INL/EXT—20-59914
Revision 0

Light Water Reactor Sustainability Program

Concrete Structure Health Monitoring Using Vibro-acoustic Testing and Machine Learning

September 2020

U.S. Department of Energy
Office of Nuclear Energy

INL/EXT-XX-XXXXX

Light Water Reactor Sustainability Program

Concrete Structure Health Monitoring Using Vibro-acoustic Testing and Machine Learning

Sarah Miele, Pranav Karve, Sankaran Mahadevan,
Vanderbilt University

Vivek Agarwal
Idaho National Laboratory

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

ABSTRACT

Assessment and management of aging concrete structures in nuclear power plants require a systematic approach rather than simple reliance on margins of safety in existing codes. Research is required to deploy online monitoring to assess the health of concrete structures in nuclear power plants and perform predictive analytics based on heterogeneous measurements, in order to reduce the operation and maintenance costs.

The research effort described in this report focuses on the health monitoring of concrete slabs containing reactive aggregates using the vibro-acoustic modulation (VAM) technique and predictive modeling. The research is developing an approach to enable detection, localization, and estimation of the extent of degradation in concrete slabs due to alkali-silica reaction (ASR). The project outcomes are generic and extendable to other degradation modes in concrete. A concrete slab without reinforcements was cast at Vanderbilt University and cured in aggressive conditions to accelerate ASR degradation.

In this report, we develop a methodology for the automation of VAM testing and the construction of machine learning models to facilitate damage localization in concrete structures. We investigate how machine learning models based on physics simulation data can be effectively utilized for damage localization (damage location and depth) and estimation in a concrete-slab specimen. Our study indicates that these models could potentially support damage localization and estimation for NPP structures.

EXECUTIVE SUMMARY

A challenge facing the current fleet of light water reactors in the United States is the age-related degradation of their passive assets, including concrete, cables, piping, and the reactor-pressure vessel. Since reactors within the current fleet of nuclear power plants (NPPs) have continued to operate for 60 years or more, it is important to understand the current and future conditions of passive assets under different operating conditions to support operational and maintenance decisions. To ensure safe and reliable long-term operation of the current fleet, the U.S. Department of Energy's Office of Nuclear Energy funds the Light Water Reactor Sustainability Program (LWRS) to develop the scientific basis for extending the operation of commercial light water reactors beyond the current license extension period.

Concrete structures—one among many passive assets of interest—are investigated in this research project. Reinforced-concrete structures found in NPPs can be grouped into the following categories: (1) primary containment, (2) containment internal structures, (3) secondary containments/reactor buildings, and (4) spent-fuel pool and cooling towers. These concrete structures are influenced by a variety of degradation mechanisms that are related to chemical, physical, and mechanical causes, and by irradiation.

The structural health of concrete must be monitored to assess the current condition of the structure, and to provide high-confidence, actionable information regarding structural integrity and reliability. Vanderbilt University, in collaboration with Idaho National Laboratory (INL) and Oak Ridge National Laboratory (ORNL), has investigated a probabilistic framework for structural health monitoring and managing the condition of aging concrete structures in NPPs. This integrated framework includes four elements: (1) monitoring, (2) data analytics, (3) uncertainty quantification, and (4) prognosis.

This report focuses on concrete degradation caused by alkali-silica reaction (ASR). The research team prepared concrete specimens to study accelerated ASR degradation in a laboratory setting. Non-destructive examination (NDE) techniques, which include thermography, mechanical-deformation measurements, nonlinear impact resonance-acoustic spectroscopy, and vibroacoustic modulation (VAM), were studied to detect the damage caused by ASR on concrete slabs cured at Vanderbilt University and were documented in an earlier report. In particular, damage localization using vibro-acoustic modulation was investigated, and the effects of different characteristics of the dual-frequency vibration tests (such as excitation frequencies, amplitudes, and locations) on the damage localization results were examined. This report discusses the automation of the VAM testing procedure to produce damage maps that indicate likely areas of damage.

The performance of physics-informed, machine learning (ML) models for ASR damage localization is investigated in this report. It is difficult to develop supervised learning models using only experimental data since the manufacturing of slabs with known hidden cracks are expensive and sufficient experimental training data is typically unavailable. 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. The ML-based diagnostic models are trained using wave physics simulations for two-dimensional (2D) domains (under plane strain conditions) containing hidden cracks. The performance of simulation-data-driven ML models is evaluated using both synthetic

2D simulation data and experimentally obtained data from a plain cement slab specimen (seeded with ASR damage). The report evaluates and compares the utility of prediction as well as classification models for damage localization. The ML models constructed using physics simulation data show encouraging results in the validation tests performed using laboratory data on a concrete-slab specimen. The proposed damage detection and localization methodology has the potential to support the continuous assessment of concrete performance and enhance the prognostics and health management framework of NPP concrete structures.

ACKNOWLEDGMENTS

This report was made possible through funding by the U.S. Department of Energy's Light Water Reactor Sustainability Program. We are grateful to Alison Hahn of the U.S. Department of Energy and Bruce Hallbert and Craig A. Primer at Idaho National Laboratory for championing this effort. We are also thankful to David Koester, Garrett Thorne, and Rich Tiesing at Vanderbilt University for their help with casting, coring, and testing the concrete specimens. We appreciate the contribution of Katie Stokes at Idaho National Laboratory for technical editing and formatting of the report.

CONTENTS

ABSTRACT iii

EXECUTIVE SUMMARY v

ACKNOWLEDGMENTS vii

ACRONYMS xiv

1. INTRODUCTION 1

2. TECHNICAL BACKGROUND 2
   - 2.1 Alkali-Silica Reaction Damage Assessment 2
   - 2.2 Plain Concrete Slab Specimen 2
   - 2.3 Vibro-acoustic Modulation (VAM) Testing 4
     - 2.3.1 Description of VAM 4
     - 2.3.2 VAM test procedure 4
     - 2.3.3 Automation of the VAM test procedure 5
   - 2.4 ASR damage localization for plain concrete specimen 6
     - 2.4.1 VAM tests 6
     - 2.4.2 Petrographic examination 7
     - 2.4.3 Visual inspection of the broken-down specimen 8

3. MACHINE LEARNING FOR DAMAGE LOCALIZATION 10
   - 3.1 Training Data Generation- Finite Element Analysis 11
   - 3.2 Damage Diagnosis Methodology 14
     - 3.2.1 Damage localization using the prediction model 14
     - 3.2.2 Damage localization using classification models 15

4. RESULTS AND DISCUSSION 19
   - 4.1 Validation using Numerical Experiments 19
     - 4.1.1 Damage location identification using the prediction model 20
     - 4.1.2 Damage location identification using classification and prediction models 22
   - 4.2 Validation using laboratory experiments 27

5. CONCLUSIONS AND FUTURE WORK 31

6. REFERENCES 32

FIGURES

Figure 1. The plain concrete slab specimen (dimensions 2 ft × 2 ft × 6 in.) 3

Figure 2. Pockets of aggregate in the slab during casting with red squares identifying visually observed crack and effluence locations on the side of the slab. 3

Figure 3. LS Plot for an accelerometer, labeled to show the values used in calculating SBSum. SBSum = AmpS1 + AmpS2. 5
Figure 4: Display of the automated testing process: test status and results 6
Figure 5: Pump, probe, and accelerometer locations for both halves of the slab (labeled in gray) 7
Figure 6: Damage map directly based on VAM data (model-free) [11] The red region highlights locations of sensors showing high average SBSum values, indicating potential ASR-damage locations 7
Figure 7: Locations of 4 cores taken from the plain concrete slab 8
Figure 8: Shifting of aggregates and the presence of small hidden cracks in each quadrant of the plain concrete slab. a.) The shift of Maine aggregates in quadrant 1. b.) The amorphous silica in quadrant 2. c.) The spread of Ontario aggregates in quadrant 3. d.) The shift of New Mexico aggregates in quadrant 4. 9
Figure 9: Results of post destruction visual inspection of the plain concrete slab showed a shift in the location the aggregates were placed before curing. a) Aggregate locations during casting. b) Estimated aggregate locations (as observed in the broken-down specimen). 10
Figure 10: Example of an artificial neural network. 11
Figure 11: The domain geometry, and pump and probe locations used in the numerical simulations 13
Figure 12: The domain geometry and crack locations used in the numerical simulations. Only one crack is present in the specimen domain for a given test 13
Figure 13: Schematic of ANN structure with six inputs, one output, and two hidden layers of five neurons each. 14
Figure 14: Sensors located in damaged and undamaged areas (sensors in region N are located in the damaged zone, sensors in regions A and B are not located in the damaged zone) 16
Figure 15: Classification Model A 17
Figure 16: Classification Model B 17
Figure 17: Classification Model C 18
Figure 18: Classification Model D 18
Figure 19: Classification Model E 19
Figure 20: Crack locations for validation using numerical experiments (cracks 5, and 21) 20
Figure 21: Bayesian estimation of damage for the validation crack 5 (197 sensors) 21
Figure 22: Damage probability for crack 5 using Bayesian damage diagnosis 21
Figure 23: Bayesian estimation of damage for the external validation crack 5 (29 sensors) 21
Figure 24: Bayesian estimation of damage for the external validation crack 5 (15 sensors) 22
Figure 25: Bayesian estimation of damage for the external validation crack 5 (all pump and probe actuator locations, 15 sensors) 22
Figure 26: Validation of Model A on cracks 5 and 21. 23
Figure 27: Validation of Model B on cracks 5 and 21. 23
Figure 28: Validation of Model C on cracks 5 and 21. 24

Figure 29: Validation of Model D on cracks 5, and 21. 24
Figure 30: Validation of Model E on cracks 5 and 21. 25
Figure 31: Validation of Model E on cracks 5 and 21 for 15 sensor locations along the surfaces of the specimens and posterior for the depth calculated using the estimated crack centroids and lengths for each validation specimen....26
Figure 32: Experimental setup for Vanderbilt ASR induced specimen....28
Figure 33: Slice locations for experimental analysis....28
Figure 34: Bayesian estimation of damage for slice 6 using a frequency ratio of 20, all amplitude ratios, and all pump and probe actuator locations at 7 sensors 29
Figure 35: Damage probability for the plane concrete specimen using the prediction ANN model and Bayesian estimation....29
Figure 36: Average damage classification for each slice of the plain concrete specimen 30

TABLES

Table 1. Material properties used in the 2D FEA model. 12
Table 2: Test parameters for the 2D FEA model 12
Table 3: Different ANN models to be evaluated 16
Table 4: Test parameters for the 2D FEA models used for generating validation test data 20
Table 5: Validation for multiple models using Crack 5 27
Table 6: Validation for multiple models using Crack 21 27

ACRONYMS

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

Concrete Structural Health Monitoring Using Vibro-acoustic Testing and Machine Learning

1. INTRODUCTION

The majority of existing nuclear power plants (NPPs) continue to operate beyond their initial license periods. As they continue to operate, the passive structures, systems, and components of NPPs suffer deterioration that influences their 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.

This project focuses on concrete structures in NPPs. These are grouped into the following categories: (1) primary containment, (2) containment internal structures, (3) secondary containment/reactor buildings, and (4) other structures, such as used fuel pools, dry storage casks, and cooling towers. These concrete structures are affected by a variety of chemical, physical, and mechanical degradation mechanisms, such as alkali-silica reaction (ASR), chloride penetration, sulfate attack, carbonation, freeze-thaw cycles, shrinkage, and mechanical loading [1]. The age-related deterioration of concrete results in continuous microstructural changes (e.g., slow hydration, crystallization of amorphous constituents, and reactions between cement paste and aggregates). It is important that changes over long periods be measured and monitored, and that their impacts on the integrity of components are analyzed to best support long-term operations and maintenance decisions for the existing fleet of nuclear reactors.

Structural health monitoring (SHM) can produce actionable information regarding structural integrity that, when conveyed to the decision-maker, enables risk management concerning structural integrity and performance. The SHM methods and technologies include measurements, monitoring, and analysis of aging structures under different operating conditions. In addition to data from the specific system being monitored, information may also be available for similar or nominally identical systems in an operational NPP fleet, as well as legacy systems. Therefore, to take advantage of this valuable information, Christensen [2] suggested that assessment and management of aging concrete structures in NPPs require a more systematic and dynamic approach than simple reliance on existing code margins of safety.

Through the Light Water Reactor Sustainability Program, national laboratories (Idaho National Laboratory and Oak Ridge National Laboratory) and universities (Vanderbilt University, University of Nebraska–Lincoln, University of Alabama, University of Tennessee, University of South Carolina, and Georgia Tech University) have been investigating concrete SHM techniques. The ultimate goal of this research is to enable plant operators to make risk-informed decisions on structural integrity, remaining useful life, and performance of concrete structures across the NPP fleet. The project's long-term research objective is to develop methods and tools that produce actionable information regarding structural integrity, individualized for structures of interest, and their performance goals. Additionally, the project supports the research objectives of three pathways under the Light Water Reactor Sustainability Program (i.e., the Plant Modernization Pathway, the Materials Research Pathway, and the Risk-Informed System Analysis Pathway).

A Prognostics and Health Management (PHM) framework was proposed in [3] to investigate concrete structure degradation by integrating the following technical elements: (1) health condition monitoring, (2) data analytics, (3) uncertainty quantification, and (4) prognosis. The framework will help plant operators make risk-informed decisions on structural integrity, remaining useful life, and concrete structure performance. In this context, Vanderbilt University has been studying various techniques to assess ASR degradation in controlled concrete specimens, as reported in [4], [5], [6], [7].

The objective of the current report is to examine the use of machine learning (ML) techniques for vibration-based non-destructive evaluation (NDE) in informing the PHM framework. The ML models

were trained based on simulation data generated by finite-element analysis. In addition, experiments were conducted at Vanderbilt University to provide concrete degradation data that were used to evaluate the performance of the ML models. The technical background, experimental setting, data processing, model building, significant results, technical findings, and conclusions are included in the remainder of the report and are organized as follows:

- Section 2 discusses the technical basics of ASR development and the dual-frequency vibration-based technique used to assess the effects of ASR on the integrity of concrete samples. Further, it discusses an automation procedure developed for VAM testing that produces damage maps indicating areas of likely damage.
- Section 3 discusses the proposed ML-based diagnostic methodology. It explores the utility of predictive and classification models, built using the training data from physics simulations, for ASR damage localization.
- Section 4 examines the results of numerical and laboratory experiments utilizing ML-based damage diagnosis.
- Section 5 summarizes the research and discusses future needs.

2. TECHNICAL BACKGROUND

2.1 Alkali-Silica Reaction Damage Assessment

The 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 is called ASR. Given sufficient moisture, 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 the 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 [8], [9]. In the presence of water, the ASR gel increases in volume and exerts an expansive pressure inside the material, causing micro- to macrocracks. As a result, ASR reduces the stiffness and tensile strength of concrete because these properties are particularly sensitive to microcracking.

Our past work considered data-driven, model-free and physics-model-based NDE methods for localization of ASR-induced damage in concrete (diffused micro- and macrocracking) [10], [11]. In this report, we explore the utilization of physics-informed ML models to perform damage localization. We first provide a brief overview of the experimental program conducted for casting a plain cement slab (to be used for validation of the proposed methodology), the salient features of the nonlinear dynamics-based NDE methodology (vibroacoustic modulation, or VAM), and the results of petrographic investigations as well as VAM tests. The performance of the proposed ML approach will be evaluated using petrographic investigations and VAM test results.

2.2 Plain Concrete-Slab Specimen

VAM-based damage localization was studied using a plain concrete specimen with aggregate pockets, constructed and cured in the laboratory. A concrete slab with dimensions of 2 ft × 2 ft × 6 in. was cast on December 21, 2015, and cured at Vanderbilt University. The details of the casting and curing process of this specimen are discussed in detail in a report by Mahadevan et al. [4]. Figure 1 shows an image of the slab immediately after the mold was removed.

Figure 1. The plain concrete-slab specimen (dimensions 2 ft × 2 ft × 6 in.).

Four types of aggregate were placed in pockets at a depth of 3 in. in the four quadrants of the slab (Figure 2). The aggregates were placed in pockets instead of being dispersed throughout the slab so that the reactivity of each aggregate can be determined independently. Additionally, since the locations of the pockets of aggregate are known, this information was used to validate the localization of ASR from vibration-based tests. The four types of aggregates used were:

1. Pure silica powder from local ceramic shop
2. Coarse aggregate from Maine, donated by the University of Alabama
3. Coarse aggregate from New Mexico, donated by the University of Alabama
4. Coarse aggregate from a quarry in Ontario, Canada, donated by the Ontario Ministry of Transportation.

In October 2016, researchers observed the first visual evidence of degradation due to ASR. Since then, degradation-related damage became increasingly pronounced. Hairline cracks were observed on the surface of the slab, and later, an ASR-gel effluent and whitish powder seeped out of the slab. In Figure 2, the red squares identify the locations where seepage of ASR-gel effluent was observed on the sides of the slab.

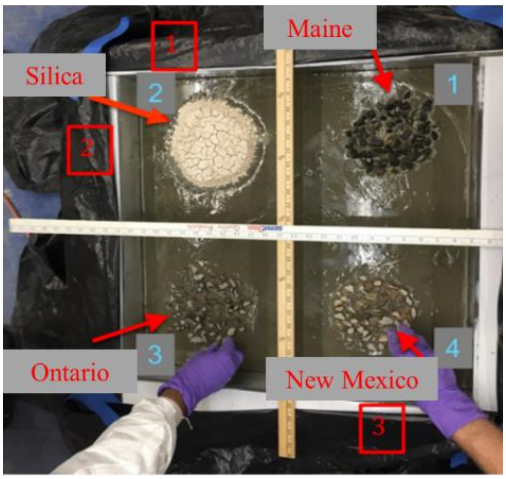


Figure 2. Pockets of aggregate in the slab during casting with red squares identifying visually observed crack and effluence locations on the side of the slab.

2.3 Vibro-acoustic Modulation (VAM) Testing

2.3.1 Description of VAM

Vibroacoustic modulation (VAM), also known as nonlinear wave modulation spectroscopy, is a non-destructive evaluation (NDE) technique that relies on detecting the dynamic signature of nonlinear structural behavior as the primary indicator of damage. Specifically, VAM aims at the detection of modulation of a higher excitation frequency by a lower excitation frequency caused by delamination or cracks in structural components. The utility of VAM for detecting debonding flaws and cracks in composites, metals, as well as ASR-induced cracks in concrete has been demonstrated in the past [12], [13].

In the VAM technique, the structural component of interest is excited simultaneously using a combination of two signals of specific frequencies, and 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" [14]. The geometric or material nonlinearity in the form of variable contact area or nonlinear adhesive bond at the surfaces of cracks or delamination causes modulation of the probing frequency (f_{pr}) by the pumping frequency (f_{pu}). This modulation, and hence, the presence of the flaw is seen in the frequency spectra of the measured response as peaks of higher magnitude (sidebands) around the probe frequency. The effect of nonlinearities is pronounced near the location of the flaw; thus, the relative magnitude of a sidebands-based damage index enables localization of the flaw [15]. VAM tests were conducted on the plain concrete-slab specimen (Section 2.2) to perform ASR damage localization. Details of the test procedure are described next.

2.3.2 VAM test procedure

In VAM tests, the pump and probe excitations are delivered using piezo-stack actuators. The locations of these actuators, as well as the frequencies at which they operate (i.e., the pump and probe frequencies) are varied. The response of the structural component of interest is measured using a finite number of accelerometers placed on the surface of the component. 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 performance of the VAM test 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_{pu}.

In the literature it has been reported that the probing frequency, f_{pr}, should be at least 10 to 20 times the pumping frequency [16]. When the f_{pr} 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. 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. A maximum of 21 accelerometers at a time were 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.

In laboratory experiments, the slab was excited using the pump and probe frequencies. The nonlinearity in the structural response, or the modulation of the f_{pr} by the f_{pu}, is seen as sidebands around the f_{pr} ("peaks" in the linear spectrum, or linear spectrum [LS]) at frequencies equal to ($f_{pr} \pm$

f_{pu}). After computing the LS, the values of AmpS1 and AmpS2 (as shown in Figure 3) were used to calculate the sideband sums (denoted as SBSum). In this report, we discuss the mapping of ASR-induced cracks based on the SBSum metric, SBSum = AmpS1 + AmpS2 (see Figure 3). Refer to [11] for a detailed description of the VAM test, data processing, and damage localization procedures.

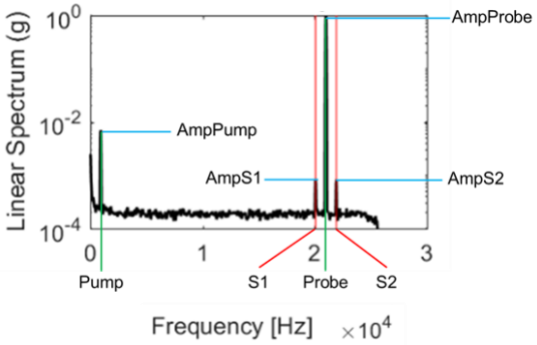


Figure 3. LS Plot for an accelerometer, labeled to show the values used in calculating SBSum. SBSum = AmpS1 + AmpS2.

2.3.3 Automation of the VAM test procedure

The optimal VAM test parameters (probe frequency, pump/probe amplitudes, pump/probe actuator locations) depend on the location and size of the damage. Since this is typically unknown at the time of test performance, the proposed methodology employs a multiconfiguration 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 multiconfiguration testing approach, the research team made efforts to automate the VAM testing process and minimize human intervention during testing. In the automated test procedure, the testing engineer's involvement is needed to install accelerometers and actuators on the specimen. This led to the development of 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). Thus, after the physical setup of testing equipment, 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.

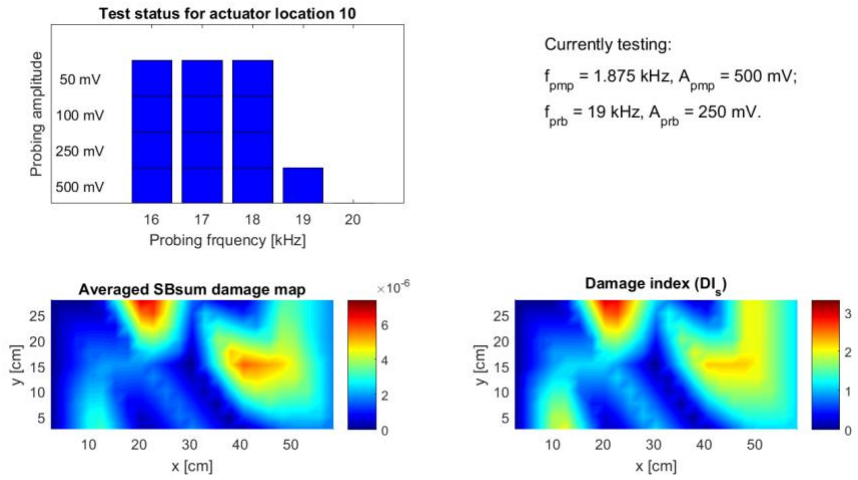


Figure 4: Display of the automated testing process: test status and results.

Note that this program can be modified to develop a fully automated testing software. This will be especially beneficial if a non-contact sensing system (such as laser Doppler velocimetry) could be used to measure the nonlinear dynamic response of the structure of interest. In this case, test engineer intervention may only be required to install the pump/probe actuator at a few locations on the testing surface.

2.4 ASR Damage Localization for Plain Concrete Specimen

In this section, we discuss the result of damage localization tests performed on the plain concrete specimen. A series of non-destructive (VAM) and destructive tests were performed. First, VAM tests were conducted, and the corresponding damage maps were obtained. Next, four cores were taken out of the specimen and examined using petrographic analysis to confirm the presence of ASR and microcracking. Finally, the slab was broken down to find out the locations of reactive aggregate and macrocracks. The results of the ML-based damage localization approach were validated using previously reported data-driven, model-free approach results as well as destructive test results.

2.4.1 VAM tests

VAM tests were conducted on the plain concrete specimen as per the procedure described in Section 2.3.2. Accelerometer locations and pump and probe configurations for the specimen are displayed in Figure 5. Figure 6 shows the average SBSum plot obtained for the Vanderbilt University sample by averaging the values of SBSum at each sensor over all five different pump and probe locations with probing amplitudes of 500, 250, 100, and 50 mV, probing frequencies of 10–21 kHz, and a pumping frequency of 920 Hz with 500 mV amplitude. These results show the highest SBSum values in quadrants 3 and 4, with the large crack between quadrants 1 and 4 also being highlighted. The damage maps were obtained using a model-free, data-driven approach. That is, a model of the governing physics for the structural system of interest (nonlinear wave physics for the concrete slab) was not used for obtaining the mean SBSum map. The details of the damage map computation procedure can be found in [11].

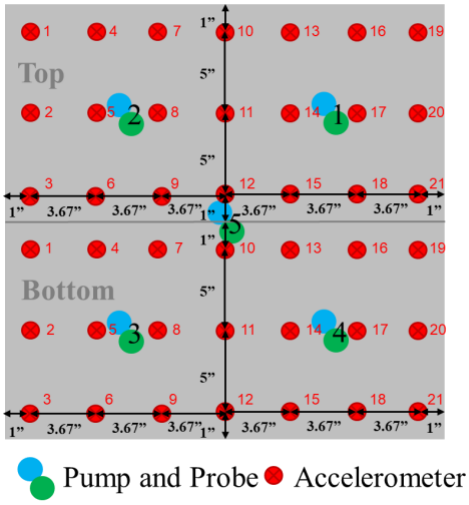


Figure 5: Pump, probe, and accelerometer locations for both halves of the slab (labeled in gray).

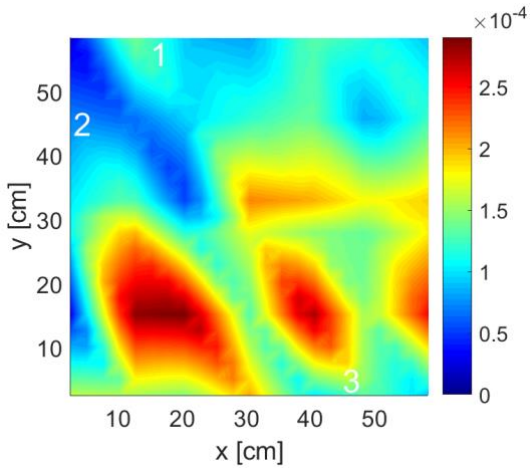


Figure 6: Damage map directly based on VAM data (model-free) [11]. The red region highlights locations of sensors showing high average SBSum values, indicating potential ASR damage locations.

2.4.2 Petrographic examination

Four 4-inch diameter cores were taken from the center of each of the four quadrants ([11] Figure 7) in the plain concrete-slab specimen, and petrography testing was performed at R.J. Lee Group. Cores taken through the amorphous silica and Maine aggregate 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, implying they had shifted elsewhere in the specimen. Trace evidence of ASR gel was observed lining a void on a fracture surface of the Maine core. Cores were taken in the New Mexico and Ontario quadrants, and both show the presence of ASR and ASR-induced cracking. The petrographic examination thus confirmed the presence of ASR-induced microcracking in quadrants 3 and 4. A detailed account of the petrographic study is found in [11].

Figure 7: Locations of four cores taken from the plain concrete slab.

2.4.3 Visual inspection of the broken-down specimen

After completing the petrographic examination, the specimen was broken down to observe the locations of internal cracks and aggregate. It was observed that the aggregates shifted to the outer corners of the slab (likely during the pouring of cement slurry). This shift explained why the core taken in quadrant 1 showed none of the Maine aggregates. These aggregates shifted to the top and slightly to the right. Damage localization in quadrant 2 is difficult using VAM because the opening created by the silica layer was too large to open and close due to the applied dynamic excitation. The Ontario and New Mexico aggregate also shifted slightly to the outer corners, but the cores were able to maintain most of the aggregates. Figure 9 depicts the approximate locations of the aggregates in each quadrant pre- and post-curing. The observed locations of shifted aggregate in quadrant 1 and small, crack-like voids in quadrant 2 indicated that small hidden cracks might be present at these locations.

a.

b.

c.

d.

Figure 8: Shifting of aggregates and the presence of small hidden cracks in each quadrant of the plain concrete slab. (a) The shift of Maine aggregates in quadrant 1. (b) The amorphous silica in quadrant 2. (c) The spread of Ontario aggregates in quadrant 3. (d) The shift of New Mexico aggregates in quadrant 4.

a. b.

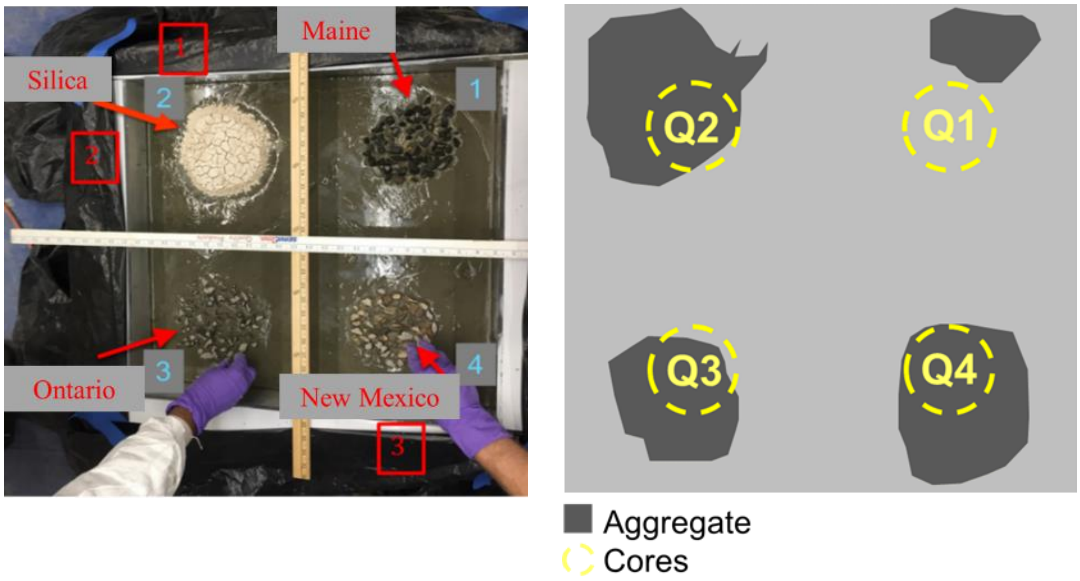


Figure 9: Results of post-destruction visual inspection of the plain concrete slab showed a shift in the location the aggregates were placed before curing. (a) Aggregate locations during casting. (b) Estimated aggregate locations (as observed in the broken-down specimen).

3. MACHINE LEARNING FOR DAMAGE LOCALIZATION

Non-destructive evaluation methodologies typically rely on a model of the structural system of interest to support inference in damage diagnosis. These models can be physics-based, data-driven, or hybrid models. Previously, we explored the physics-based (finite element) modeling approach as well as a model-free, data-driven approach for VAM-based damage localization. The computational cost involved in numerical simulation of the nonlinear phenomenon of interest, for heterogeneous, 3D domains, could be prohibitive in real-world applications. In previous work, the damage index threshold selection for the data-driven approach was made based on expert/analyst elicitation. Hence, in this work, we investigate whether simulation data-driven models can be effectively utilized to reduce the computational cost of damage diagnosis and eliminate the need for analyst-defined damage index threshold. It is difficult to develop a data-driven model using only laboratory or real-world test data because this data is typically unavailable and/or expensive to obtain. To overcome this challenge, we investigate the utility of low-fidelity physics models to train supervised ML models. Specifically, we opt for a finite-element model of a 2D elastic domain containing hidden cracks, since nonlinear dynamics simulations for 3D domains are computationally expensive. 2D finite-element models allow us to generate sufficient training data.

We use a popular data-driven model called an artificial neural network (ANN). ANN uses a series of connected layers with multiple nodes and neurons intended to represent synapse connections in the brain. The model does not require previous knowledge or empirical formulas, making it useful for many applications, such as image and speech recognition, video games, consumer-use prediction, and damage estimation. ANNs have the advantage of being able to process large amounts of data, with multiple inputs that are likely to be available in monitoring actual NPP structures. Input values feed into the ANN and pass through the hidden layers before arriving at the output value(s) for that set of inputs. The learning utilizes gradient-based methods and error backpropagation for efficient computation of gradients [17]. The process is repeated, and the model parameters (weights and biases) are updated until either the

desired number of runs (epochs) is completed, or the value of a loss function converges. Figure 10 displays a shallow neural network with only one hidden layer.

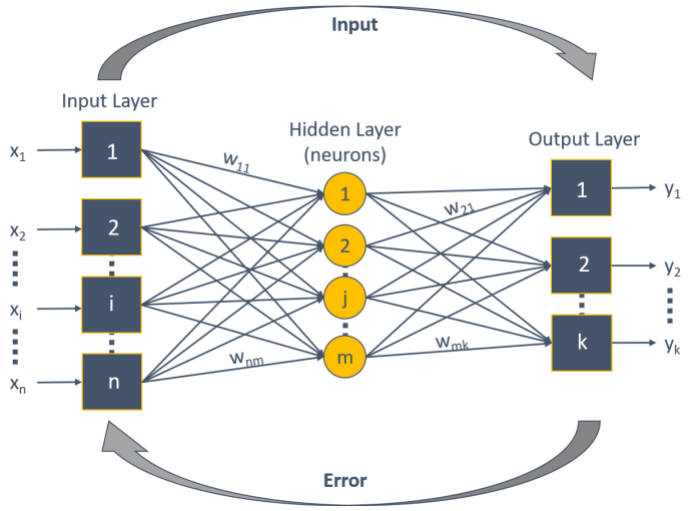


Figure 10: Example of an artificial neural network.

ML models have an enormous capacity to learn complex, nonlinear phenomena [18] and are well-suited for modeling nonlinear dynamics for complex, heterogeneous material like concrete. Both supervised and unsupervised techniques have been effectively used for concrete SHM [19]. Unsupervised learning is used to extract the damage-sensitive features from the diagnostic data [20]. For supervised learning, the model needs labeled training data consisting of the chosen model inputs (damage-sensitive features) for various damage levels in the real-world structure of interest. Since these data are typically not available, most of the aforementioned investigations use numerical simulations or scaled-down laboratory specimens for generating training as well as validating data [21]–[23].

Here, we train two simulation-data-driven models to perform damage localization. The first ANN model is a classification model. It identifies whether the damage is present or absent at/near a sensor location using the SBSum value at that sensor as an input. The second model estimates the SBSum metric at a sensor given the damage location, extent, and VAM test parameters. This model is used to compute the likelihood of measured data in a Bayesian estimation framework. The Bayesian estimation framework, in turn, enables damage localization and associated uncertainty quantification. We validate the performance of simulation-data-driven models using laboratory test data. Next, we describe the process of training data generation.

3.1 Training Data Generation- Finite-Element Analysis

Modeling wave propagation in heterogeneous media like concrete (with cement and randomly distributed aggregate) is a computationally challenging task. Thus, we neglect the material heterogeneity and simulate wave propagation in a homogeneous elastic specimen with concrete-like properties. Even with this simplifying assumption, the simulation of nonlinear wave propagation in 3D 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. To overcome this difficulty, we modeled the wave propagation in a 2D domain under plane strain assumptions.

Our previous work indicated this model is suitable for computing the sensitivity and specificity for VAM tests with different test parameters as well as measurement noise. The domain geometry, the crack locations, and the pump and probe locations used in our numerical simulations are shown in Figure 11.

The computational domain is 60.96-cm wide and 15.24-cm thick. It contains a crack of 12.7 cm in length at a mid-thickness (7.62 cm from the top surface). We used a commercial finite-element program (Abaqus [10]) to perform a numerical simulation. We discretized the domain using a structured mesh of 8-noded finite elements. We employed an implicit scheme for time integration of governing equations and ensured at least ten computational nodes were available over the smallest wavelength. We modeled the interaction at the crack interface using a hard, frictionless contact condition. The material properties used in our simulations are given in Table 1. The test parameters used in conducting VAM on this specimen are in Table 2, and the visual of the individual crack locations are combined in Figure 12. Data is taken at N-sensor locations along the specimen in the x-direction and used to calculate the LS and SBSum values (as described in Section 2.3.2).

Table 1. 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 2: Test parameters for the 2D FEA model.

Crack #	**Centroid (x-direction) [m]**	**Length [m]**	**Depth from Surface (z-direction) [m]**	**Amplitude Ratio (probe/pump)**	**Frequency Ratio (probe/pump)**	**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

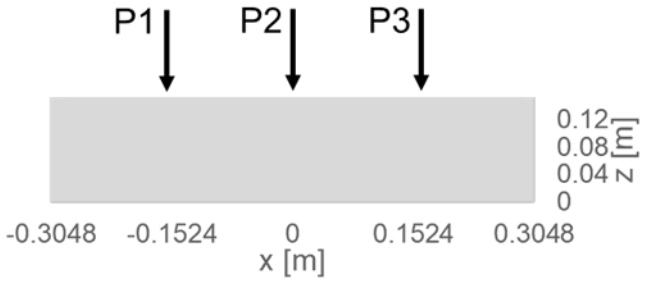
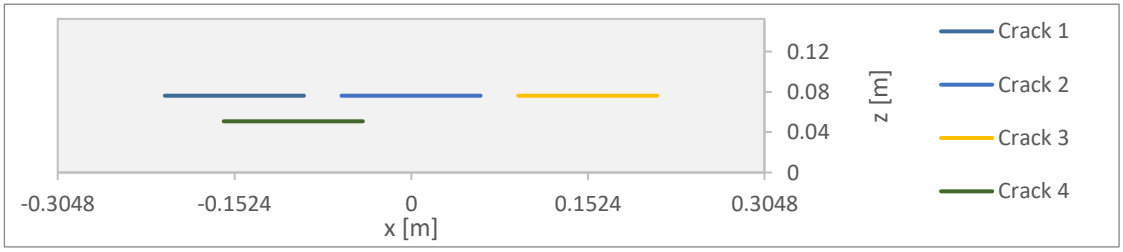
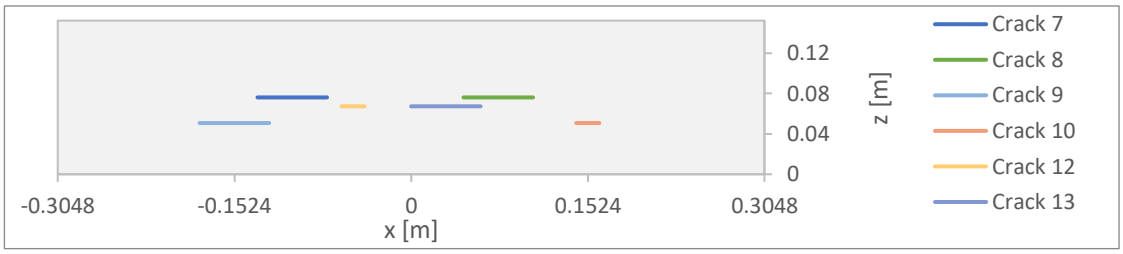


Figure 11: The domain geometry, and pump and probe locations used in the numerical simulations.

a.



b.



c.

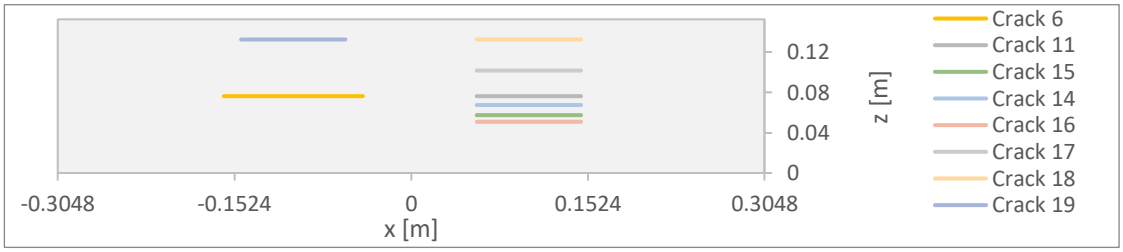


Figure 12: The domain geometry and crack locations used in the numerical simulations. Only one crack is present in the specimen domain for a given test.

3.2 Damage Diagnosis Methodology

We evaluate the performance of two neural network models built using training data retrieved from 2D finite-element analysis. The first model is a prediction model that estimates the SBSum values at all sensor locations given the VAM test parameters and a (candidate) damage location. This model and VAM test (SBSum) data can be used in an iterative damage diagnosis algorithm, where a search is done for the damage location that best fits the observed VAM test (SBSum) data. The second model uses the VAM test parameters as well as the measured SBSum values at a particular sensor to classify each sensor as one showing or not showing the signs of damage. This is the classification approach where each sensor location is classified into one of the two categories: damage present or damage absent. The details of the two types of models are discussed in detail in the following sections.

3.2.1 Damage localization using the prediction model

This model uses the (candidate) crack locations and VAM test parameters to predict the SBSum value at N-sensors for a VAM test. The inputs include the frequency ratio of the pump and probe, the amplitude ratio or the pump and probe, the x-location of the centroid of the crack, the depth of the crack from the surface, the x-location of the pumping and probing actuators and the x-location of the sensor (Table 2). The output is the SBSum value. In the data preprocessing phase, all inputs and outputs are normalized with a mean of zero and a standard deviation of 1. Additional validation data (Table 4) will use the same normalization factor for each variable to distribute the data in the same space as the training data. The regression feed-forward neural network to estimate SBSum (Figure 13) consists of two layers. The first layer has 16 nodes, and the second has 8. The Levenberg-Marquardt least squares method was used as the training algorithm associated with the loss function of the network. This network converged at 488 epochs and had an MSE of 0.123 using the 15% randomly selected cross-validation data.

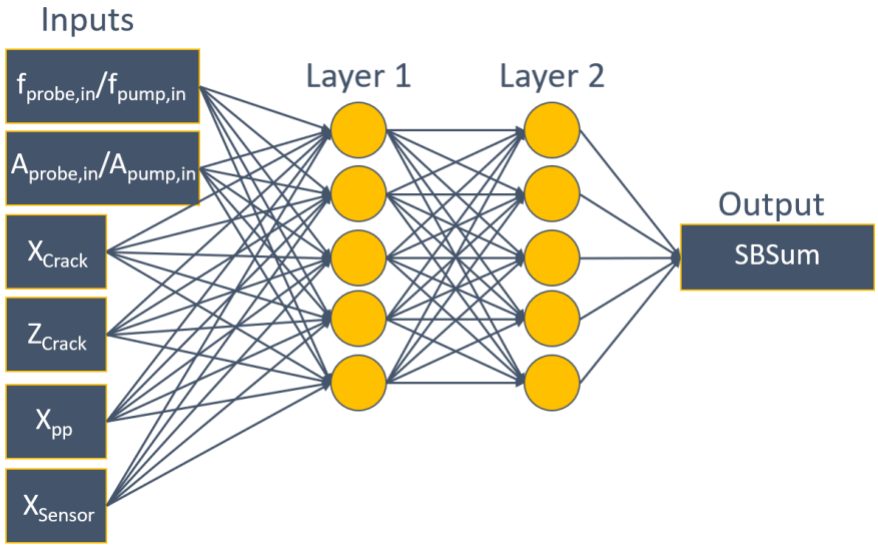


Figure 13: Schematic of ANN structure with six inputs, one output, and two hidden layers of five neurons each.

3.2.1.1 Bayesian Estimation for Damage Diagnosis

The prediction 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 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 this work, the x-location of the crack and the depth of the crack are the calibrated model inputs. The Bayesian update equation for the calibration can be written as,

$$P(\theta|SBSum) \propto P(SBSum|\theta) * P(\theta)$$

where $P(\theta_i)$ is the prior, $P(SBSum|\theta_i)$ is the likelihood, and $P(\theta_i|SBSum)$ is the posterior. $\theta_1 = X_{crack}$, $\theta_2 = Z_{crack}$ are the two model inputs to be calibrated. We used a uniform prior for both $P(\theta_1)$ and $P(\theta_2)$ with possible values along the entire length and width of the FEM since we assume the crack location to be unknown and that there will be an equal probability of damage throughout the 2D domain of interest. Using the regression neural network model and Markov Chain Monte Carlo (MCMC) simulations, we calculated the posterior distribution for values of θ_1 and θ_2. 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 algorithm uses a Markov process to approach a stationary distribution. This method generates the Markov chain using a proposal density for new steps, some of which are rejected by the method. The Metropolis-Hastings algorithm is popular in multidimension distribution sampling [24], [25]. The principle of Monte Carlo simulations is to take a set of samples from a target density in a particular state space and to use these samples to transition the target density for a different state. After each transition, the probabilities are converted into the current state. After a certain number of transitions, the probability converges and becomes stationary. This stable distribution is the posterior. Samples are accepted if they fit the posterior and rejected if they do not. All accepted samples form the targeted distribution [26].

3.2.2 Damage localization using classification models

This approach uses the SBSum values and VAM test parameters to determine whether the damage is present or absent in the vicinity of each given sensor. Here, we discuss the performance of different ANN classifiers for damage localization. The main difference between these classifiers is the number of inputs used and the presence/absence of down-sampling of the training data (to alleviate class imbalance). To obtain a set of important input variables, an analysis of variance was conducted on many possible model inputs, including the SBSum, frequency ratio, amplitude ratio, pump and probe locations, mean SBSum, and sensor locations, with the output being the damage indicator {0,1}. All of these inputs had p-values smaller than 0.05, supporting that they all were significant in damage estimation. The input parameters with the smallest p-values were the SBSum and amplitude ratios, which had values smaller than $1x10^{-5}$. The relative importance of other test parameters was judged by building different ANN classifiers with different inputs and comparing their validation accuracy (for crack 5).

3.2.2.1 Down-sampling

All crack sizes used for generating the training data are smaller than about a third of the computational domain length. 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 the class damage = 0 (absent), whereas only 16% of nodes belong to the other class (damage present). To alleviate the effects of class imbalance on model training, we down-sample the training data. Down-sampling is the process of learning on a subset of data where an even amount of training data is extracted from each of the classes. In our models, this suggests 50% of the data should come from the damaged area, while the remaining 50% should be from the undamaged sections. Down-sampling is used to ensure that the model does not overclassify into one group because there is more data from that set. For our data, we took the number of sensors (N) for each crack that had the output value as showing the presence of damage, where the crack was located. We sampled the same

number of sensors from the areas in which no damage was present (marked A and B in Figure 14). The sampled sensors were evenly distributed throughout the union of these areas. The number of sensors located in section N is approximately equal to the number of sensors located in sections A and B combined. The total number of data points after down-sampling from the 18 crack locations is 38,370. The structures of each of the candidate classification models (Table 3) are described next.

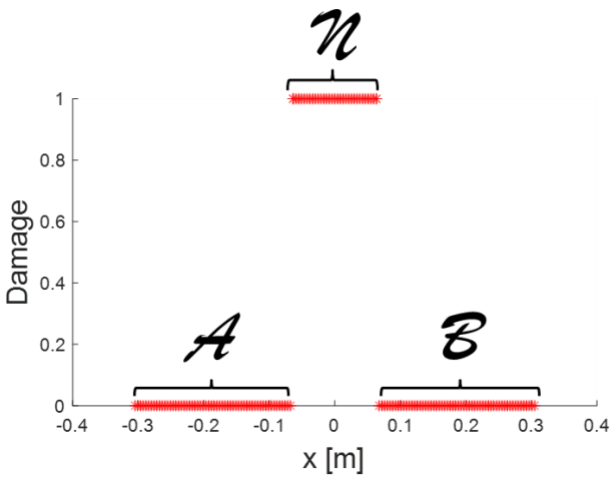


Figure 14: Sensors located in damaged and undamaged areas (sensors in region N are located in the damaged zone, sensors in regions A and B are not located in the damaged zone).

Table 3: Different ANN models to be evaluated.

Model Label	**Inputs**						**Down-sampling**
	$f_{probe,in}$/ $f_{pump,in}$	**$A_{probe,in}$/ $A_{pump,in}$**	**X_{pp}**	**X_{Sensor}**	**SBSum**	**μ_{SBSum}**	
A	X	X			X		
B	X	X			X	X	
C	X	X	X	X	X	X	
D	X	X			X	X	X
E	X	X	X	X	X	X	X

3.2.2.2 Model A

The damage classification feed-forward neural network (Figure 15) consists of two layers. The first layer has 16 nodes, and the second has 8. Scaled conjugate gradient was used as the training algorithm with 115,538 data points randomly separated into 70% for training, 15% for testing, and 15% for validation. This network converged at 172 epochs and had an overall training accuracy of 86%, sensitivity of 3%, and specificity of 99%.

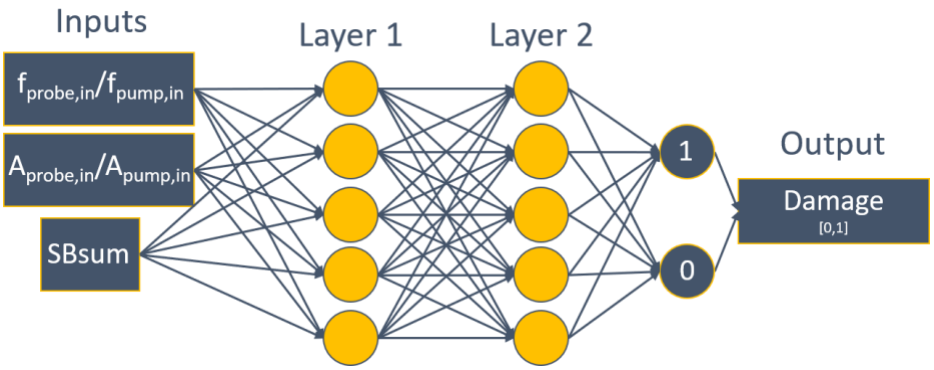


Figure 15: Classification Model A.

3.2.2.3 Model B

The damage classification feed-forward neural network (Figure 16) consists of two layers. The first layer has 16 nodes, and the second has 8. Conjugate gradient with Powell-Beale restarts was used as the training algorithm with 115,538 data points randomly separated into 70% for training, 15% for testing, and 15% for validation. This network converged at 173 epochs and had an overall training accuracy of 86%, sensitivity of 17%, and specificity of 99%.

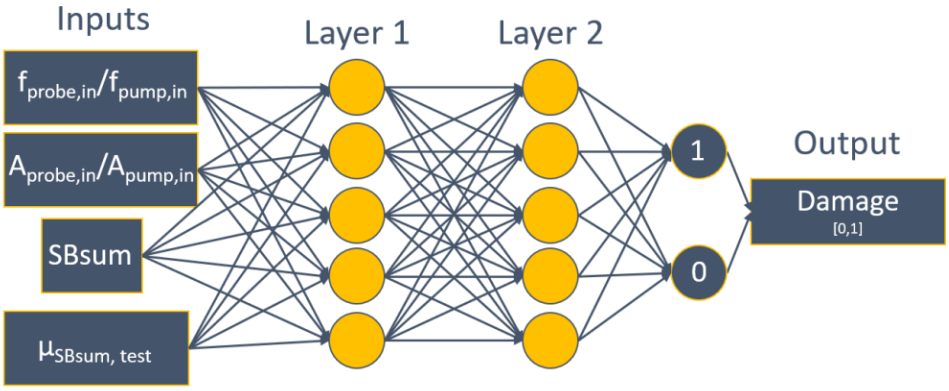


Figure 16: Classification Model B .

3.2.2.4 Model C

The damage classification feed-forward neural network (Figure 17) consists of two layers. The first layer has 32 nodes, and the second has 16. Conjugate gradient with Powell-Beale restarts was used as the training algorithm of the network with 115,538 data points randomly separated into 70% for training, 15% for testing, and 15% for validation. This network converged at 392 epochs and had an overall training accuracy of 90%, sensitivity of 54%, and specificity of 97%.

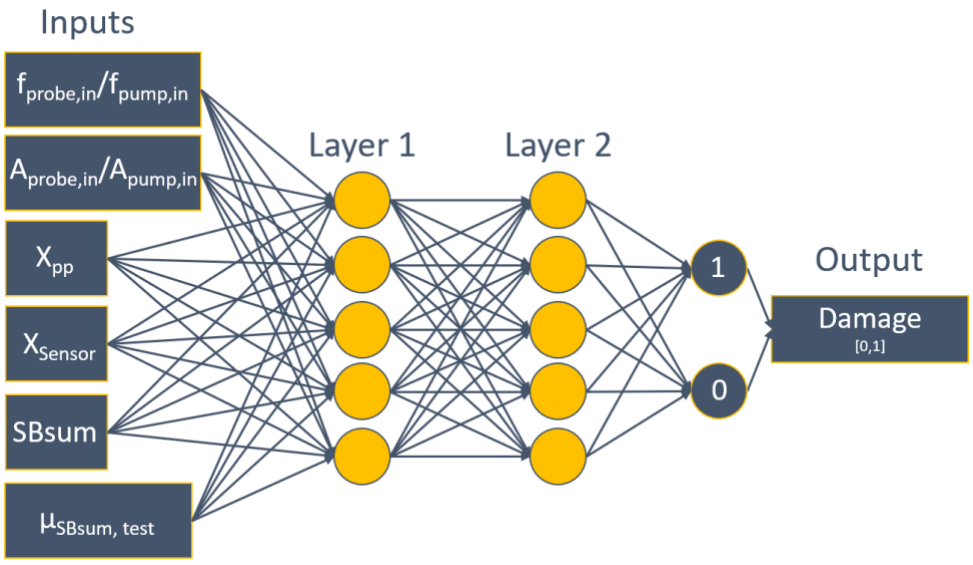


Figure 17: Classification Model C.

3.2.2.5 Model D

The damage classification feed-forward neural network (Figure 18) consists of two layers. The first layer has 32 nodes, and the second has 16. Conjugate gradient with Powell-Beale restarts was used as the training algorithm with 38,370 data points from down-sampling randomly separated into 70% for training, 15% for testing, and 15% for validation. This network converged at 222 epochs and had an overall training accuracy of 67%, sensitivity of 72%, and specificity of 60%.

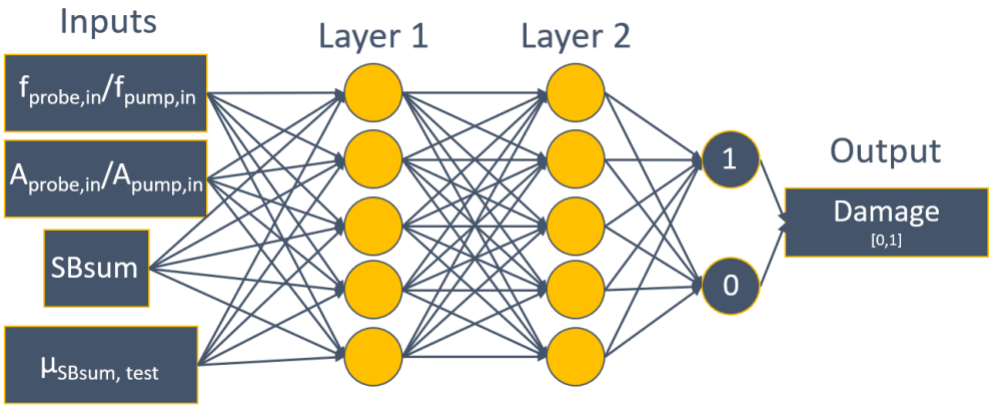


Figure 18: Classification Model D.

3.2.2.6 Model E

The damage classification feed-forward neural network (Figure 19) consists of two layers. The first layer has 32 nodes, and the second has 16. Conjugate gradient with Powell-Beale restarts was used as the training algorithm of the network with 38,370 data points from down-sampling randomly separated into 70% for training, 15% for testing, and 15% for validation. This network converged at 306 epochs and had an overall training accuracy of 84%, sensitivity of 90%, and specificity of 78%.

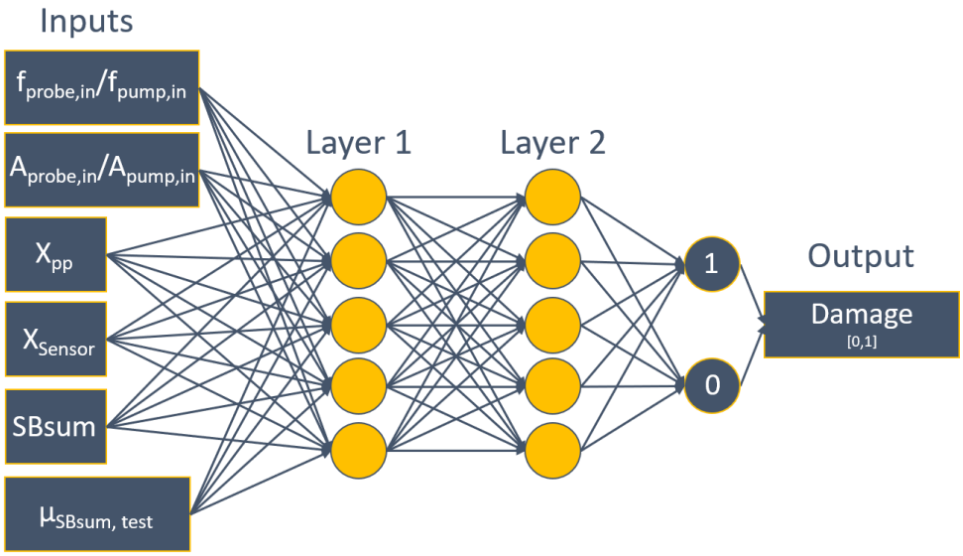


Figure 19: Classification Model E.

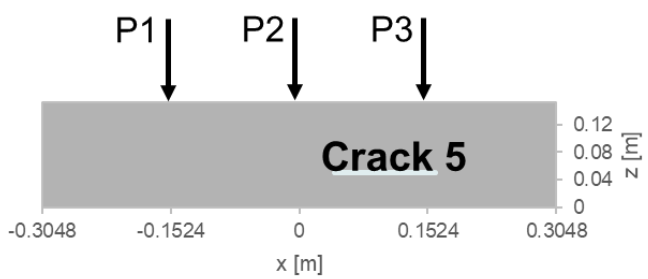
4. RESULTS AND DISCUSSION

In this section, we discuss the results for the ANNs constructed in Section 3.2, when they are applied to external validation data sets.

4.1 Validation Using Numerical Experiments

The proposed physics-informed ML framework was first validated by using simulation data for two cracks not used in the training of the ANNs. The locations of these cracks and testing parameters are described in Figure 20, Table 4. Two different damage diagnosis approaches were investigated. First, Bayesian estimation is used (along with the trained prediction ANN) to determine posteriors for the crack centroid, length, and depth. Next, the damage classification model is used to determine the centroid for the crack and crack length. The Bayesian estimation framework is then used to determine the posterior for damage depth. The second approach is not as computationally demanding as the first approach. For both of these applications, the number of sensors that could be used in a laboratory setting is used to conduct the VAM tests.

a.



b.

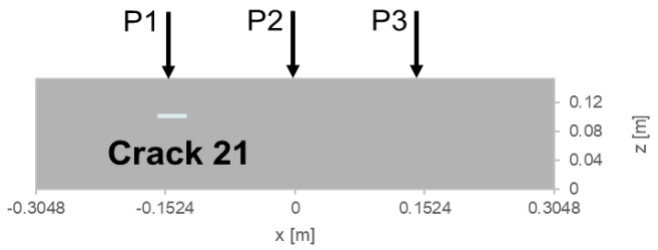


Figure 20: Crack locations for validation using numerical experiments (cracks 5 and 21).

Table 4: Test parameters for the 2D FEA models used for generating validation test data.

Parameter	**Crack 5**	**Crack 21**
Frequency Ratio (probe/pump)	10,12,14,16,18,20	18,20,22
Amplitude Ratio (probe/pump)	1, 0.5, 0.2, 0.1	1, 0.5, 0.2, 0.1
Pump/Probe Location (m)	-0.1524, 0, 0.1524	-0.1524, 0, 0.1524
Crack Centroid (x-direction) (m)	0.1016	-0.1448
Crack Depth from Surface (m)	0.1016	0.0508
Crack Size (m)	0.12	0.03

4.1.1 Damage location identification using the prediction model

For crack 5, Bayesian estimation was conducted using test data obtained from numerical experiments (Figure 20, Table 4) to determine the posterior distributions of the crack centroid, length, and depth. The prediction model was used to compute the likelihood of the *measured* SBSum values given the VAM test parameters (frequency ratio of 20, all amplitudes, and a pump and probe location of 0.1524 m). The prior distribution for the centroid was a uniform distribution over the entire length of the specimen (-.03048 to 0.3048 m), the prior distribution for crack size was a uniform distribution over half of the specimen length (0.01 to 0.3048 m), and the prior distribution for the depth was the entire depth (0 to 0.15 m). Metropolis-Hasting (MH) algorithm was used to generate 1,000 samples and were used in MCMC simulations with a 20% burn-in. Simulations were conducted using either 197, 29, or 15 sensor locations along the surface. One-hundred-ninety-seven sensors correspond to the number of finite-element nodes in the 2D finite-element model, and do not represent a realistic number of sensors for this slab. We tested two candidate sensor densities, with a sensor spacing of about 0.02 m and 0.043 m. The results show that the uncertainty in the posterior increased after a decrease in the sensor density.

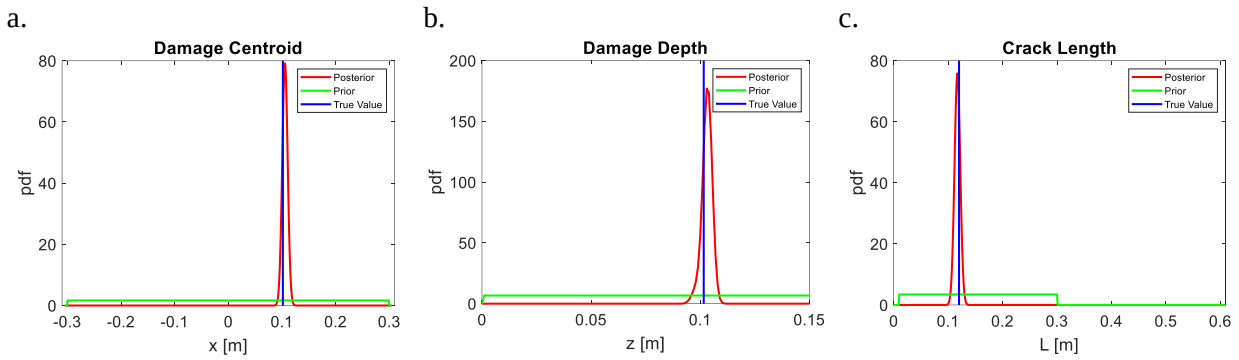


Figure 21: Bayesian estimation of damage for the validation crack 5 (197 sensors).

In order to obtain a probability of damage at each sensor location, the posteriors of the centroid and crack length were sampled. Specifically, 1,000 samples were obtained for the three posteriors to create 1,000 different damage maps. Every sensor located within half of the estimated size of the crack on each side of the centroid was considered to be damaged (1), and all other locations were assumed to be undamaged (0). These damage maps were averaged to obtain an overall damage probability, shown in Figure 22.

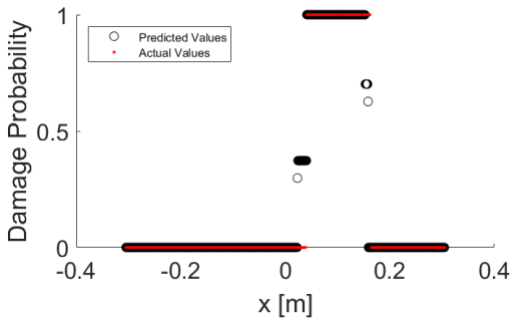


Figure 22: Damage probability for crack 5 using Bayesian damage diagnosis.

The results for the high-sensor density case show that the proposed methodology is successfully able to estimate damage location as well as its extent with very low uncertainty when a large amount of data is available (Figure 21 and Figure 22). Since the sensor density will not be available in real-world applications, we performed the damage estimation using lower sensor densities (with 29 and 15 sensors). Figure 23 and Figure 24 show results for the lower sensor density cases. It can be observed that the damage diagnosis methodology performed well, even for the lower sensor density cases.

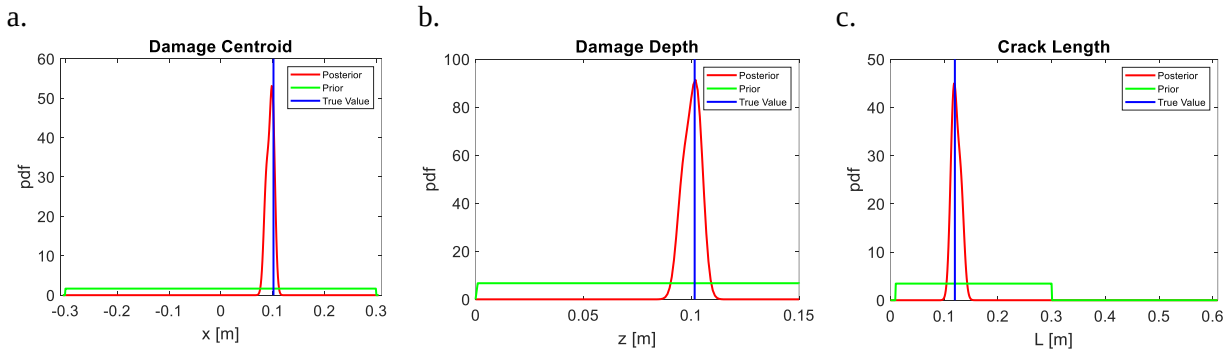


Figure 23: Bayesian estimation of damage for the external validation crack 5 (29 sensors).

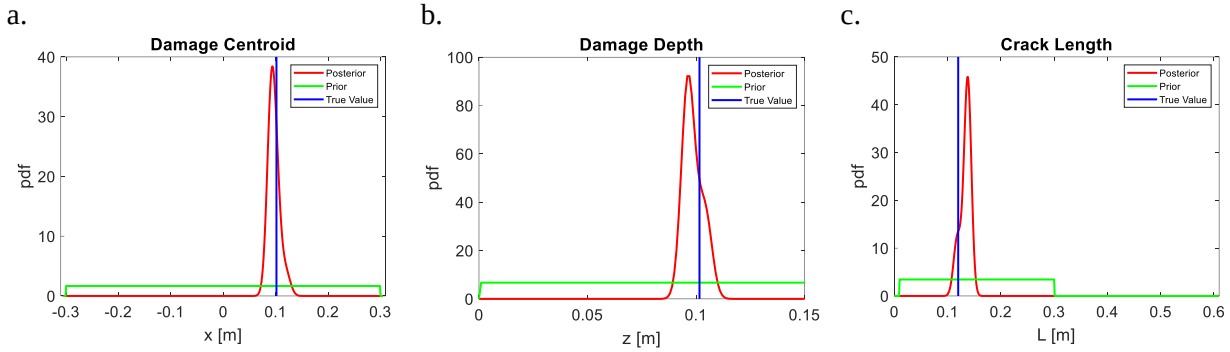


Figure 24: Bayesian estimation of damage for the external validation crack 5 (15 sensors).

It can be seen that the estimates for the lower density case (Figure 24) exhibited higher diagnosis uncertainty. It is relevant to understand whether this uncertainty can be reduced by including data from other VAM tests. To that end, VAM test data from two additional pump and probe excitation actuator locations were used for damage diagnosis. The results, shown in Figure 25, indicate that the diagnosis uncertainty can be reduced by including more test data. This is an important result for future, real-world applications of the methodology, where diagnosis uncertainty reduction may be desired to improve decision making.

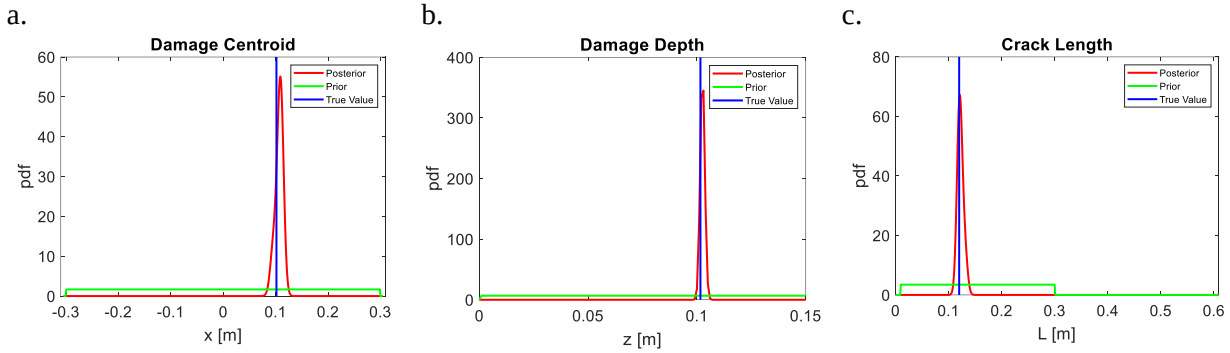


Figure 25: Bayesian estimation of damage for the external validation crack 5 (all pump and probe actuator locations, 15 sensors).

4.1.2 Damage location identification using classification and prediction models

The damage classification model for damage localization determines an approximate centroid for the crack and crack length. This model used all frequency ratios, all amplitude ratios, and all pump and probe locations for each validation crack.

4.1.2.1 Model A

First, Model A, trained using all the data, was analyzed. Both cracks were used for validation. For crack 5, 72 different VAM tests were performed using 197 sensors along the surface of the specimen. At each sensor location, the classification results were averaged to obtain the assimilated damage metric (Figure 26 (a)). The overall validation accuracy for this model was 80%, the sensitivity was 2%, and the specificity was 99.9%. For crack 21, 36 different VAM tests were performed at 195 sensors along the surface of the specimen and results are averaged at each sensor location (Figure 26 (b)). The overall validation accuracy for this model was 93%, the sensitivity was 6%, and the specificity was 99%. The average assimilated damage value for this model was very low (less than 0.3). This model contained no information on how the SBSum value at a sensor compared with other sensors in the same test. The

model also likely suffered from a class imbalance in the training data. The effects are seen in the poor sensitivity of the model, which indicates improvements are needed.

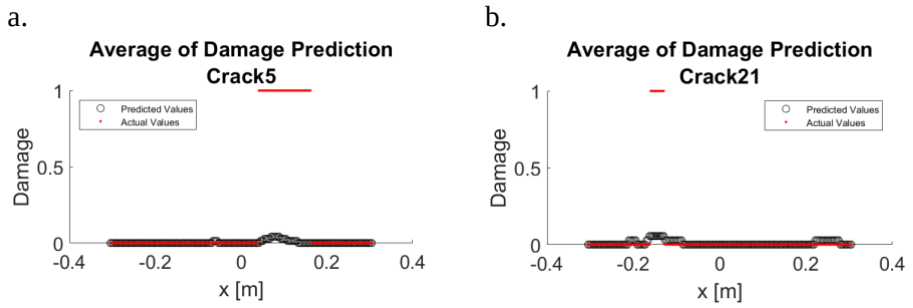


Figure 26: Validation of Model A on cracks 5 and 21.

4.1.2.2 Model B

Compared to Model A, Model B included an additional input: a metric that may inform the average SBSum value in a VAM test (μ_{SBSum}). Both cracks (numbers 5 and 21) were used for validation. For crack 5, 72 different VAM tests were performed using 197 sensors along the surface of the specimen. At each sensor location, the damage classification results were averaged to obtain the assimilated damage metric (Figure 27 (a)). The overall validation accuracy for this model was 81%, the sensitivity was 10%, and the specificity was 99%. For crack 21, 36 different VAM tests were conducted using 195 sensors along the surface of the specimen, and classification results were averaged at each sensor location (Figure 27 (b)). The overall validation accuracy for this model was 94%, the sensitivity was 31%, and the specificity was 99%. By the addition of the μ_{SBSum} input, Model B performed better than Model A. Although the damage location was accurately identified, the assimilated damage metric was still too low to conclude the damage was significant.

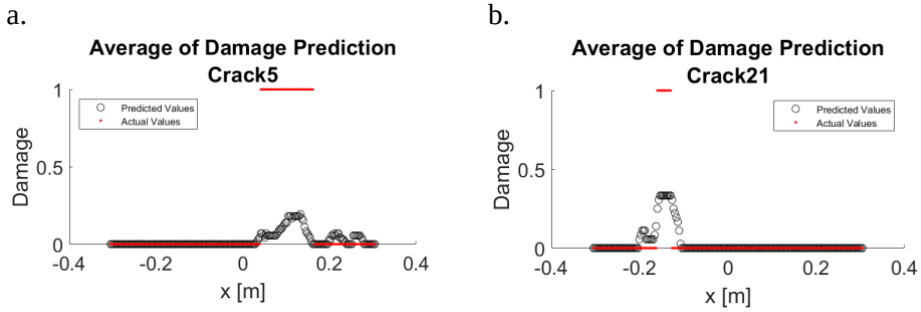


Figure 27: Validation of Model B on cracks 5 and 21.

4.1.2.3 Model C

The location of a sensor relative to the actuator may significantly affect the wave propagation and the SBSum value recorded at a sensor. Model C thus incorporated these two additional inputs (sensor and actuator location). Both cracks were used for validation. For crack 5, 72 different VAM tests were performed using 197 sensors along the surface of the specimen. At each sensor location, the classification results from 72 tests were averaged (Figure 28(a)). The overall validation accuracy for this model was 83%, the sensitivity was 32%, and the specificity was 97%. For crack 21, 36 different VAM tests were

conducted using 195 sensors along the surface of the specimen, and the results were averaged at each sensor location (Figure 28 (b)). The overall validation accuracy for this model was 90%, the sensitivity was 70%, and the specificity was 92%. The addition of the two inputs (x_{pp} and x_{Sensor}) in the training model significantly increased the sensitivity of the method. We explore next whether alleviating class imbalance in the training data could further improve results.

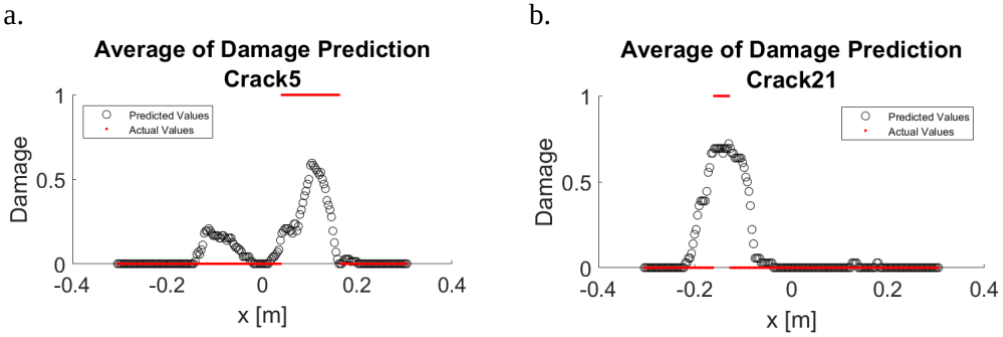


Figure 28: Validation of Model C on cracks 5 and 21.

4.1.2.4 Model D

Model D is similar to Model B trained using the down-sampled data. Both cracks were used for validation. For crack 5, 72 different VAM tests were performed using 197 sensors along the surface of the specimen. At each sensor location, the classification results from 72 tests were averaged (Figure 29 (a)) to obtain the assimilated damage metric. The overall validation accuracy for this model was 62%, the sensitivity was 63%, and the specificity was 62%. For crack 21, 36 different VAM tests were performed using 195 sensors along the surface of the specimen and the results were averaged at each sensor location (Figure 29 (b)). The overall validation accuracy for this model was 50%, the sensitivity was 78%, and the specificity was 47%. Although using the down-sampled data in training increased the sensitivity of the model, both the overall accuracy and specificity greatly decreased. The overall average of damage metric value increased to almost 1 for Crack 21. However, this model suffered from a significant reduction in the specificity as compared to Model B.

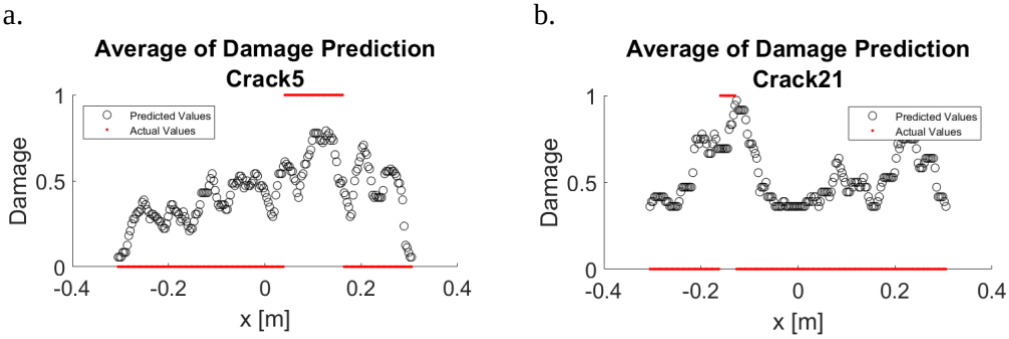


Figure 29: Validation of Model D on cracks 5, and 21.

4.1.2.5 Model E

Model E was the same as Model C—trained using the down-sampled data. Both cracks were used for validation. For crack 5, 72 different VAM tests were performed using 197 sensors along the surface of the specimen. At each sensor location, the classification test results were averaged (Figure 30(a)) to obtain

the assimilated damage metric. The overall validation accuracy for this model was 81%, the sensitivity was 75%, and the specificity was 82%. For crack 21, 36 different VAM tests were conducted using 195 sensors along the surface of the specimen and the results were averaged at each sensor location (Figure 30 (b)). The overall validation accuracy for this model was 73%, the sensitivity was 100%, and the specificity was 71%. Although specificity for this model was higher than Models B and D, and although it was able to achieve the highest sensitivity amongst all the models, the assimilated damage metric value (which can be viewed as damage probability) at each sensor above the hidden crack was near or equal to 1. This model overestimated the size of the damage but predicted damage nearly 100% of the time.

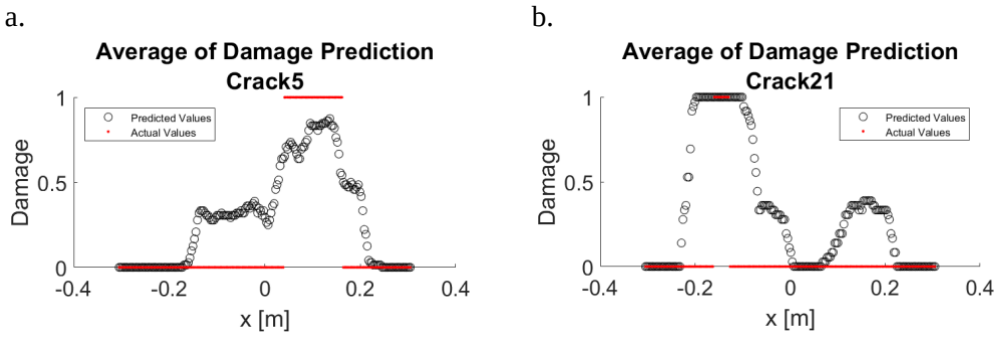
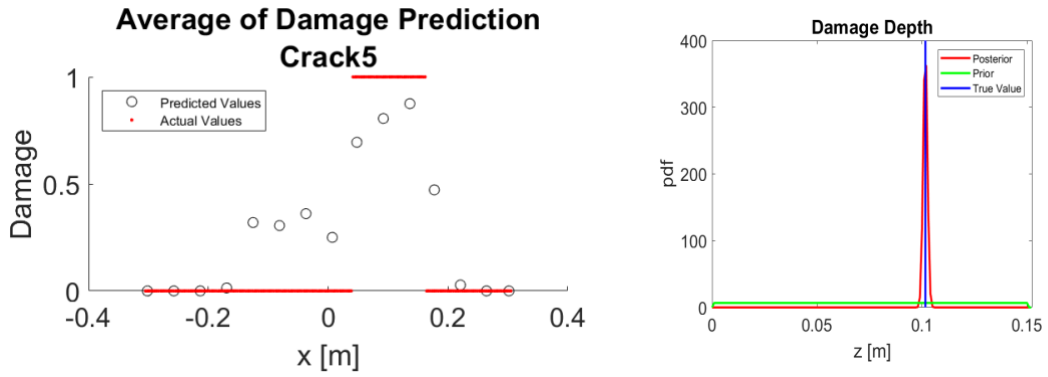


Figure 30: Validation of Model E on cracks 5 and 21.

Since Model E showed the best validation performance amongst all models considered in this work, it was chosen to investigate the effect of sensor density on diagnosis uncertainty. To that end, damage classification was performed using only 15 sensors spaced every 0.043 m along each of the specimens. This sensor distribution was similar to what would be used in experimental VAM testing [11]. Sensors indicating locations where the average of the classification results from all VAM tests were greater than 0.5 were assumed to indicate the presence of damage. For crack 5, the damage identification results using classification were averaged for all 72 VAM tests at the 15 sensors (Figure 31 (a)). The estimated crack centroid was 0.0884 m (true=0.1016) and the length was 0.0921 m (0.12). For crack 21, the damage identification results using classification are averaged for all 36 different VAM tests at the 15 sensors (Figure 31 (b)). The estimated crack centroid is −0.154 m (−0.1448 m), and the length is 0.125 m (0.03 m). Next, prediction model-based Bayesian estimation is performed to estimate the damage depth. VAM test results for a frequency ratio of 20, all amplitudes, and all pump and probe actuator locations were used for both validation cracks. The estimated crack centroid and length for each validation crack are used to define the crack location, and a search is done for crack depth. The prior for the crack depth is assumed to be the entire depth of the specimen. The models used 1,000 MH samples and a 20% burn-in. The results of the combined classification and prediction model estimates are shown in Figure 31.

a.



b.

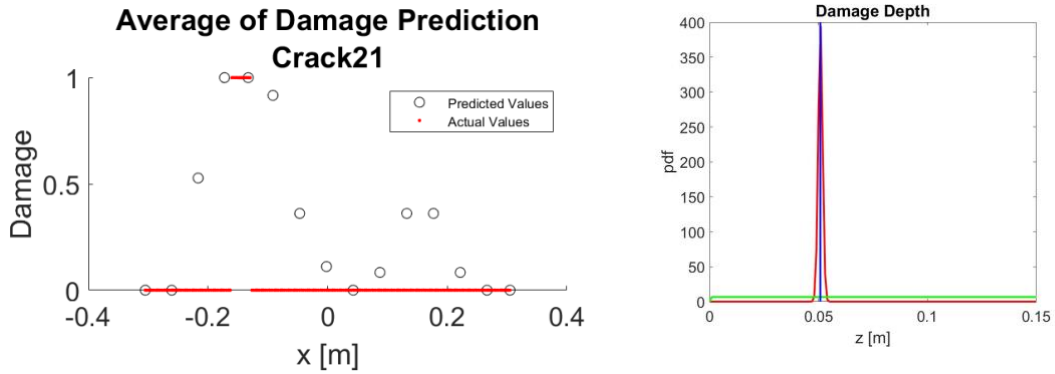


Figure 31: Validation of Model E on cracks 5 and 21 for 15 sensor locations along the surfaces of the specimens and posterior for the depth calculated using the estimated crack centroids and lengths for each validation specimen.

Based on the performance of Models A–E, we can see that the addition of physically significant input variables and down-sampling the training data increased the sensitivity of the damage diagnosis methodology while maintaining high specificity. Table 5 and Table 6 give a summary of model performance on the three external validation cracks. For each crack, the overall validation accuracy and specificity remained consistently high for Models A, B, and C, even while new inputs were being incorporated. The addition of inputs did significantly affect the sensitivities in those models. By adding the μ_{SBSum} to Model A, Model B had a sensitivity that was 3–5 times higher for each crack, and the addition of X_{pp} and X_{Sensor} only improved it further.

Table 5: Validation for multiple models using Crack 5.

Model Label	Inputs						Down-sampling	Model Performance for Crack 5		
	$f_{probe,in}/f_{pump,in}$	$A_{probe,in}/A_{pump,in}$	X_{pp}	X_{Sensor}	SBSum	μ_{SBSum}		Accuracy	Sensitivity	Specificity
A	X	X			X			0.80	0.02	0.99
B	X	X			X	X		0.81	0.10	0.99
C	X	X	X	X	X	X		0.83	0.32	0.97
D	X	X			X	X	X	0.62	0.63	0.62
E	X	X	X	X	X	X	X	0.81	0.75	0.82

Table 6: Validation for multiple models using Crack 21.

Model Label	Inputs						Down-sampling	Model Performance for Crack 21		
	$f_{probe,in}/f_{pump,in}$	$A_{probe,in}/A_{pump,in}$	X_{pp}	X_{Sensor}	SBSum	μ_{SBSum}		Accuracy	Sensitivity	Specificity
A	X	X			X			0.93	0.06	0.99
B	X	X			X	X		0.94	0.31	0.99
C	X	X	X	X	X	X		0.90	0.70	0.92
D	X	X			X	X	X	0.50	0.78	0.47
E	X	X	X	X	X	X	X	0.73	1.00	0.71

The results of the numerical experiment show that the prediction model, classification model, and Bayesian estimation framework can be successfully used for damage localization. In addition, to provide the expected damage localization parameters, the proposed methodology can also estimate diagnosis uncertainty. Adding diagnostic information (by conducting more VAM tests) reduces the diagnosis uncertainty. Next, we perform validation using laboratory experiments.

4.2 Validation Using Laboratory Experiments

The plain concrete slab described in Section 2.2 is used for validation. The details of VAM tests performed on this specimen can be found in [11]. A schematic of the sensors on the top surface of the specimen is shown in Figure 32. Six vertical slices were taken along sensor locations (parallel to the x-axis) for the cement slab specimen (Figure 33). Damage localization using only the prediction model (Bayesian estimation for three damage parameters), as well as damage localization using the classification model and prediction model (Bayesian estimation for damage depth) were performed for each (2D) slice.

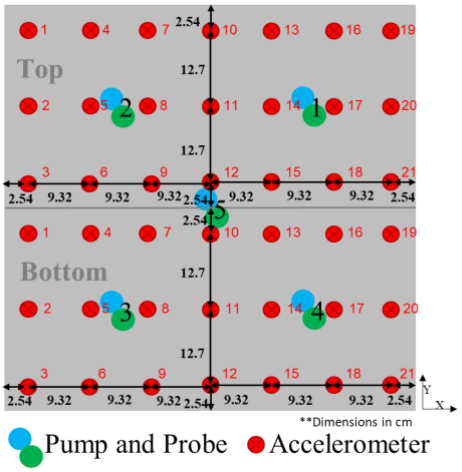


Figure 32: Experimental setup for Vanderbilt ASR-induced specimen.

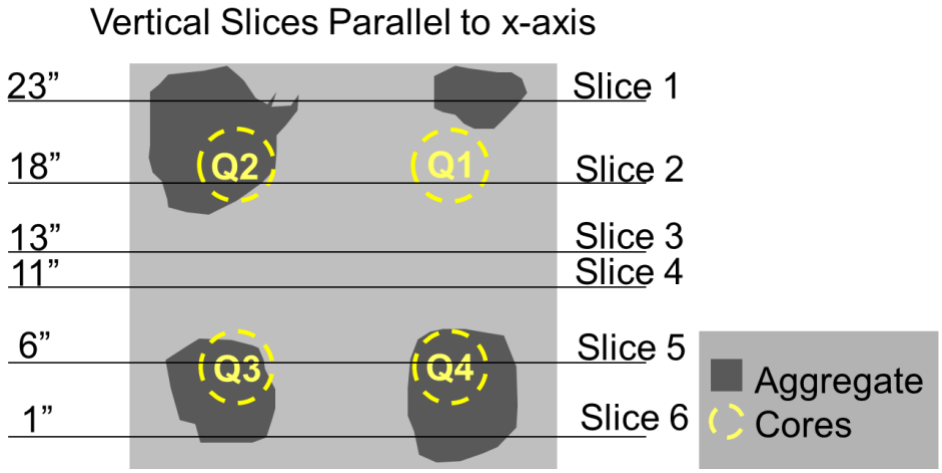


Figure 33: Slice locations for experimental analysis.

The Bayesian estimation provides the posterior distributions for the crack centroid, length, and depth. The model used the SBSum values for the crack given VAM test parameters of the frequency ratio of 20, all amplitudes, and all pump and probe locations (x-direction). The prior distribution for the centroid was a uniform distribution over the entire length of the specimen (−.03048 to 0.3048 m), the prior distribution for crack size was a uniform distribution over half of the specimen length (0.01 to 0.3048 m), and the prior distribution for the depth was the entire depth (0 to 0.15 m). 1,000 MH samples were used in MCMC simulations with a 20% burn-in.

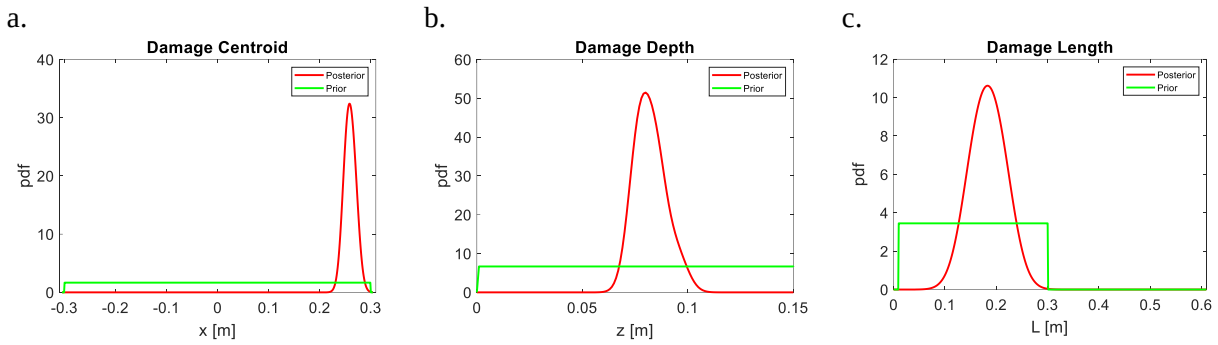


Figure 34: Bayesian estimation of damage for slice 6 using a frequency ratio of 20, all amplitude ratios, and all pump and probe actuator locations at seven sensors.

In order to obtain a 2D damage map for the slab, the posteriors of the centroid and crack length were sampled 1,000 times for each 2D (vertical) slice to create 1,000 different damage maps. Every sensor located within half of the size of the crack on each side of the centroid was considered to be damaged (1), and all other locations were assumed to be undamaged (0). These damage maps were used to obtain the overall damage probability (Figure 35). The results for quadrants 3 and 4 are consistent with the petrography results discussed in Section 2.4.1, and the visual inspection of damage is discussed in Section 2.4.3. That is, the damage location predictions from destructive tests and VAM tests are in agreement in the bottom half of 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 6). The low sensor density (7 sensors along the length of the slab), as well as the use of (vertical) 2D slices for damage diagnosis, may be possible reasons for the observed difference. The prediction ANN model did successfully diagnose damage around the edges of the void in quadrant 2, as visually observed in the broken-down slab.

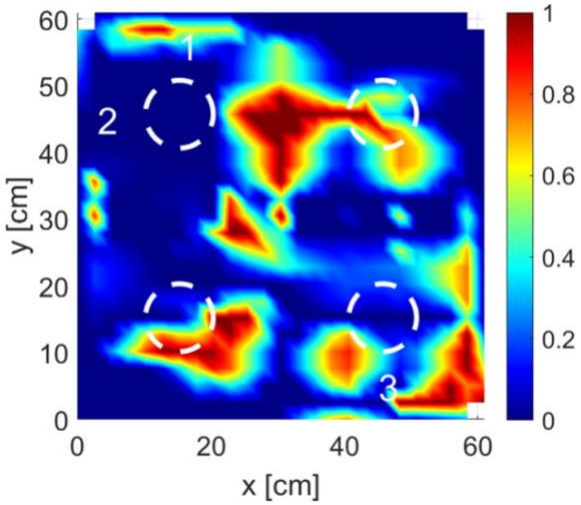


Figure 35: Damage probability for the plane concrete specimen using the prediction ANN model and Bayesian estimation.

Next, the classification model trained using the down-sampled data (Model E) was used to predict damage for each vertical 2D slice. Damage prediction for each slice was interpolated from slice to slice to create an overall damage prediction map (Figure 36). The classification model estimate showed a lower probability of damage as compared to the prediction model results. The model successfully localized damage in quadrants 3 and 4. However, the area of damage in quadrant 2 was larger than we expected to detect with VAM, and the high-probability damaged area appeared 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 results of multiple 2D cross-sections of a slab to obtain a damage map.

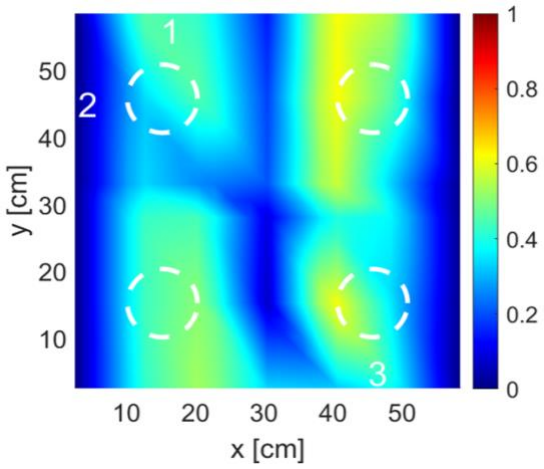


Figure 36: Average damage classification for each slice of the plain concrete specimen.

In this section, we discussed the validation of VAM-based damage diagnosis using ML models built as discussed in Section 3.2. The prediction model estimates the damage index for given values test parameters, damage location, and damage extent The prediction model thus 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 thus computationally expensive to use this model for damage diagnosis. The classification model is a diagnostic model, which provides the state of damage at (near) 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 prediction model and a parameter estimation algorithm. If only the damage presence and damage extent are needed, then the classification model can potentially provide the estimates quickly.

The performance of both diagnosis methods was compared with numerical experiments and with 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 ML model results for the bottom half of the slab matched better than for the top half of the previously reported experimental results (Figure 6, Figure 35, Figure 36). Of the two models studied, the prediction model performed slightly better than the classification model as far as results for the bottom half were concerned. Both prediction and classification model-based damage diagnoses showed high damage probability regions in the top half of the slab. These regions were not highlighted as damaged zones by the data-driven approach.

We suggest that the slice-by-slice diagnosis procedure could be the main reason for this discrepancy. The data-driven method used VAM test results (SBSum values) on the entire slab surface to determine the sensors that show *higher* or *lower* SBSum values. The ML methods on the other hand considered only a subset of sensors (seven sensors in a row) to decide which sensors showed high SBSum values. If a sensor has a *high* SBSum value compared to six sensors in its row, even though it has a *low* value when all 42 sensors are considered (as previously reported, Figure 6), then it will still be highlighted as a sensor *showing signs of damage* in a slice-by-slice diagnosis but will be regarded as *not showing signs of*

damage when all 42 sensors are considered together. Hence the observed propensity of ML models for showing false-positive results could be mitigated by incorporating training data from three-dimensional finite-element simulations. This may enable VAM-based diagnosis to use the mean SBSum value from all sensors installed on the testing surface and also help to build a multifidelity diagnostic methodology that utilizes a slice-by-slice strategy in a more accurate manner.

5. CONCLUSIONS AND FUTURE WORK

The objectives of this report were to examine the utility of physics-informed ML for VAM-based ASR damage diagnosis. The VAM technique was applied to a 24 in. × 24 in. × 6 in. slab cast at Vanderbilt University with four pockets of reactive aggregates.

The main outcomes of the experiments and subsequent analyses include:

1. The VAM testing procedure was automated to conduct the tests at multiple test parameters with minimal human intervention, producing in real-time a damage map that shows likely areas of damage.
2. The vibration-based technique successfully detected and localized ASR damage in the medium-sized plain cement slab using VAM tests.
3. Data-driven ANN models can be used for damage estimation, given numerical simulations of VAM results for 2D FEA specimens. Both the damage classification model and damage prediction model were successful in localizing damage. Damage classification is computationally inexpensive to run to quickly determine where damage may be located along the surface of the specimen. The damage prediction model still needs to be used to determine damage depth.
4. The proposed ML-based approach provides damage depth, which was not predicted using any of the previously reported damage localization methods.

Future work needs to focus on the following issues:

1. The vibration techniques, enhanced by ML, have generally shown good performance in terms of identifying the damage locations. Future work needs to focus on scaling up the technique to field implementation for damage diagnosis in NPP concrete structures. One of the key challenges is the number of sensors (accelerometers) needed. For a large structure, it is not feasible to use a large number of accelerometers; therefore, non-contact sensing, full-field observation techniques (such as laser velocimetry) might be beneficial.
2. A suitable full-field observation technique needs to be integrated with the automated VAM testing tool to enable fast, reliable VAM-based damage localization for real-world structures. This can be used in NPPs requiring frequent monitoring of ASR damage growth.
3. This work considered a plain concrete specimen. Future work needs to investigate the VAM and ML methodologies in reinforced concrete specimens.
4. In this report, we showed a ML approach to construct a data-driven model, based on 2D numerical simulation of VAM tests. Future models will include additional crack sizes, multiple cracks in 2D, 3D FEA model data, and experimental specimen data. Future work should leverage such a model, in combination with a physics model for material degradation, to perform prognosis and uncertainty quantifications of future damage evolution.
5. Future work needs to investigate the incorporation of damage diagnosis and uncertainty quantification in developing a robust prognostics and health management framework.

6. REFERENCES

[1] D. J. Naus, "Activities in Support of Continuing the Service of Nuclear Power Plant Safety Related Concrete Structures," in *Infrastructure Systems for Nuclear Energy*, John Wiley & Sons, Ltd, 2013, pp. 475–495.

[2] J. A. Christensen, "NPAR approach to controlling aging in nuclear power plants," 1990.

[3] S. Mahadevan, V. Agarwal, K. Neal, D. Kosson, and D. Adams, "Interim Report on Concrete Degradation Mechanisms and Online Monitoring Techniques, INL/EXT 14 33134," Nov. 2014.

[4] S. Mahadevan *et al.*, "A Demonstration of Concrete Structural Health Monitoring Framework for Degredation due to Alkali-Silica Reaction, INL/EXT 16 38602," Nov. 2016.

[5] S. Mahadevan *et al.*, "Interrogation of Alkali-Silica Reaction Degraded Concrete Samples using Acoustic and Thermal Techniques to Support Development of a Structural Health Monitoring Framework, INL/EXT 17 41852," Nov. 2017.

[6] V. Agarwal, S. Mahadevan, S. Miele, P. Karve, and J. Finfrock, "Enhancement of Structural Health Monitoring Framework by Optimizing Vibro-Acoustic Modulation Technique to Localize Alkali-Silica Reaction Degradation in Medium-Sized Concrete Samples," 2018.

[7] S. Miele *et al.*, "Vibration-Based Techniques for Concrete Structural Health Monitoring," 2019.

[8] 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.

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

[10] 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.

[11] P. Karve *et al.*, "Vibro-acoustic modulation and data fusion for localizing alkali-silica reaction-induced damage in concrete," *Struct. Heal. Monit.*, p. 147592172090550, Feb. 2020.

[12] X. J. Chen, J.-Y. Kim, K. E. Kurtis, J. Qu, C. W. Shen, and L. J. Jacobs, "Characterization of progressive microcracking in Portland cement mortar using nonlinear ultrasonics," *NDT E Int.*, vol. 41, no. 2, pp. 112–118, 2008.

[13] 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.

[14] S. Kim *et al.*, "Crack detection technique for operating wind turbine blades using Vibro-Acoustic Modulation," *Struct. Heal. Monit.*, vol. 13, no. 6, pp. 660–670, 2014.

[15] 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.

[16] 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.

[17] A. S. Lapedes and R. M. Farber, "How neural nets work," in *Neural information processing systems*, 1988, pp. 442–456.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[26] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, “An introduction to MCMC for machine learning,” *Mach. Learn.*, vol. 50, no. 1–2, pp. 5–43, 2003.