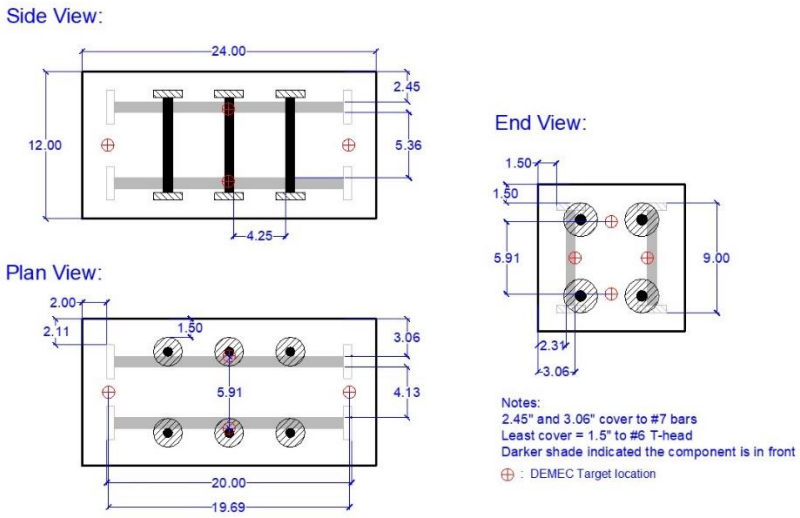


Figure 57: Rebar locations for the alkali-boosted bidirectionally reinforced concrete specimen.

Figure 58: Plan view of two-directional rebar reinforcement (with ties) for the alkali-boosted bidirectionally reinforced concrete specimen.

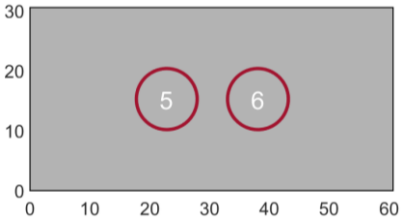


Figure 59: Core locations for the alkali-boosted bidirectionally reinforced concrete specimen (cm).

Table 20: Core locations for each specimen and DRIs.

Core Label	Centroid (cm)	DRI
EXP-B-C1	(20.32, 15.24)	73
EXP-B-C2	(35.00, 22.86)	58
EXP-BR1-C3	(21.59, 15.24)	29
EXP-BR1-C4	(42.55, 20.17)	34
EXP-BR2-C5	(22.86, 15.24)	70
EXP-BR2-C6	(38.10, 15.24)	51

5.7. Performance Evaluation of PIML Model

5.7.1. Laboratory Experiments

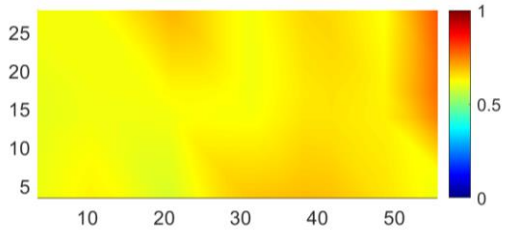
In conducting VAM testing on these experimental specimens, compared to the slab specimen tested in [48], the ASR-induced cracks are not likely to be present near a pocket of reactive

aggregates, but could be present anywhere in the specimen. To account for this, and given the key conclusions from [108] where the SBSum was best used for damage localization when the pumping and probing actuators were located above a damage location, 10 pumping and probing configurations were used for each specimen. The probing excitation frequency was {16, 17, 18, 19, 20} kHz and the pumping excitation frequency was 1kHz for all specimens, The amplitude ratios used were {1, 0.5, 0.2, 0.1} probe/pump.

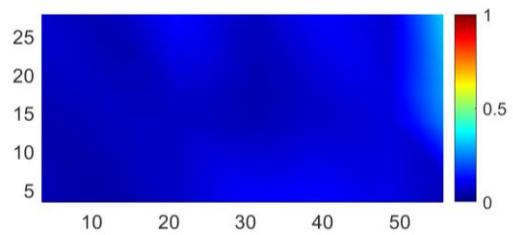
5.7.2. Control

Each of the eight classification models trained in Section 5.5.1 were evaluated on the control specimen. The results are displayed in Figure 60. Since the specimen showed no deformation or signs of ASR-induced damage, we expect to see a low to no damage probability. It is clear that Model I performed poorly, as it identified damage. The reason Model I provides a high damage probability is because the model is trained to look at the relationship between the SBSum and the mean of the SBSum for that specimen without accounting for the true value of the SBSum. Although the value of SBSum is small in this case, the relationship between the value at a given sensor and the mean value causes the model to predict the presence of damage at some sensors. The model was not trained for the zero-probability case. Models V and VI performed the best and had damage probabilities of zero. All other models also successfully did not identify much damage.

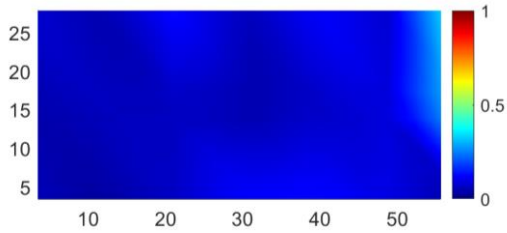
Model I



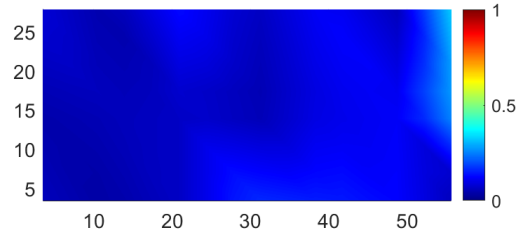
Model II



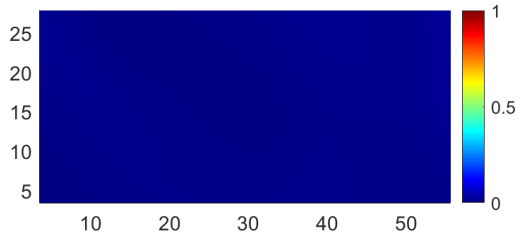
Model III



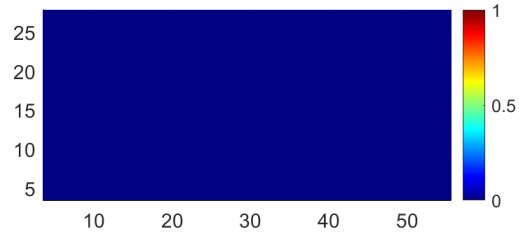
Model IV



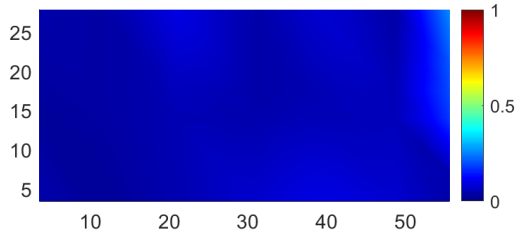
Model V



Model VI



Model VII



Model VIII

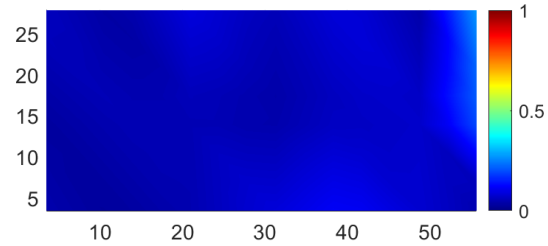
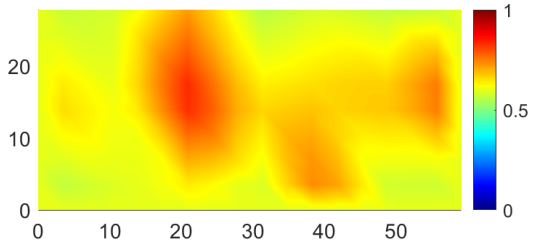


Figure 60: Results of the eight classification models evaluated on the control specimen.

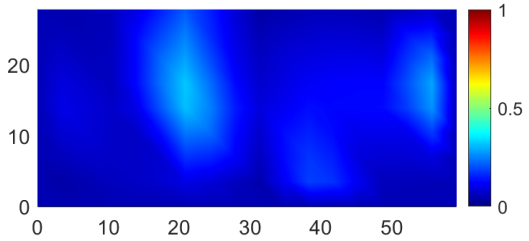
5.7.3. Alkali-boosted Non-reinforced Specimen

Each of the eight classification models trained in Section 5.5.1 was evaluated on the alkali-boosted non-reinforced specimen. The results are displayed in Figure 61. Since the specimen had a DRI of 73 for core 1 and 58 for core 2, we expect to detect a high probability of damage over core 1, and a slightly lower value near core 2. Model V and VI performed the worst as almost no damage was predicted about core 1. Model I performed the best, and the other models had a higher probability above core 1, despite their low probabilities overall.

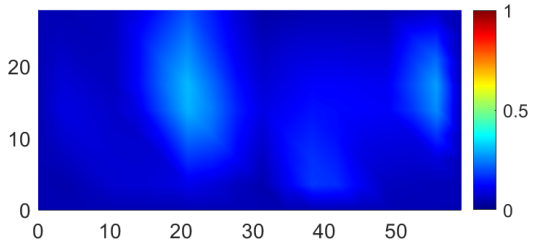
Model I



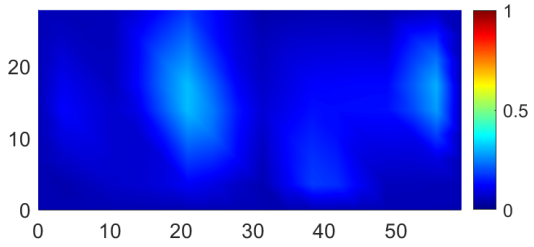
Model II



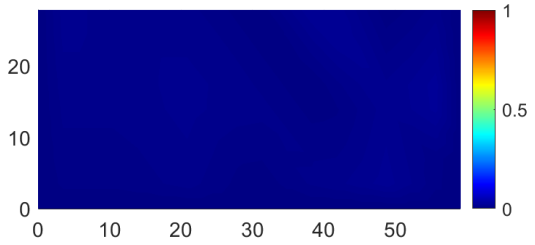
Model III



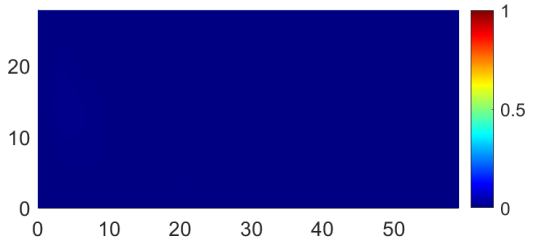
Model IV



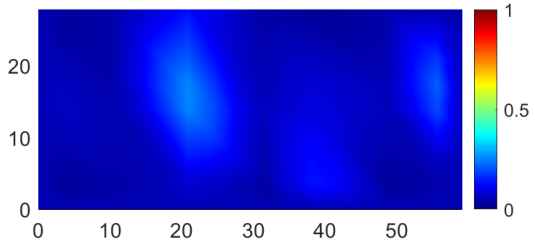
Model V



Model VI



Model VII



Model VIII

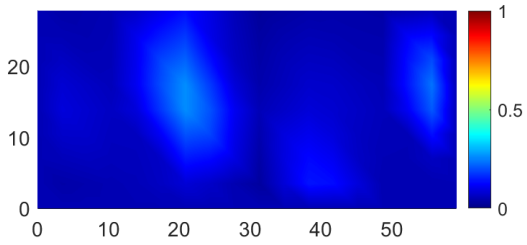


Figure 61: Results of the eight classification models evaluated on the alkali-boosted non-reinforced specimen.

Given the damage probability map for Model I, a crack is assumed to be located near the region of high damage probability. As per the methodology discussed in Section 3.3.1.4 a vertical slice is taken along the center of the specimen (containing the selected crack) at y=15.24 cm (Figure 62 (a)). The centroid is 20.32 cm in the x-direction, and the length is estimated as 10.16 cm. Figure 62 (b) shows the vertical slice cross-section containing the crack with the unknown depth, z_{crack}. 5000 MH samples are taken, and the prior of z_{crack} is assumed to be a uniform distribution, U[0.00001, 30.48] cm. The crack's depth is estimated at about 13 cm from the surface (Figure 63).

(a)

(b)

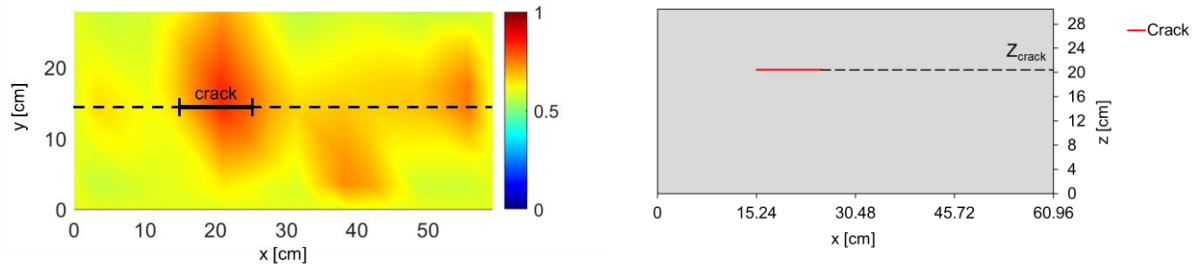


Figure 62: (a) Damage probability map for Model I with a vertical slice at y=15.24 cm; (b) Cross-section of the vertical slice with a 10.16 cm crack at centroid x=20.32 cm and an unknown depth, z_{crack}.

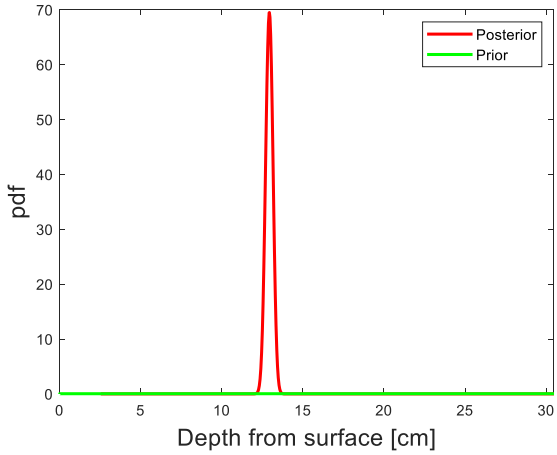
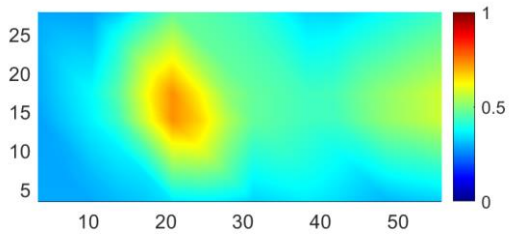


Figure 63: Posterior for damage depth of a crack estimated in the alkali-boosted non-reinforced specimen.

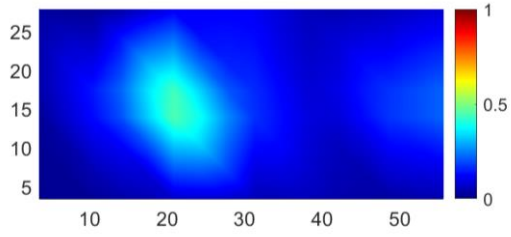
5.7.4. Alkali-boosted Unidirectionally Reinforced Specimen

Each of the eight classification models trained in Section 5.5.1 was evaluated on the unidirectionally reinforced specimen. The results are displayed in Figure 64. Since the specimen had a DRI of 29 for core 3 and 34 for core 4, we expect to detect a low probability of damage at both core locations. Models V and VII show almost no damage probability for the specimen. Model 1 has a fairly high damage probability, but that is to be expected since significant ASR damage is present in the specimen. The remaining models show a much higher damage probability at the core 3 location but otherwise detect very little damage elsewhere in the specimen.

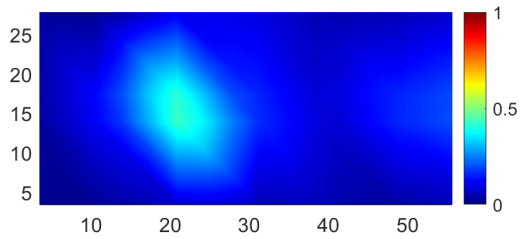
Model I



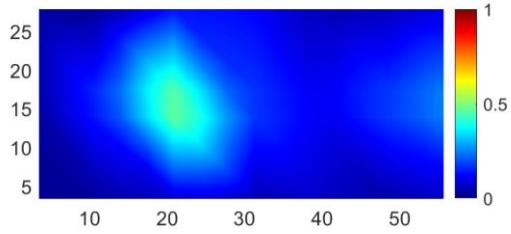
Model II



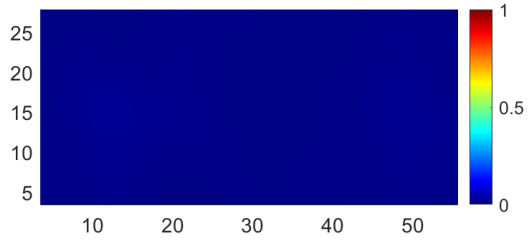
Model III



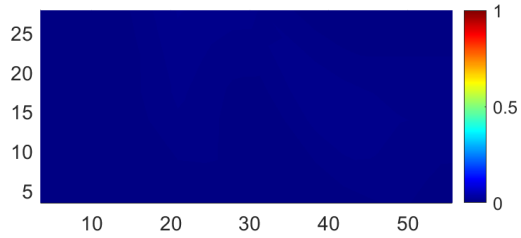
Model IV



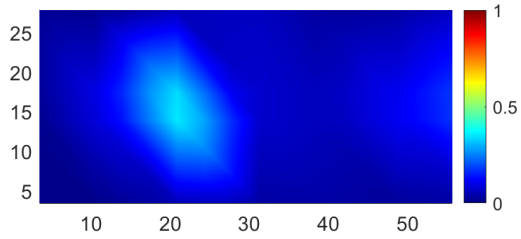
Model V



Model VI



Model VII



Model V III

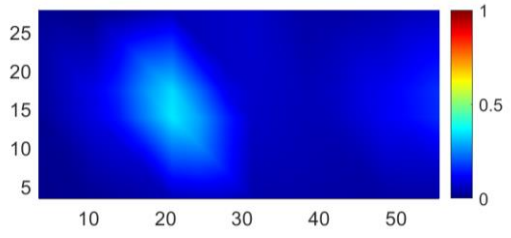


Figure 64: Results of the eight classification models evaluated on the alkali-boosted unidirectionally reinforced specimen.

Given the damage probability map for Model I, a crack is assumed to be located near the region of high damage probability. A vertical slice is taken along the center of the specimen (containing the selected crack) at y=15.24 cm (Figure 65 (a)). The centroid is 23 cm in the x-direction, and the length is estimated as 7.6 cm. Figure 65 (b) shows the vertical slice cross-section containing the crack with the unknown depth, z_{crack}. 5000 MH samples are taken, and the prior of z_{crack} is assumed to be a uniform distribution, U[0.00001, 30.48] cm. The crack's depth is estimated at 10 cm from the surface (Figure 66).

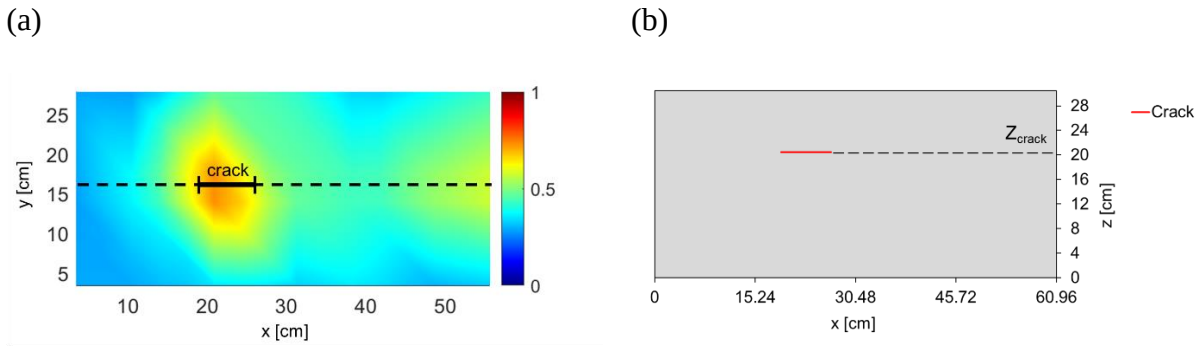


Figure 65: (a) Damage probability map for Model I with a vertical slice at y=15.24 cm; (b) Cross-section of the vertical slice with a 7.6 cm crack at centroid x=23 cm and an unknown depth, z_{crack}.

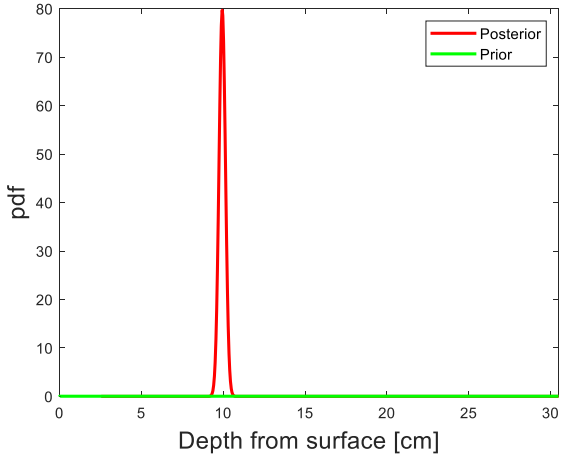
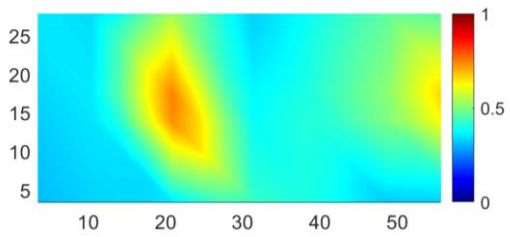


Figure 66: Posterior for damage depth of a crack estimated in the alkali-boosted unidirectionally reinforced specimen.

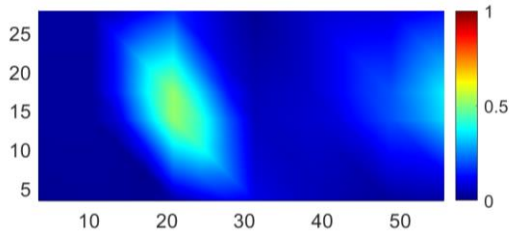
5.7.5. Alkali-boosted Bidirectionally Reinforced Specimen

Each of the eight classification models trained in Section 5.5.1 was evaluated on the unidirectionally reinforced specimen. The results are displayed in Figure 67. Since the specimen had a DRI of 70 for core 5 and 51 for core 6, we expect to detect a high probability of damage at both core locations, especially at core 5. Model I performs best for damage localization. Models II, III, and IV also have fairly high damage probabilities, and they are higher than the unidirectionally reinforced case, which is what we expect given the DRI values.

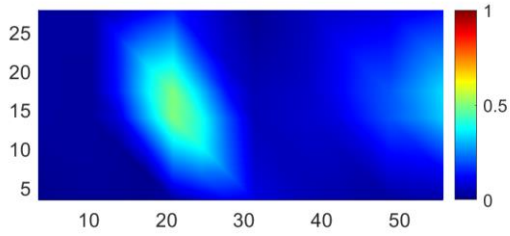
Model I



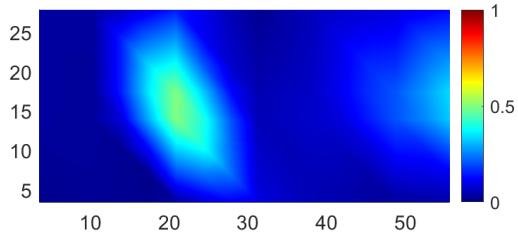
Model II



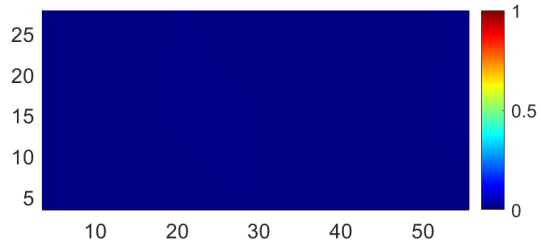
Model III



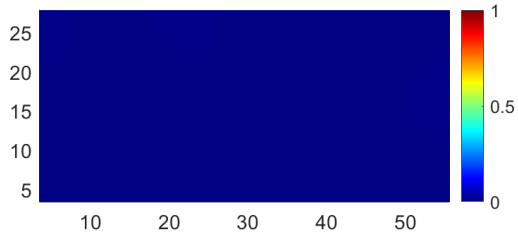
Model IV



Model V



Model VI



Model VII

Model V III

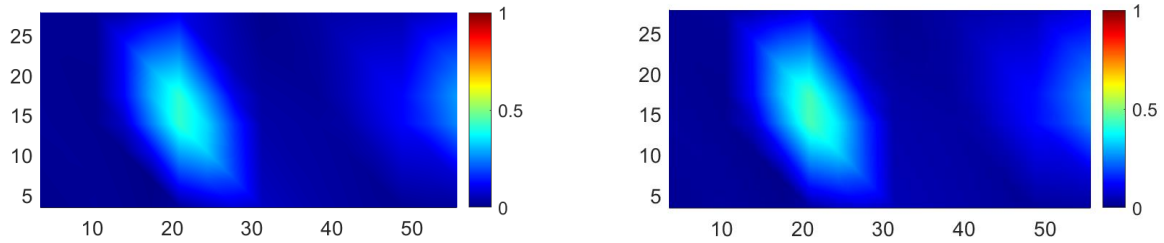


Figure 67: Results of the eight classification models evaluated on the alkali-boosted bidirectionally reinforced specimen.

Given the damage probability map for Model I, a crack is assumed to be located near the region of high damage probability. A vertical slice is taken along the center of the specimen (containing the selected crack) at y=15.24 cm (Figure 68 (a)). The centroid is 22 cm in the x-direction, and the length is estimated as 7.6 cm. Figure 68 (b) shows the vertical slice cross-section containing the crack with the unknown depth, z_{crack}. 5000 MH samples are taken, and the prior of z_{crack} is assumed to be a uniform distribution, U[0.00001, 30.48] cm. The crack's depth is estimated at 6.5 cm from the surface (Figure 69).

(a)

(b)

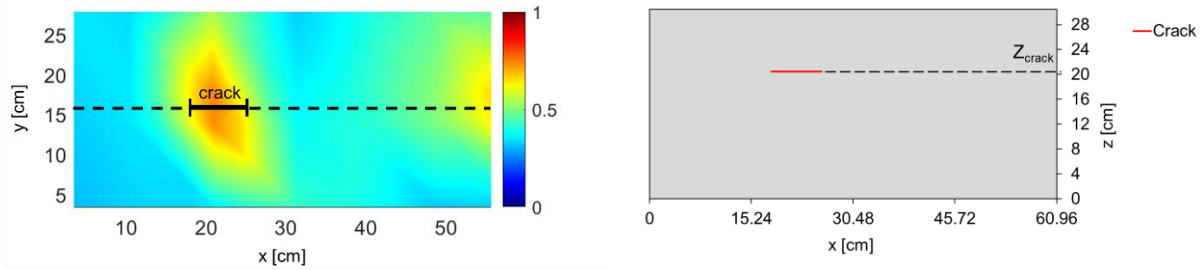


Figure 68: (a) Damage probability map for Model I with a vertical slice at y=15.24 cm; (b) Cross-section of the vertical slice with a 7.6 cm crack at centroid x=22 cm and an unknown depth, z_{crack}.

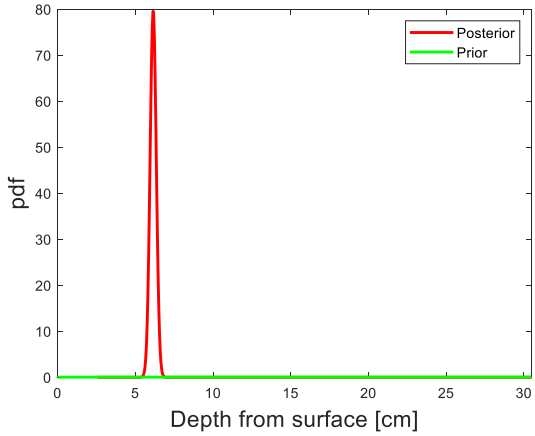


Figure 69: Posterior for damage depth of a crack estimated in the alkali-boosted bidirectionally reinforced specimen.

5.7.6. Discussion of Results

The damage probability calculated from the PIML models for each core location is compared to the DRI for that core in Table 21. Since DRI values above 50 are generally indicative of severe ASR-induced damage in concrete, we can expect the best-performing models to have a high probability of damage at the core locations that have DRI values over 50. Using this metric for comparison, Model I performs the best for localizing the ASR-induced damage. That is, the damage probability value at the core location (in percentage) is high when the DRI value for the core is high. Additionally, the damage probabilities for Models II, III, IV, VII, and VIII are smaller for the significantly damaged cores than Model I; however, the relationship between a higher probability of damage and a higher DRI is still present. Model I has a higher damage probability because the training data was less "diluted" with other VAM test parameters. The training data for Model I had more information coming from VAM tests of different cracks and not just other parameters (R_{freq}, Ramp, etc.), which may not be as important in the ML models as the variation in damage locations. The main drawback of Model I is its poor performance for the control

specimen; the model predicted a fairly high damage probability. It can be noted that the absolute value of SBSum is an order of magnitude smaller in the control specimen than in all of the specimens with induced damage. A filter can be applied before the data is analyzed with the ML model to ensure the SBSum value is high enough to continue with the model, or an additional input adjusting for the difference between SBSum values across samples needs to be included. Note that the damage depth predicted using regression models indicates that the damage was closer to the top surface (in the top half of the core, Figure 63, Figure 66, and Figure 69). The cracks modeled in 2D-B and 2D-BR are primarily located deeper within the specimen. Additional finite element simulations for shallower cracks in the two computational domains (2D-B and 2D-BR) should be included within the training data to improve these models.

Table 21: Damage probability for each PIML model compared to the DRI at each core.

Specimen	**Core Number**	**Model Number**	**Damage Probability (from PIML)**	**DRI**
Non-Reinforced (EXP-B)	1	I	0.76	73
		II	0.23	
		III	0.22	
		IV	0.23	
		V	0.01	
		VI	0.00	
		VII	0.19	
		VIII	0.20	
	2	I	0.64	58
		II	0.07	
		III	0.08	
		IV	0.13	
		V	0.01	
		VI	0.00	
		VII	0.05	
		VIII	0.04	
Unidirectional Reinforcement (EXP-BR1)	3	I	0.73	29
		II	0.44	
		III	0.42	
		IV	0.46	
		V	0.01	
		VI	0.01	
		VII	0.34	
		VIII	0.34	

	4	I	0.47	34
		II	0.17	
		III	0.16	
		IV	0.17	
		V	0.00	
		VI	0.00	
		VII	0.07	
		VIII	0.07	
Bidirectional Reinforcement (EXP-BR2)	5	I	0.75	70
		II	0.52	
		III	0.50	
		IV	0.52	
		V	0.01	
		VI	0.00	
		VII	0.42	
		VIII	0.44	
	6	I	0.39	51
		II	0.06	
		III	0.06	
		IV	0.05	
		V	0.00	
		VI	0.00	
		VII	0.03	
		VIII	0.03	

5.8. Conclusion

Damage localization in concrete is a complex problem in which machine learning can help to aid in automation and accuracy. This chapter investigated the performance of PIML models using VAM on realistic concrete structures. We examined classification and regression models trained using 2-dimensional FE simulations and some experimental test data to localize hidden cracks in (3-dimensional) concrete test specimens, with dispersed reactive aggregates and with unidirectional and bidirectional steel reinforcement. Overall, rebar did not appear to adversely affect the diagnostic performance; however, when adding rebars in the 2D FE simulations, model performance decreased in validation. Model I, trained using numerous low-fidelity (2D) FE simulation data (and no experimental data), appeared to perform the best on all validation specimens except the control specimen. This might be because data in 2D-A contained many

different crack locations with less variability in the other testing parameters (such as R_{freq}); more damage locations trained a stronger model. Although some additional models did not yield accurate results for all cases, the addition of 2D-B and EXP-A data did improve the verification results. These models were able to identify the areas of higher damage probability.

To improve this methodology for real world applications, the process needs to be automated, and damage prediction results should be reported in real-time. Currently, giving the classification result on the surface can happen quickly for a VAM test, but the depth prediction model is computationally expensive. The high cost for MCMC sampling could be decreased by exploring a different sampling technique, such as No U-Turn Sampler. Additionally, the models could be improved with the continued addition of physical specimens representing real-world, concrete structures used in training and examining an additional input or filter to ensure a low damage probability when no damage is present (in the entire specimen).

Chapter 6

Summary and Future work

6.1. Summary of accomplishments

This dissertation research investigated VAM for ASR-induced damage detection in concrete structures as large as 0.057 m^3 with unknown damage locations and rebar reinforcement. As noted in the literature review, a probabilistic localization of ASR induced damage in concrete had not been developed in earlier studies. The majority of techniques could only detect ASR induced damage or give a (qualitative, non-probabilistic) zone of possible damage. Thus, we focused on quantitative, probabilistic localization of ASR-induced damage in concrete structures. In Chapter 2 we examined the utility of VAM for a 15.24 cm thick plain concrete specimen with known damage locations and fused the damage index data for all different test parameters to identify the damage locations. The averaging and Bayesian fusion techniques localized the damage based on the sum of magnitudes of side-bands (SBSum) but still had a few drawbacks. The averaging technique did not provide a probabilistic value, and the Bayesian fusion method needed an expert analyst's input to decide a threshold value for SBSum. We addressed this problem in Chapter 3, which developed a classification- and regression-based PIML approach for damage prediction. Since 3D FEA calculations are computationally expensive, the models in this chapter were trained with 2D FE simulations. Both classification and regression models successfully localized the damage for the 15.24 cm thick specimen with localized areas of ASR damage inducement; the regression model also allowed us to localize damage in all three dimensions. Chapter 4 attempted

to improve the models from Chapter 3 by adding data from 3D FE simulations, and investigated how to optimize the computational effort in adding additional simulation data to a training set. We found that adding 50 additional low-fidelity (LF) simulations improved the diagnosis performance more than adding one high-fidelity simulation to the training data. We also discovered that 50 LF simulations were enough to localize the damage correctly. In Chapter 5, we applied the PIML methodology to plain and reinforced concrete specimens with uniformly dispersed aggregates, i.e., unknown ASR damage locations. In Chapter 5 we built the ML models with training data from additional 2D FE simulations for specimens of a larger depth and with simulated rebar reinforcement to successfully localize ASR-induced damage in both plain and reinforced concrete structures with dispersed ASR damage inducement.

Notable contributions of this research include:

- Localized ASR-related damage in a concrete specimen using VAM.
- Examined the optimization of test parameters to improve VAM damage localization in large concrete specimens.
- Built ML models to facilitate VAM internal damage localization.
- Computational physics simulations were used to obtain adequate training data for the ML models.
- Utilized 2D-simulation-data-driven ML models to perform damage diagnosis for real-world (3D) structural components.
- Built a discriminative ML model that classifies binary damage state at a sensor location to minimize analyst intervention (i.e., selecting the threshold).
- Investigated the utility of performing Bayesian damage diagnosis for an ML prediction model built to expedite damage likelihood computation given VAM test data.

- Developed a multi-fidelity PIML model for structural damage diagnosis using a transfer learning approach.
- Analyzed of the trade-offs between computational cost (to perform the computational physics simulations used to train the diagnostic model) and model performance in damage diagnosis.
- Localized damage in plain and reinforced concrete specimens with dispersed and unknown ASR damage inducement.

6.2. Automation

An important aspect of the work conducted and future work discussed in Section 6.3 is the automation of the methodology to localize damage in large concrete specimens. The optimal VAM test parameters (R_{freq}, R_{amp}, pump/probe actuator locations) depend on the location and size of the damage. Since this is typically unknown at the time of performing the test, the proposed methodology employs a multi-configuration approach. That is, VAM tests are performed for multiple test parameter values, and the diagnostic information obtained from multiple VAM tests is fused using a Bayesian method or a simple averaging method to obtain the damage map. To facilitate the multi-configuration testing approach, efforts were made to automate the VAM testing process and minimize human intervention during testing. A MATLAB® program, which communicates with the testing hardware (actuators and accelerometers) and sweeps through the pre-defined values of test parameters (probe frequencies and pump/probe amplitudes) was developed. Thus, after the physical setup of testing equipment is completed, the test engineer can execute the program to obtain the test data corresponding to desired test parameters. The program also provides instant visualization of the VAM test results, as shown in Figure 70.

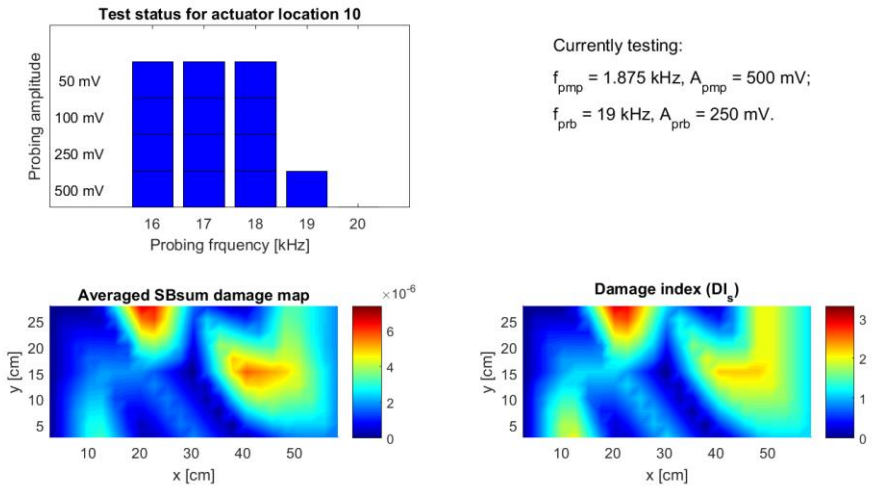


Figure 70: Display of the automated testing process: test status and results

In the above automated test procedure, the testing engineer's involvement is still needed to install accelerometers and actuators on the specimen. A fully automated, non-contact sensing system could be developed using laser Doppler velocimetry. In this case, test engineer intervention may only be required to install the pump/probe actuator at a few locations on the testing surface.

6.3. Future work

The VAM technique, enhanced by ML, is, in general, showing promise in terms of identifying the damage locations. In the future, both diagnostic models (classification and regression) could be improved by adding training data by including different VAM test parameters and additional crack locations and sizes. From an application perspective, future work needs to focus on scaling the technique to field implementation for damage diagnosis in real-size concrete structures. One of the key challenges is the number of sensors (accelerometers) needed. It is not feasible for a large structure to use many accelerometers; therefore, non-contact sensing and full-field observation

techniques (such as laser velocimetry) might be beneficial. Another challenge in testing larger structures is attenuation. VAM tests on larger structures will require higher amplitudes or multiple pump/probe actuator locations to ensure sufficiently strong signals. From a methodological perspective, future work needs to augment the low-fidelity data from 2D FE analyses with a few higher-fidelity 3D FE analyses for generating a multi-fidelity training data set. Including multi-fidelity physics model simulation for generating training data will help build a more accurate machine, learning-based SHM framework; however, as seen in Chapter 4, the selection of the high-fidelity simulation settings makes a substantial difference.

In Chapter 4, the diagnostic test and damage parameters for the high-fidelity physics simulations were selected based on expert knowledge and past research. Future research could formulate and solve an optimization problem for the systematic selection of parameters defining the high-fidelity physics simulations used for training the models to maximize the diagnostic performance within the computational resource constraints. The high-fidelity simulations have a large parameter space, and the criteria for optimal parameter selection involve diagnostic performance, thus making the optimization more complex; previous studies on optimizing the high-fidelity runs have been for forward prediction models [97], [98]. Additionally, this research can be extended by including multiple damage locations and types and varying specimen shapes for generating physics simulation-based training data. The multi-fidelity framework also allows for the inclusion of experimental data for training an additional layer while constructing the multi-fidelity model. All of these analyses will help improve the confidence in applying the developed PIML methodology to the damage diagnosis of real-world structures.

To improve this methodology for real world applications, the process needs to be automated, and damage diagnosis results need to be reported in real-time. Currently, the damage localization

on the 2D plane is fast for a VAM test, but the depth prediction model is computationally expensive due to the use of Bayesian inference and MCMC sampling. The high cost of MCMC sampling could be decreased by exploring a different sampling technique, such as a No U-Turn Sampler. Additionally, the models could be improved with the continued addition of physical specimens representing real-world, concrete structures used in training.

It should be noted that VAM only detects breathing cracks; it cannot detect voids or large openings. Also, it is not possible to determine whether the damage is due to ASR, delamination, shrinkage, or mechanical loading (such as seismic load, large live load, etc.). Fusing information gained from multiple damage diagnosis approaches can help better understand the type and extent of the damage for further localization, damage size estimation, and prognosis.

REFERENCES

[1] M. Azimi, A. D. Eslamlou, and G. Pekcan, "Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review," *Sensors*, vol. 20, no. 10, p. 2778, 2020.

[2] A. Rytter, "Vibrational based inspection of civil engineering structures," Dept. of Building Technology and Structural Engineering, Aalborg University, 1993.

[3] F. J. Ulm, O. Coussy, L. Kefei, and C. Larive, "Thermo-chemo-mechanics of ASR expansion in concrete structures," *J. Eng. Mech.*, vol. 126, no. 3, pp. 233–242, 2000.

[4] K. Kreitman, "Nondestructive Evaluation of Reinforced Concrete Structures Affected by Alkali Silica Reaction and Delayed Ettringite Formation," University of Texas at Austin, 2011.

[5] P. J. Shull, *Nondestructive evaluation: theory, techniques, and applications*, 1st ed. CRC Press, 2016.

[6] S. Multon and F. Toutlemonde, "Effect of applied stresses on alkalia-silica reaction-induced expansions," *Cem. Concr. Res.*, vol. 36, pp. 912–920, Nov. 2006.

[7] P. Bruck, T. Esselman, and M. Fallin, "Digital Image Correlation for Nuclear." Nov-2012.

[8] M. Abdelrahman, M. K. ElBatanouny, P. Ziehl, J. Fasl, C. J. Larosche, and J. Fraczek, "Classification of alkali--silica reaction damage using acoustic emission: A proof-of-concept study," *Constr. Build. Mater.*, vol. 95, pp. 406–413, 2015.

[9] A. A. Pollock, "Classical wave theory in practical AE testing," *Prog. Acoust. Emiss. III-JAP Soc. Non-Destructive Test.*, pp. 708–721, 1986.

[10] ASTM C215, "Standard Test Method for Fundamental Transverse, Longitudinal, and Torsional Resonant Frequencies of Concrete Specimens," 2014.

[11] ASTM C597-16, "Standard Test Method for Pulse Velocity Through Concrete," 2016.

[12] T. Akashi, N. Amasaki, and N. Takagi, "The estimate for deterioration due to alkali aggregate reaction by ultrasonic methods," in *7th International conference on AAR*, 1986, pp. 10–17.

[13] G. Gallo and J. Popovics, "Application of ultrasonic surface waves to characterize ASR damage in concrete.," in *7th Conference of the structural material technology for highways and bridges*, 2006, pp. 301–308.

[14] L. Khazanovich, K. Freeseman, L. Salles, and D. Clayton, "Nondestructive analysis of alkali-silica reaction damage in concrete slabs using shear waves," *AIP Conf. Proc.*, vol. 1949, no. 1, p. 40003, 2018.

[15] K. J. Leśnicki, J.-Y. Kim, K. E. Kurtis, and L. J. Jacobs, "Characterization of ASR damage in concrete using nonlinear impact resonance acoustic spectroscopy technique," *NDT E Int.*, vol. 44, no. 8, pp. 721–727, 2011.

[16] J. Chen, A. Jayapalan, J. Kim, K. Kurtis, and L. Jacobs, "Rapid evaluation of alkali–silica reactivity of aggregates using a nonlinear resonance spectroscopy technique," *Cem. Concr. Res.*, vol. 40, no. 6, pp. 914–923, 2010.

[17] M. Rashidi, M.C. Knapp, A. Hashemi, J.Y. Kim, K.M. Donnell, R. Zoughi, L.J. Jacobs, K.E. Kurtis, "Detecting alkali-silica reaction: A multi-physics approach," *Cem. Concr. Compos.*, vol. 73, pp. 123–135, 2016.

[18] A. K. Singh, B.-Y. Chen, V. B. C. Tan, T.-E. Tay, and H.-P. Lee, "Finite element modeling of nonlinear acoustics/ultrasonics for the detection of closed delaminations in composites," *Ultrasonics*, vol. 74, pp. 89–98, 2017.

[19] D. M. Donskoy and A. M. Sutin, “Vibro-Acoustic Modulation Nondestructive Evaluation Technique,” *J. Intell. Mater. Syst. Struct.*, vol. 9, no. 9, pp. 765–771, 1998.
[20] D. Donskoy, A. Sutin, and A. Ekimov, “Nonlinear acoustic interaction on contact interfaces and its use for nondestructive testing,” *NDT E Int.*, vol. 34, no. 4, pp. 231–238, 2001.
[21] M. Sargolzahi, S. A. Kodjo, P. Rivard, and J. Rhazi, “Effectiveness of nondestructive testing for the evaluation of alkali–silica reaction in concrete,” *Constr. Build. Mater.*, vol. 24, no. 8, pp. 1398–1403, 2010.
[22] L. Pieczonka, L. Zietek, A. Klepka, W. J. Staszewski, F. Aymerich, and T. Uhl, “Damage imaging in composites using nonlinear vibro-acoustic wave modulations,” *Struct. Control Heal. Monit.*, vol. 25, no. 2, p. e2063, 2017.
[23] F. Moradi-Marani, S. A. Kodjo, P. Rivard, and C.-P. Lamarche, “Nonlinear Acoustic Technique of Time Shift for Evaluation of Alkali-Silica Reaction Damage in Concrete Structures,” *ACI Mater. J.*, vol. 111, no. 5, pp. 581–592, 2014.
[24] Kobayashi K. and N. and Banthia, “Corrosion detection in reinforced concrete using induction heating and infrared thermography,” *J. Civ. Struct. Heal. Monit.*, vol. 1, no. 1, pp. 25–35, Jun. 2011.
[25] Y. Bao and S. Mahadevan, “Uncertainty quantification of thermal image-based concrete diagnosis,” *Int. J. Sustain. Mater. Struct. Syst.*, vol. 2, no. 1/2, pp. 77–95, 2015.
[26] J. Chen, A. R. Jayapalan, J. Y. Kim, K. E. Kurtis, and L. J. Jacobs, “Nonlinear wave modulation spectroscopy method for ultra accelerated alkali-silica reaction assessment,” *ACI Mater. J.*, vol. 106, no. 6, pp. 340–348, 2009.
[27] M. Meo and G. Zumpano, “Nonlinear elastic wave spectroscopy identification of impact damage on a sandwich plate,” *Compos. Struct.*, vol. 71, p. 469, 2005.
[28] U. Polimeno, M. Meo, and D. P. Almond, “Smart Nonlinear Acoustic Based Structural Health Monitoring System,” in *Emboding Intelligence in Structures and Integrated Systems*, 2009, vol. 56, pp. 426–434.
[29] F. Aymerich and W. J. Staszewski, “Experimental study of impact-damage detection in composite laminates using a cross-modulation vibro-acoustic technique,” *Struct. Heal. Monit.*, vol. 9, p. 541, 2010.
[30] N. C. Yoder and D. E. Adams, “Vibro-acoustic modulation utilizing a swept probing signal for robust crack detection,” *Struct. Heal. Monit.*, vol. 9, p. 257, 2010.
[31] A. Klepka, L. Pieczonka, W. J. Staszewski, and F. Aymerich, “Impact damage detection in laminated composites by non-linear vibro-acoustic wave modulations,” *Compos. Part B Eng.*, vol. 65, p. 99, 2014.
[32] S. Kim, D.E. Adams, H. Sohn, G. Rodriguez-Rivera, N. Myrent, R. Bond, J. Vitek, S. Carr, A. Grama, J.J. Meyer, “Crack detection technique for operating wind turbine blades using Vibro-Acoustic Modulation,” *Struct. Heal. Monit.*, vol. 13, no. 6, pp. 660–670, 2014.
[33] B.-Y. Chen, S.-K. Soh, H.-P. Lee, T.-E. Tay, and V. B. C. Tan, “A vibro-acoustic modulation method for the detection of delamination and kissing bond in composites,” *J. Compos. Mater.*, vol. 50, no. 22, pp. 3089–3104, 2016.
[34] A. K. Singh, B. Chen, V. B. C. Tan, T. E. Tay, and H. P. Lee, “A theoretical and numerical study on the mechanics of vibro-acoustic modulation,” *J. Acoust. Soc. Am.*, vol. 141, no. 4, pp. 2821–2831, 2017.
[35] P. Karve and S. Mahadevan, “On the performance of vibro-acoustic-modulation-based diagnosis of breathing cracks in thick, elastic slabs,” *Struct. Control Heal. Monit.*, vol. 27, no. 3, 2020.

[36] Z. Tian, L. Huo, W. Gao, H. Li, and G. Song, “Modeling of the attenuation of stress waves in concrete based on the Rayleigh damping model using time-reversal and PZT transducers,” *Smart Mater. Struct.*, vol. 26, no. 10, p. 105030, 2017.
[37] P. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967.
[38] A. Nuttall, “Some windows with very good sidelobe behavior,” *IEEE Trans. Acoust.*, vol. 29, no. 1, pp. 84–91, 1981.
[39] V. Abaqus, “6.14 Documentation,” *Dassault Syst. Simulia Corp.*, vol. 651, pp. 2–6, 2014.
[40] M. Kawamura and K. Iwahori, “ASR gel composition and expansive pressure in mortars under restraint,” *Cem. Concr. Compos.*, vol. 26, no. 1, pp. 47–56, 2004.
[41] X. Hou, L. J. Struble, and R. J. Kirkpatrick, “Formation of ASR gel and the roles of C-S-H and portlandite,” *Cem. Concr. Res.*, vol. 34, no. 9, pp. 1683–1696, 2004.
[42] T. Knudsen, N. Thaulow, and M. Thomas, “Quantitative microanalyses of alkali-silica gel in concrete,” in *Cement and Concrete Research*, vol. 5, no. 5, J. P. Skalny, J. Gebauer, and I. Odler, Eds. The address of the publisher: American Ceramic Society, 1975, pp. 443–454.
[43] M. Thomas, “The role of calcium hydroxide in alkali recycling in concrete,” in *Materials Science of Concrete: Calcium Hydroxide in Concrete, Special Volume*, J. P. Skalny, J. Gebauer, and I. Odler, Eds. The address of the publisher: American Ceramic Society, 2001, pp. 225–236.
[44] S. Sankararaman and S. Mahadevan, “Bayesian methodology for diagnosis uncertainty quantification and health monitoring,” *Struct. Control Heal. Monit.*, vol. 20, no. 1, pp. 88–106, 2013.
[45] Y. Qian and A. Mita, “Acceleration-based damage indicators for building structures using neural network emulators,” *Struct. Control Heal. Monit.*, vol. 15, pp. 901–920, 2008.
[46] C. S. Byington, M. Watson, and D. Edwards, “Data-driven neural network methodology to remaining life predictions for aircraft actuator components,” in *IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720)*, 2004, vol. 6, pp. 3581----3589 Vol.6.
[47] S. W. Wegerich, “Similarity based modeling of time synchronous averaged vibration signals for machinery health monitoring,” in *IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720)*, 2004, vol. 6, pp. 3654----3662 Vol.6.
[48] P. Karve, S. Miele., K. Neal, S. Mahadevan, V. Agarwal, E.R. Giannini, and P. Kyslinger, “Vibro-acoustic modulation and data fusion for localizing alkali–silica reaction–induced damage in concrete,” *Struct. Heal. Monit.*, vol. 0, no. 0, p. 147592172090550, Feb. 2020.
[49] W. Z. Taffese and E. Sistonen, “Machine learning for durability and service-life assessment of reinforced concrete structures: Recent advances and future directions,” *Autom. Constr.*, vol. 77, pp. 1–14, 2017.
[50] Z. Wang and Y. Cha, “Automated damage-sensitive feature extraction using unsupervised convolutional neural networks,” in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018*, 2018, vol. 10598, p. 105981J.
[51] M. H. Rafiei and H. Adeli, “A novel machine learning-based algorithm to detect damage in high-rise building structures,” *Lat. Am. J. Solids Struct.*, vol. 26, no. 18, p. e1400, 2017.
[52] R. P. Finotti, A. A. Cury, and F. de S. Barbosa, “An SHM approach using machine learning and statistical indicators extracted from raw dynamic measurements,” *Struct. Des. tall Spec. Build.*, vol. 16, no. 2, 2019.

[53] B. Guldur and J. Hajjar, “Automated Classification of Detected Surface Damage from Point Clouds with Supervised Learning,” *Proc. 33rd ISARC*, pp. 307–313, 2016.
[54] S. Li and L. Sun, “Detectability of Bridge-Structural Damage Based on Fiber-Optic Sensing through Deep-Convolutional Neural Networks,” *J. Bridg. Eng.*, vol. 25, no. 4, p. 4020012, 2020.
[55] P. Castellini and G. M. Revel, “An experimental technique for structural diagnostic based on laser vibrometry and neural networks,” *Shock Vib.*, vol. 7, no. 6, pp. 381–397, 2000.
[56] C. V. Dung and L. D. Anh, “Autonomous concrete crack detection using deep fully convolutional neural network,” *Autom. Constr.*, vol. 99, pp. 52–58, 2019.
[57] O. R. de Lautour and P. Omenzetter, “Prediction of seismic-induced structural damage using artificial neural networks,” *Eng. Struct.*, vol. 31, no. 2, pp. 600–606, 2009.
[58] A. Athanasiou, A. Ebrahimkhanlou, J. Zaborac, T. Hrynyk, and S. Salamone, “A machine learning approach based on multifractal features for crack assessment of reinforced concrete shells,” *Comput. Civ. Infrastruct. Eng.*, vol. 35, no. 6, pp. 565–578, 2020.
[59] M. Li, Z. Wu, H. Yang, and H. Huang, “Direct damage index based on inverse finite element method for structural damage identification,” *Ocean Eng.*, vol. 221, p. 108545, 2021.
[60] S. Das, S. Dutta, C. Putcha, S. Majumdar, and D. Adak, “A data-driven physics-informed method for prognosis of infrastructure systems: Theory and application to crack prediction,” *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.*, vol. 6, no. 2, p. 4020013, 2020.
[61] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, and D. J. Inman, “Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks,” *J. Sound Vib.*, vol. 424, pp. 158–172, 2018.
[62] A. Rai and M. Mitra, “A hybrid physics-assisted machine-learning-based damage detection using Lamb wave,” *Sādhanā*, vol. 46, no. 2, pp. 1–11, 2021.
[63] F.-G. Yuan, S. A. Zargar, Q. Chen, and S. Wang, “Machine learning for structural health monitoring: challenges and opportunities,” in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2020*, 2020, vol. 11379, p. 1137903.
[64] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.
[65] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, and P. Perdikaris, “Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data,” *J. Comput. Phys.*, vol. 394, pp. 56–81, 2019.
[66] B. Kapusuzoglu and S. Mahadevan, “Physics-informed and hybrid machine learning in additive manufacturing: Application to fused filament fabrication,” *JOM*, vol. 72, no. 12, pp. 4695–4705, 2020.
[67] D. Liu and Y. Wang, “Multi-fidelity physics-constrained neural network and its application in materials modeling,” *J. Mech. Des.*, vol. 141, no. 12, 2019.
[68] C. Zhang and A. Shafieezadeh, “Simulation-free reliability analysis with active learning and Physics-Informed Neural Network,” *Reliab. Eng. Syst. Saf.*, vol. 226, p. 108716, 2022.
[69] E. Figueiredo, I. Moldovan, A. Santos, P. Campos, and J. C. W. A. Costa, “Finite element--based machine-learning approach to detect damage in bridges under operational and environmental variations,” *J. Bridg. Eng.*, vol. 24, no. 7, p. 4019061, 2019.
[70] P. Seventekidis, D. Giagopoulos, A. Arailopoulos, and O. Markogiannaki, “Structural Health Monitoring using deep learning with optimal finite element model generated data,” *Mech. Syst. Signal Process.*, vol. 145, p. 106972, 2020.

[71] A. S. Lapedes and R. M. Farber, “How neural nets work,” in *Neural information processing systems*, 1988, pp. 442–456.
[72] ASTM E1316-13c, “The estimate for deterioration due to alkali aggregate reaction by ultrasonic methods,” *Struct. Heal. Monit.*, vol. 9, no. 2, pp. 1–28, Nov. 2014.
[73] D. Broda, W. J. Staszewski, A. Martowicz, T. Uhl, and V. V Silberschmidt, “Modelling of nonlinear crack--wave interactions for damage detection based on ultrasound—A review,” *J. Sound Vib.*, vol. 333, no. 4, pp. 1097–1118, 2014.
[74] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.
[75] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
[76] K. Levenberg, “A method for the solution of certain non-linear problems in least squares,” *Q. Appl. Math.*, vol. 2, no. 2, pp. 164–168, 1944.
[77] D. W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.
[78] L. Prechelt, “Early stopping-but when?,” in *Neural Networks: Tricks of the trade*, Springer, 1998, pp. 55–69.
[79] Z. Li, D. Tao, M. Li, Z. Shu, S. Jing, M. He, P. Qi, “Prediction of damage accumulation effect of wood structural members under long-term service: A machine learning approach,” *Materials (Basel).*, vol. 12, no. 8, p. 1243, 2019.
[80] M. J. D. Powell, “Restart procedures for the conjugate gradient method,” *Math. Program.*, vol. 12, no. 1, pp. 241–254, 1977.
[81] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
[82] S. Pawar, O. San, P. Vedula, A. Rasheed, and T. Kvamsdal, “Multi-fidelity information fusion with concatenated neural networks,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–13, 2022.
[83] X. Zhang, F. Xie, T. Ji, Z. Zhu, and Y. Zheng, “Multi-fidelity deep neural network surrogate model for aerodynamic shape optimization,” *Comput. Methods Appl. Mech. Eng.*, vol. 373, p. 113485, 2021.
[84] P. Nath, M. Sato, P. Karve, and S. Mahadevan, “Multi-fidelity Modeling for Uncertainty Quantification in Laser Powder Bed Fusion Additive Manufacturing,” *Integr. Mater. Manuf. Innov.*, pp. 1–20, 2022.
[85] G. N. Absi and S. Mahadevan, “Multi-fidelity approach to dynamics model calibration,” *Mech. Syst. Signal Process.*, vol. 68, pp. 189–206, 2016.
[86] M. G. Fernández-Godino, C. Park, N.-H. Kim, and R. T. Haftka, “Review of multi-fidelity models,” *arXiv Prepr. arXiv1609.07196*, 2016.
[87] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis, “Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling,” *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 473, no. 2198, p. 20160751, 2017.
[88] K. Zhou and J. Tang, “Efficient characterization of dynamic response variation using multi-fidelity data fusion through composite neural network,” *Eng. Struct.*, vol. 232, p. 111878, 2021.
[89] S. Miele, P. M. Karve, S. Mahadevan, and V. Agarwal, “Diagnosis of internal cracks in concrete using vibro-acoustic modulation and machine learning,” *Struct. Heal. Monit.*, p. 14759217211047900, 2022.

[90] H. Adeli and X. Jiang, *Intelligent infrastructure: neural networks, wavelets, and chaos theory for intelligent transportation systems and smart structures*. CRC press, 2008.
[91] A. De Luca, D. Perfetto, A. De Fenza, G. Petrone, and F. Caputo, "Guided wave SHM system for damage detection in complex composite structure," *Theor. Appl. Fract. Mech.*, vol. 105, p. 102408, 2020.
[92] J. P. Amezquita-Sanchez and H. Adeli, "Signal processing techniques for vibration-based health monitoring of smart structures," *Arch. Comput. Methods Eng.*, vol. 23, no. 1, pp. 1–15, 2016.
[93] Y. Bao and S. Mahadevan, "Diagnosis of interior damage with a convolutional neural network using simulation and measurement data," *Struct. Heal. Monit.*, p. 14759217211056574, 2021.
[94] X. Wang, C. Shen, M. Xia, D. Wang, J. Zhu, and Z. Zhu, "Multi-scale deep intra-class transfer learning for bearing fault diagnosis," *Reliab. Eng. Syst. Saf.*, vol. 202, p. 107050, 2020.
[95] D. Chakraborty, N. Kovvali, B. Chakraborty, A. Papandreou-Suppappola, and A. Chattopadhyay, "Structural damage detection with insufficient data using transfer learning techniques," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2011*, 2011, vol. 7981, p. 798147.
[96] M. Torzoni, A. Manzoni, and S. Mariani, "Health monitoring of civil structures: A MCMC approach based on a multi-fidelity deep neural network surrogate," in *Presented at the 1st Online Conference on Algorithms*, 2021.
[97] G. N. Absi and S. Mahadevan, "Simulation and Sensor Optimization for Multifidelity Dynamics Model Calibration," *AIAA J.*, vol. 58, no. 2, pp. 879–888, 2020.
[98] W. Sisson, S. Mahadevan, and B. P. Smarslok, "Optimization of Information Gain in Multifidelity High-Speed Pressure Predictions," *AIAA J.*, vol. 59, no. 8, pp. 3096–3105, 2021.
[99] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2009.
[100] V. Pandiyan *et al.*, "Deep transfer learning of additive manufacturing mechanisms across materials in metal-based laser powder bed fusion process," *J. Mater. Process. Technol.*, vol. 303, p. 117531, 2022.
[101] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 692–706, 2019.
[102] P. Karve, "Digital twin approach for damage-tolerant mission planning under uncertainty," *Eng. Fract. Mech.*, 2020.
[103] L. F. M. Sanchez, T. Drimalas, and B. Fournier, "Assessing condition of concrete affected by internal swelling reactions (ISR) through the Damage Rating Index (DRI)," *Cement*, vol. 1, p. 100001, 2020.
[104] H. Song, S. B. Feldman, and J. S. Popovics, "In situ detection and characterization of alkali-silica reaction damage in concrete using contactless ultrasonic wavefield imaging," *Cem. Concr. Compos.*, vol. 133, p. 104661, 2022.
[105] N. D. B. Ezell, H. Sun, S. Venkatakrishnan, and S. Sabatino, "Ultrasonic Model Based Iterative Reconstruction of Experimental Concrete Specimens at EPRI," *ORNL*, p. 2318, 2022.

[106] N. D. B. Ezell, A. Albright, D. Floyd, D. Clayton, and L. Khazanovich, “A novel use of frequency-banded synthetic aperture focusing technique for reconstructions of alkali-silica reaction in thick-reinforced concrete structures,” in *AIP Conference Proceedings*, 2019, vol. 2102, no. 1, p. 110003.
[107] S. Miele, P. Karve, and S. Mahadevan, “Multi-Fidelity Physics-Informed Machine Learning for Probabilistic Damage Diagnosis,” *Reliab. Eng. Syst. Saf.*, vol. Under Review, 2022.
[108] P. M. Karve and S. Mahadevan, “Bayesian information fusion for fatigue crack growth diagnosis using ultrasonic guided wave pitch-catch in a piezoelectric actuator-sensor network,” in *Structural Health Monitoring 2019: Enabling Intelligent Life-Cycle Health Management for Industry Internet of Things (IIOT) - Proceedings of the 12th International Workshop on Structural Health Monitoring*, 2019, vol. 2.
[109] P. E. Grattan-Bellew and A. Danay, “Comparison of laboratory and field evaluation of AAR in large dams,” in *Proceedings of the International Conference on Concrete Alkali--Aggregate Reactions in Hydroelectric Plants and Dams, Fredericton, NB*, 1992, vol. 28.
[110] ASTM C856-02(2003), “Standard Practice for Petrographic Examination of Hardened Concrete,” 2003.
[111] B. Fournier, M.-A. Bérubé, K. J. Folliard, M. Thomas, and others, “Report on the diagnosis, prognosis, and mitigation of Alkali-Silica Reaction (ASR) in transportation structures,” 2010.