

University of Texas at Arlington

**MavMatrix**

---

Civil Engineering Dissertations

Civil Engineering Department

---

2023

## Infrastructure damage detection leveraging advanced AI techniques

Yin Chao Wu

Follow this and additional works at: [https://mavmatrix.uta.edu/civilengineering\\_dissertations](https://mavmatrix.uta.edu/civilengineering_dissertations)



Part of the [Civil Engineering Commons](#)

---

### Recommended Citation

Wu, Yin Chao, "Infrastructure damage detection leveraging advanced AI techniques" (2023). *Civil Engineering Dissertations*. 497.

[https://mavmatrix.uta.edu/civilengineering\\_dissertations/497](https://mavmatrix.uta.edu/civilengineering_dissertations/497)

This Dissertation is brought to you for free and open access by the Civil Engineering Department at MavMatrix. It has been accepted for inclusion in Civil Engineering Dissertations by an authorized administrator of MavMatrix. For more information, please contact [leah.mccurdy@uta.edu](mailto:leah.mccurdy@uta.edu), [erica.rousseau@uta.edu](mailto:erica.rousseau@uta.edu), [vanessa.garrett@uta.edu](mailto:vanessa.garrett@uta.edu).

Infrastructure damage detection leveraging advanced AI techniques

BY

YinChao Wu

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2023

## ABSTRACT

The dissertation explores into structural damage identification and prediction through the application of machine learning. Firstly, it enhances impact echo delamination mapping by incorporating two signal image parameters—signal duration and starting time after zero-crossing. The research also explores the efficiency of machine learning concerning image resolution and the number of images. This approach refines the evaluation of bridge damage in field tests. Two crucial signal parameters are investigated: signal duration and signal starting time after the zero-crossing point. These parameters exert a direct impact on the frequency domain, a crucial metric in impact echo analysis. This metric discerns signals reflected from areas with delamination and those from non-delaminated areas by assessing the accumulated low-frequency energy within the range of 1kHz to 5kHz. Considering the complexity of signals collected during field tests, artificial neural network (ANN) and convolution neural network (CNN) are employed to identify delamination signals using these two parameters. The study reveals that a signal duration of 1 millisecond with a starting time at 0.1 millisecond yields the highest accuracy in damage identification. A comparative analysis between delamination maps before and after applying deep learning (DL) suggested signal parameters demonstrates the updated results' enhanced accuracy in distinguishing delamination areas correctly.

Secondly, the study differentiates between homogeneous and inhomogeneous mediums with random aggregate sizes and distributions, utilizing wave scattering models and wave response variations. The internal crack geometry is identified through wave response variation with machine learning, and the finite element model design for random aggregate is detailed. Analyzing cracks at different depths, the wave response variation (WRV) study observes lower impulse frequencies for shallower cracks, computed from forward and incident waves, and higher frequencies for deeper cracks. Inhomogeneous medium (IHM) complexity increases with random aggregate size and distribution, where larger aggregates significantly impact the WRV pattern, causing more energy attenuation in the forward wave than smaller aggregates. Machine learning (ML) techniques accurately predict cracks and elucidate the black box process,

emphasizing the impact of aggregate information on WRV. The study outlines the framework of internal damage in IHM using ML technology.

Lastly, the dissertation employs bridge weight-in-motion signals to predict varying levels of structural damage. A finite element model using the kinetic contact method is developed and verified with bridge-vehicle motion theory. The study focusses on the development of moving load FE model. It introduces a comprehensive assessment of 1) a unique finite element (FE) simulation approach, which leverages the kinematic contact enforcement (KCE) method, verified with the vehicle-bridge interaction (VBI) theory, 2) laboratory tests and field test were conducted and performed to calibrate and verify the FE model, and 3) ML techniques to identify and automatically predict structural damages from the structural response. The KCE method is a new approach to simulating vehicle motion in a BWIM model, which is used to carry out actual structural responses to motion. By providing contact conditions between elements, including contact type, material properties, and element speed, the KCE method enables realistic simulation of both vehicle motion and structural response. The responses generated by the FE simulation are further analyzed using a feature selection method that ranks the importance of various ML models. Specifically, the prediction model includes a decision tree, a support vector machine (SVM), backpropagation (BP), and XGBoost. The results show that XGBoost with its assembly decision tree provides the most reliable outcomes. Three distinct studies leverage machine learning algorithms for tasks such as signal identification, damage identification, and the prediction of structural damages.

## ACKNOWLEDGEMENTS

Thank you to everyone who has helped and supported me over the past few years. I express my gratitude to Dr. Suyun Ham for providing me the opportunity to earn a Ph.D. degree and for dedicating much time to teach me both within and outside the academic realm. I extend my thanks to the committee members, Dr. Shin-ho Chao, Dr. Xinbao Yu, and Dr. Chen Kan, for their service and valuable feedback at different stages of the research.

I am grateful to every lab mate who worked together, including Dr. Sanggoo Kang, Dr. David Dafnik, and Dr. Yeongsok Jeong, for imparting knowledge, skills, and sharing life experiences. Special thanks to undergrad and graduate students Avishkar, Yuan, Felix, Emanuel, and Lance for their extra time and effort working with me.

My sincere thanks to my mom, dad, sister, and friends who have patiently listened to my arguments. A heartfelt thank you to my wife and her parents for their unwavering support even when I had nothing. Lastly, I appreciate my younger self for persevering and preparing for English tests during military service, as well as for not giving up during the PhD degree journey.

# TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION .....	1
1.1	Problem Statement .....	1
1.2	Objectives and Approach .....	2
CHAPTER 2	BACKGROUND .....	3
2.1	Overview .....	3
2.2	Bridge Defects .....	3
2.2.1	Delamination and cracks .....	3
2.2.2	Structural Damage .....	4
2.3	Nondestructive Testing .....	5
2.3.1	Wave propagation .....	5
2.3.2	Wave attenuation .....	6
2.3.3	Impact-echo .....	6
2.4	Signal Identification & Damage Prediction .....	7
2.4.1	Machine learning .....	7
2.4.2	Artificial neural network .....	8
2.4.3	Support vector machine .....	9
2.4.4	Decision tree .....	10
2.4.5	Extreme gradient boosting .....	11
2.4.6	Convolution neural network .....	12
2.4.7	Accumulated local effects .....	13

2.4.8	Generative adversarial network.....	13
2.4.9	Confusion matrix.....	14
CHAPTER 3 IE DAMAGE IDENTIFICATION & CNN EFFICIENCY .....		15
3.1	Chapter overview .....	15
3.2	Data preparation.....	16
3.2.1	Impact echo laboratory test data .....	16
3.2.2	Multi-input data of damage identification model.....	19
3.2.3	Graphical-CNN efficiency study data.....	20
3.3	Result and discussion.....	21
3.3.1	Impact echo laboratory testing result .....	22
3.3.2	Damage identification result .....	23
3.3.3	CNN efficiency analysis result.....	32
CHAPTER 4 IE SIGNAL EXTRACTION STUDY .....		40
4.1	Chapter overview .....	40
4.2	Field test environment.....	41
4.3	Signal extraction ML algorithm.....	44
4.3.1	ANN.....	44
4.3.2	CNN.....	46
4.4	Results and discussions.....	48
4.4.1	CNN Signal extraction result .....	49
4.4.2	Delamination results from field test.....	55

CHAPTER 5	INTERNAL CRACK IDENTIFICATION IN IHM MEDIUM .....	59
5.1	Chapter overview .....	59
5.2	Wave response variation .....	60
5.2.1	Mathematical formulation with the relation between the wave scattering and various material-crack properties in an elastic medium .....	60
5.2.2	Comparative study of wave response variation in different geometry sizes using an infinite element in homogeneous medium.....	64
5.2.3	Develop the damage prediction model using ML technology .....	67
5.2.4	Laboratory test .....	69
5.3	Develop the damage prediction model.....	71
5.3.1	Develop the damage prediction model using ML .....	71
5.3.2	Field test verification .....	74
5.4	Result and discussion.....	75
5.4.1	WRV study in homogenous medium.....	76
5.4.2	IHM numerical simulation model verified with laboratory test.....	79
5.4.3	Field test damage prediction .....	89
CHAPTER 6	STRCUTRUAL DAMAGE PREDICTION .....	97
6.1	Chapter overview .....	97
6.2	Develop moving load FE model .....	98
6.2.1	Kinetic contact enforcement method .....	98
6.2.2	Analytical solutions for moving vehicle .....	102
6.2.3	Feature selection and damage prediction model .....	104

6.3	Verification of FEM model.....	106
6.3.1	Laboratory sensor calibration test.....	106
6.3.2	Field test.....	109
6.4	Result and discussion.....	110
6.4.1	FEM simulation and analytical solution result.....	110
6.4.2	Laboratory test results.....	114
6.4.3	Field test results .....	115
6.4.4	Feature selection results .....	116
6.4.5	Damage prediction results.....	118
CHAPTER 7	CONCLUSION.....	122
REFERENCE.....		127

## LIST OF FIGURES

<i>Figure 2.1. The delamination and cracks. Delamination and cracks are interconnected, with delamination weakening the material and contributing to crack initiation, while cracks can exacerbate delamination by providing pathways for further separation between layers (Reilly et al., 2014).</i> .....	4
<i>Figure 2.2. The example of structural damage (Peris-Sayol et al., 2017).</i> .....	5
<i>Figure 2.3. The ground surface motion of Rayleigh waves (Hishamuddin et al., 2021).</i> .....	6
<i>Figure 2.4. The IE testing configuration (Carino, 2001).</i> .....	7
<i>Figure 2.5. The dataflow of ML (Sarker, 2021).</i> .....	8
<i>Figure 2.6. The structure of ANN (Kulkarni et al., 2017).</i> .....	9
<i>Figure 2.7. The concept of SVM. The distance between features and hyper-plane can calculate the optimized hyper-plane with the maximum margin range. The optimized hyper-plane has a maximum margin range, which defines the best classification border of features. Therefore, the optimized hyper-plane can classify features with high accuracy.</i> .....	10
<i>Figure 2.8. The simple binary decision tree model. Each node displays one attribute, each branch displays one decision based on the entropy threshold calculated in nodes, and each leaf indicates one output. ....</i>	11
<i>Figure 2.9. XGBoost model with tree ensemble (T. Chen &amp; Guestrin, 2016).</i> .....	12
<i>Figure 2.10. The structure of CNN (Hou &amp; Li, 2020)</i> .....	12
<i>Figure 2.11. The structure of GANs (Aggarwal et al., 2021)</i> .....	13
<i>Figure 3.1. The flowchart of Chapter 3. The research pursues two key objectives: firstly, employing the collected data, it explores the realm of multi-input CNN to enhance damage identification precision. This innovative approach involves harnessing the power of advanced ML techniques, specifically the multi-input CNN model, to generate detailed and accurate delamination maps, illuminating the structural integrity of the materials studied. Concurrently, the study focuses into the nuanced interplay between precision and computation time within CNN efficiency.</i> .....	16
<i>Figure 3.2. The artificial delamination slabs, with three different size and two deployed depths cited from (Kang et al., 2022). There are a total of 1280 testing points on the slab, arranged in a grid of 32 by 40, with a spacing of 5 centimeters.</i> .....	18
<i>Figure 3.3. The testing configuration includes the use of non-contact MEMS sensors. The data acquisition system is developed using National Instruments (NI), and LabVIEW software is utilized to save the data.</i> .....	18

Figure 3.4. Impact echo signal samples reveal delamination. Higher energy (1kHz to 5kHz) indicates severe delamination, while lower energy signifies intact structures. Shown in (a) non-delamination and (b) delamination cases..... 19

Figure 3.5. Examples of input images include: (a) frequency-domain signal input (F), (b) multi-input with time-domain and frequency-domain signals (SF), (c) multi-input with frequency-domain signal and STFT (FSTFT), and (d) multi-input with time-domain signal, frequency-domain signal, and STFT (TFSTFT).. 20

Figure 3.6. The image example of resolution study. Ten different resolutions ranging from 10 by 10 to 500 by 500 pixels, encompassing a total of 3000 images is considered. Each of the ten resolution cases consists of 300 images, enabling a comprehensive exploration of the impact of resolution on the trade-off between accuracy and computation time. .... 21

Figure 3.7. The laboratory impact echo test was conducted on a slab featuring artificial delaminations. The color bar represents the accumulated EI in the low-frequency range between 1 kHz to 5 kHz. Higher EI values are depicted in red, indicating highly delaminated areas, while lower EI values appear in blue, representing areas with minimal damage. A noteworthy observation from the test results is the way energy diffusion varies with the depth of delamination. The upper three regions (D1-D3), indicative of shallow delaminations, exhibit markedly higher energy levels than the lower three regions (D4-D6) housing deeper delaminations. .... 23

Figure 3.8. The ML training accuracy and loss curves of four different models: (a) The training accuracy, and (b) training loss. The training accuracy and loss patterns for the four distinct models are evident. Notably, F model stands out with its rapid training process compared to the other three models, swiftly reaching maximum accuracy and minimal loss. This observation underscores that the data within category F is more easily learned by the model, indicating its simplicity in training. .... 25

Figure 3.9. Validation and training curves for F model: (a) accuracy curves, and (b) loss curves. The validation curves require more epochs to reach their peak values, indicating the model's adaptability in unpredictable environments with unknown data..... 27

Figure 3.10. Validation and training curves for SF model: (a) accuracy curves, and (b) loss curves. Compared with Figure 3.9, the validation loss takes more epochs to reach its minimum value, indicating the SF model has higher complexity compared to model F. .... 27

Figure 3.11. Validation and training curves for FSTFT model: (a) accuracy curves, and (b) loss curves. The validation accuracy and loss are meet to threshold value around 40 epochs. .... 28

Figure 3.12. Validation and training curves for TFSTFT model: (a) accuracy curves, and (b) loss curves. The validation accuracy meets to threshold value around 30 epochs. .... 28

Figure 3.13. The confusion matrix of testing accuracy: (a) F model, (b) SF model, (c) FSTFT model, and (d) TFSTFT model. Overall, the FSTFT has the highest testing accuracy to show a better ability to identify delamination. .... 30

Figure 3.14. The CNN prediction delamination map: (a) F model, (b) SF model, (c) FSTFT model, and (d) TFSTFT model. The results presented in (c) demonstrate the most accurate delamination map among all cases. .... 32

Figure 3.15. The computation time and accuracy curves: (a) different number of images study, and (b) different image resolution study. The accuracy and computation time are significantly influenced by the number of images and the image resolution. However, beyond a specific threshold, such as 600x600 resolution, the computation time exponentially increases for a mere 2% improvement in accuracy. .... 34

Figure 3.16. The slope analysis of computation time with number of image study: (a) the curves of computation time and accuracy, and (b) the slope analysis of computation time. For 600x600 resolution image, the number of image threshold for computation time is 200 images. .... 35

Figure 3.17. The slope analysis of accuracy with number of image study: (a) the curves of computation time and accuracy, and (b) the slope analysis of accuracy. For 600x600 resolution image, the number of image threshold for accuracy is 60 images, while for 100x100 resolution image the threshold is 120 images. .... 36

Figure 3.18. The slope analysis of computation time with image resolution study: (a) the curves of computation time and accuracy, and (b) the slope analysis of computation time. For 100 images input, the image resolution threshold for computation time is 600x600. .... 37

Figure 3.19. The slope analysis of accuracy with image resolution study: (a) the curves of computation time and accuracy, and (b) the slope analysis of accuracy. For 100 images input, the image resolution threshold for accuracy is 200x200. .... 37

Figure 3.20. The CNN efficiency study with image resolution: (a) The efficiency curve of number of image study: (b) the curves of computation time and accuracy. The analysis indicates that efficiency values tend to be higher with fewer than 100 images. However, these values need to be carefully considered alongside accuracy and computation time thresholds. .... 38

Figure 3.21. The CNN efficiency study with number of images: (a) The efficiency curve of image resolution study. (b) the curves of computation time and accuracy. A high efficiency is noted with a 50x50 resolution image, but the accuracy remains at 53%. ..... 39

Figure 4.1. The flowchart of Chapter 4. Objective 1 focuses on signal extraction through deep learning, aiming to determine the signal parameters such as time duration and starting time, crucial for accurate damage identification. In Objective 2, these identified parameters are utilized to update the impact echo results. .... 41

Figure 4.2. Field testing photo - (a) scanned area of the bridge, and (b) system and safety control. .... 42

Figure 4.3. The ACES scanning system encompasses impact-echo and GPR technologies cited from (Kang et al., 2022). ..... 42

Figure 4.4. The surface condition highlights prevalent issues such as potholes and damaged surfaces. Potholes and surface damage can disrupt the impactor's contact with the bridge, leading to variations in the generated mechanical waves. .... 43

Figure 4.5. The impact echo algorithm aims to generate a delamination map, but it encounters a significant challenge: processing over 30 million sample data points within a single scan time and the effect of multi-impact. Identifying the precise target peaks using traditional mathematical algorithms becomes daunting due to this vast dataset. To address this challenge, DL algorithms are employed to accurately identify signals. .... 44

Figure 4.6. Schematic representation of the ANN architecture. In the training process, BP computes the error by comparing predicted and actual outputs. It then iteratively adjusts the network's weights and biases from the output layer to the input layer to minimize this error. The input features consist of frequency and time-domain signals, with the output focused on damage identification. .... 45

Figure 4.7. The CNN model comprises a convolution layer for feature extraction and a pooling layer to reduce dimensions, retaining essential features. Extracted features are connected to a flatten layer and a neural network for training and testing purposes. .... 47

Figure 4.8. The CNN inputs for signal extraction, examining various durations (Dt<sub>i</sub>) of time-domain impact echo signals. A total of 10 different durations, ranging from 0.2 milliseconds to 2 milliseconds, are analyzed. Each dataset (e.g., Dt<sub>1</sub> dataset) comprises 300 images obtained from field tests. .... 48

Figure 4.9. The study explores CNN inputs for signal extraction, varying the starting time (St<sub>i</sub>) after the zero-crossing point. A total of eight different signals are considered, ranging from the immediate zero-crossing point to 0.7 milliseconds after it. .... 48

Figure 4.10. The confusion matrix illustrates the identification of different duration signals, categorizing them into non-delamination, delamination, and insignificant signal classes. (a)  $Dt1 = 0.2ms$ , (b)  $Dt2 = 0.4ms$ , (c)  $Dt4 = 0.8ms$ , and (d)  $Dt5 = 1ms$ . ..... 50

Figure 4.11. The outcomes of the varied signal duration study are evident. Specifically, when duration is  $1ms$ , a significantly higher prediction accuracy is achieved. This is because within the  $1ms$  window, crucial impact signals reflecting structural interior damage are effectively captured. In contrast, when duration= $2ms$  is considered, the accuracy drops significantly. This decrease is attributed to the broader duration including a multitude of insignificant signals, such as tail resonance or noise, resulting in decreased accuracy. .... 51

Figure 4.12. The interpretation of signal extraction accuracy with varying duration cases reveals distinct patterns.  $Dt1$  and  $Dt2$ , characterized by limited signal information, result in lower CNN identification accuracy. In contrast,  $Dt5$  exhibits a precise signal group, leading to higher CNN accuracy. However,  $Dt6$  includes more signals, including tail insignificant signals, which adversely affect accuracy. .... 52

Figure 4.13. The outcomes of various signal starting times  $Sti$  across three different durations, including lab test delamination data, exhibit a consistent pattern. In all six scenarios, the highest accuracy is achieved when  $St2=0.1m$  or  $St3=0.2ms$ . This heightened accuracy can be attributed to the exclusion of the initial surface wave, a component that significantly influences signal behavior. .... 53

Figure 4.14. The explanation of CNN accuracy shown in Figure 4.13. At  $St1=0ms$ , both surface wave and target wave are present, but the surface wave offers similar information across all signals, making it less significant for CNN training. When  $St2=0.1ms$  and  $St3=0.2ms$ , the elimination of the surface wave focuses CNN on the target signal while introducing more tail noise. These signals are effectively identified due to their richer information. However, at  $St4=0.3ms$ , the exclusion of the target signal results in decreased accuracy. .... 54

Figure 4.15. The ANN accuracy with four features from Eq. 4.4 to Eq. 4.7. The observed pattern correlates with the results from the starting time study in Figure 4.14, showing that  $St2=0.1ms$  to  $St4=0.3ms$  provides the highest CNN accuracy. Conversely, for ANN, accuracy sharply decreases at  $St5=0.4ms$  due to the signal being influenced by tail-end noise, consistent with the observations in Figure 4.14. .... 55

Figure 4.16. The delamination maps of scanning path 1. (a) Original algorithm, and (b) CNN-suggestion signal parameter from signal extraction study (e.g.,  $Dt5=1ms$  and  $St2=0.1ms$ ). Consider point C as another example: Figure 4.16 (a) indicates a non-delamination case, whereas Figure 4.16 (b) suggests slight damage in that area. This lower energy could result from the impactor failing to strike the surface

*accurately at the correct angle, possibly due to surface conditions like potholes where the impactor may become stuck and vibrate..... 57*

*Figure 4.17. The verification of Figure 4.16. The samples in Figure 4.16 depict the time-domain and frequency-domain signals for different cases: (a) Region A, (b) Region B, (c) Region C, (d) Region D, (e) Region E, and (f) Region F. Consider sample C as an illustration; despite the relatively low energy, a distinct low-frequency peak indicates a delamination case. Utilizing the CNN-recommended signal parameters depicted in Figure 4.16 (b) results in accurate identification of the delamination in this sample. .... 58*

*Figure 5.1. Overview of research steps. In this study, a sequential research approach was conducted, divided into several tasks. The process began with a comprehensive review of wave response theory, establishing a foundation. Subsequently, the focus shifted to the mathematical relationship between wave response in HM, and WRV patterns were analyzed. Due to limited research on WRV in IHM like concrete, a FE model was developed, and IHM laboratory tests were conducted, confirming the FE model's validity with varied crack geometries. Machine learning techniques were applied to enhance the database and predict crack depth based on understood WRV patterns. Additionally, exploration of GANs was undertaken to generate more WRV data. The study culminated in a field test, evaluating the accuracy of machine learning in identifying damages..... 60*

*Figure 5.2. The wave scattering mechanism involves the incident wave energy splitting into backward and forward scattering upon encountering an internal crack. In this context, "r" and "θ" are polar coordinate components, "a" represents the depth-to-crack (distance between the surface and the top of the crack tip), and "b" signifies the crack size. .... 61*

*Figure 5.3. This FE model example includes listening nodes N1 and N2. N1 captures incident and reflected waves, while N2 captures forward waves. Ten layers of dampers are incorporated to minimize boundary reflections..... 65*

*Figure 5.4. An example of the incident wave and forward wave in the laboratory test is shown in Figure 5. The incident wave (blue solid line) has a higher amplitude, while the forward wave (red solid line) has a lower amplitude due to the presence of a crack..... 66*

*Figure 5.5. The reflection validation from FEM. The model with the infinity element (blue) and without the infinity element (orange) experiences less reflection in the front first peak. The WRV is calculated from the first peak of the incident wave and forward scattering wave..... 67*

Figure 5.6. Different aggregate size study. Two study groups were defined: L1-L6, consisting of larger aggregates exclusively, and S1-S6, comprising smaller aggregates with a random distribution. These cases maintained a similar aggregate portion ratio..... 69

Figure 5.7. Ten patterns were created using random aggregate sizes and distributions, offering diverse designs for the study..... 69

Figure 5.8. Artificial internal crack samples. (a) the specimen container with dimensions 8x8x4 inches; (b) the fresh concrete mixture; (c) a model with a constant depth-to-crack of 1 inch; (d) a top-view of diagonal depth-to-crack ranging from 0.5 inch to 2 inches; and (e) the DC motor used in the testing setup. These components were essential in simulating and studying crack patterns within IHM..... 70

Figure 5.9. The laboratory test setup involved the use of a DC motor to generate mechanical waves, while MEMS sensors were employed to capture the incident wave, reflected wave, and forward wave. .... 71

Figure 5.10. The concept of GANs model. A generator network (G) and a discriminator network (D) operate in tandem, with the generator taking a random noise vector (z) as input and generating synthetic data (x'). The generator and discriminator are trained adversarially: the generator strives to create synthetic data resembling real data, while the discriminator aims to differentiate between real and synthetic data. The input and generated data in this study is WRV pattern..... 73

Figure 5.11. The CNN model comprises a convolution layer for feature extraction and a pooling layer to reduce dimensions, retaining essential features. Extracted features are connected to a flatten layer and a neural network for training and testing purposes..... 74

Figure 5.12. The field test is conducted on an airport runway, where the bottom layer consists of a concrete slab with potential internal cracks. This scenario raises concerns about the presence of hidden defects within the structure. .... 75

Figure 5.13. The ACES system configuration cited from (Kang et al., 2022). .... 75

Figure 5.14. The comparison between experiment and FE simulation. (a) incident wave, and (b) forward wave. .... 77

Figure 5.15. Normalized WRV of FE simulation and analytical solutions in HM. On the y-axis, the normalized forward wave is depicted, representing the ratio of the first peak by dividing the forward wave  $E_{fw}$  amplitude by the incident wave amplitude  $E_{in}$ . The x-axis represents the product of "a" and wavelength. (a) the analytical solution results with different a. (b) the FE simulation results with different a. (c)-(f) shows the comparison between FE simulation and analytical solution with different a. The shallower crack case ( $a=12.7mm$ ) exhibits a peak at a lower impulse frequency, while the deeper crack

*(a=50.8mm) shows the WRV peak at a higher frequency. This variation occurs because different impulse frequencies have varying abilities to propagate through deeper mediums, influenced by their wave speed and wavelength. .... 79*

*Figure 5.16. The example of sensing data comparison between experiment and simulation. (a) the incident wave, and (b) the forward wave..... 80*

*Figure 5.17. The WRV pattern with different aggregate size with random distribution (a) the WRV with two aggregate sizes, (b) the comparison between experiment and P1 simulation (c) the boxplot of large aggregate group and (d) the boxplot of small aggregate group. It reveals a slightly lower amplitude for the large aggregate group (L1-L6) compared to the small aggregate group (S1-S6). This difference may stem from larger aggregates dissipating more energy, causing the wave to be influenced by the aggregate size before passing through completely. .... 82*

*Figure 5.18. The WRV comparison between HM and IHM. On the y-axis, the normalized forward wave is depicted, representing the ratio of the first peak by dividing the forward wave Efw amplitude by the incident wave amplitude Ein. The x-axis represents the product of "a" and wavelength. (a) a=12.7mm, (b) a=25.4mm, (c) a=38.1mm, and (d) a=50.8mm. .... 83*

*Figure 5.19. Ten pattern WRV results with 8 different "a" value. On the y-axis, the normalized forward wave is depicted, representing the ratio of the first peak by dividing the forward wave Efw amplitude by the incident wave amplitude Ein. The x-axis represents the product of "a" and wavelength. (a) P1, (b) P2, (c) P3, (d) P4, (e) P5, (f) P6, (g) P7, (h) P8, (i) P9 and (j) P10. .... 86*

*Figure 5.20. The ten-pattern comparison by different "a" value. On the y-axis, the normalized forward wave is depicted, representing the ratio of the first peak by dividing the forward wave Efw amplitude by the incident wave amplitude Ein. The x-axis represents the product of "a" and wavelength. (a) a=6.3mm, (b) a=12.7mm, (c) a=18.9mm, (d) a=25.4mm, (e) a=31.5mm, (f) a=38.1mm, (g) a=44.1mm, and (h) a=50.8mm. Within the same crack, different "a" cases exhibit similar WRV waveforms but varying amplitudes. The WRV curves are influenced by the random distribution and aggregate size. .... 87*

*Figure 5.21. The boxplot of ten pattern comparison by different "a" value. (a) a=6.3mm, (b) a=12.7mm, (c) a=18.9mm, (d) a=25.4mm, (e) a=31.5mm, (f) a=38.1mm, (g) a=44.1mm, and (h) a=50.8mm. The red plus marks denote values falling outside the standard deviation range, indicating significant differences in forward amplitude from other cases. .... 89*

*Figure 5.22. Example of training input of GANs. There are 8 different groups depends on "a" value with 10 patterns. .... 90*

Figure 5.23. The output of GANs. In each “a” value, there are 200 images, total has 1600 images. .... 90

Figure 5.24. The ALE plot: (a) the number of small size aggregate, (b) the number of large size aggregate, (c) the minimum aggregate diameter, and (d) the maximum aggregate diameter. According to the findings, the number of small aggregates exhibits a decreasing effect, while the number of large aggregates shows an increasing effect on the model's output. This observation aligns with the conclusions as depicted in Figure 5.19(a), where larger aggregates exert a more substantial impact (e.g., reducing the WRV amplitude) on the WRV pattern compared to smaller ones. Figure 5.24(c) and (d) portray the ALE results for the minimum and maximum aggregate diameters, respectively..... 92

Figure 5.25. The confusion matrix of CNN model: (a) the CNN model without using fake samples, (b) with 40 fake sample, (c) with 120 fake samples, (d) with 200 fake samples, (e) with 400 fake samples and (f) with 1080 fake samples. The increase in accuracy with the number of fake samples underscores the effectiveness of WRV data in identifying wave propagation patterns. .... 94

Figure 5.26. The CNN accuracy curves vary with different numbers of fake samples. The curves show variations with different numbers of fake samples. The accuracy is 76% without fake samples, while using 400 fake samples can boost accuracy to 86%, reaching a saturated level..... 95

Figure 5.27. The delamination results of impact echo. The yellow solid line denotes the scanning path. . 96

Figure 5.28. The comparison between CNN crack length prediction and the crack estimation from Rayleigh wave(Kang et al., 2021). (a) the area-A prediction result. The field test estimated crack depth is 0.8 inch, while the CNN prediction using WRV is 0.6 inch. (b) the area-B prediction result. The field test estimated crack depth is 0.4 inch, while the CNN prediction using WRV is 0.7 inch. .... 96

Figure 6.1. A flowchart depicting the two stages of developing a more accurate structural prediction model. Stage 1 involves data collection from FE modeling, and Stage 2 shows the implementation of AI application for feature selection and damage prediction. .... 98

Figure 6.2. The model of static load design. The design of the static load FE model with three different load locations (LL 1-3) and three different crack locations (CL1-3). The dimensions of the model are 1 m × 5 m × 0.02 m, and it features thirty sensing points (S1-S30) positioned underneath. The model simulates mild, moderate, and severe cracks with 10%, 25%, and 45% of the missing elements, respectively, in the side section area..... 99

Figure 6.3. The one-way FE model moving load applied with KCE..... 101

Figure 6.4. The 3-D model of surface-to-surface contact in KCE, where  $F_s$  and  $F_M$  denote the nodal force on both the slave surface and master surface;  $k$  is the contact stiffness between the master and slave elements, and  $F_N$  represents an equilibrium of forces. .... 101

Figure 6.5. The simple model considered passing vehicle and bridge, where  $L$  is the length of the bridge,  $E, I$ , and  $\omega_b$  are the bridge modulus of elasticity, bridge second moment of area and bridge first natural frequency, while  $m_v$ ,  $k_v$ , and  $\omega_v = k_v/m_v$  are the mass of the vehicle, vehicle first natural frequency, and vehicle stiffness, separately. .... 103

Figure 6.6. Testing configuration of the sensor performance laboratory test. .... 107

Figure 6.7. Testing sensors and DAQ. (a) LVDT, (b) linear potentiometer position sensor, (c) draw-wire displacement linear sensor, (d) VPG DAQ, and (e) NI DAQ. .... 108

Figure 6.8. A testing configuration with concrete beam in full-scale laboratory test. .... 109

Figure 6.9. The location of bridge ..... 110

Figure 6.10. The bridge displacement of (a) the FE simulation and (b) the analytical solution. .... 111

Figure 6.11. The comparison of density results between FE simulation and analytical solution. The plot shows a similar trend among different density cases, verifying the accuracy of the FE model. The displacement in the simulation results starts at a non-zero value (when  $D = 100 \text{ kgm}^3$ ) due to the higher density of the moving element ( $700 \text{ kgm}^3$ ) compared to the bridge density ( $100 \text{ kgm}^3$ ). .... 112

Figure 6.12. The comparison of depth results between simulation and analytical solutions. The results show the trend where the lower depth is shown as (e) a higher displacement. However, the waveform is substantially affected by high frequency reflected waves from (c) the bottom sides caused by (d) a thicker dimension. .... 113

Figure 6.13. The loss (%) of maximum displacement. The value between FE simulation and analytical solution demonstrates similar energy loss, which means the developed FE simulation from KCE and analytical solution has a similar pattern between moving vehicles. (a) elastic modulus study, and (b) density study. .... 114

Figure 6.14. The Laboratory test results: (a) the displacement result of the sensor calibration laboratory test. The draw-wire displacement linear sensor provides low-resolution results and, therefore, will not be used in the field test. (b) The comparison between full-scale laboratory tests and KCE simulation. .... 115

Figure 6.15. The example of field test results: (a) with truck 10mph, and (b) with truck 40mph. .... 116

*Figure 6.16. The comparison between KCE FE model and field test: (a) with van 25 mph and (b) with sedan 25 mph ..... 116*

*Figure 6.17. Examples of structural acceleration responses: (a) under static loads and (b) under moving loads, presenting more complex wave response patterns. .... 117*

*Figure 6.18. The importance ranking is calculated from the random forest with Gini impurity. The first capital alphabet presents what data used for indicators. “A” means the indicator calculated from acceleration; “D” means the indicator calculated from displacement and “S” means the indicator calculated from stress. The acceleration provides the highest importance feature for ML prediction results. The features with Gini importance over 0.3 are used in ML training data for damage prediction..... 118*

*Figure 6.19. The damage prediction results were derived by using static load FE simulation results from (a) BP, (b) SVM, (c) decision tree and (d) XGBoost algorithms. .... 120*

*Figure 6.20. The damage prediction results after processing the moving load FE simulation results from (a) BP, (b) an SVM, (c) a decision tree, and (d) the XGBoost algorithm ..... 121*

## LIST OF TABLES

<i>Table 2.1. The example of a confusion matrix</i> .....	14
<i>Table 3.1. The total number of samples correct from laboratory test.</i> .....	20
<i>Table 3.2. The accuracy of each damage of four models. Unit (%)</i> .....	31
<i>Table 5.1. Five different integral methods used in WRV analytical solution</i> .....	63
<i>Table 5.2. The damper coefficient is denoted by Damper 1 for the inner one attached to the specimen, while Damper 10 refers to the outer damper</i> .....	65
<i>Table 5.3. The aggregate detail of ten patterns</i> .....	68
<i>Table 5.4. The training sample and testing sample of CNN with different number of GANs samples.</i> .....	74
<i>Table 6.1. The locations (x-axis) in Figure 2 of the static load model's crack load (CL) and (LL) load. Unit: meter</i> .....	100
<i>Table 6.2. The setting for one-way moving load detail and material properties</i> .....	102
<i>Table 6.3. The properties of bridges and moving vehicle</i> .....	104
<i>Table 6.4. Detailed variables of parameters between FE modeling and analytical solutions</i> .....	104
<i>Table 6.5. Statistical indicator and FE modeling information used for data features (Finotti et al., 2019)</i> .....	105
<i>Table 6.6. Four commonly used impurities in random forest</i> .....	105
<i>Table 6.7. The final setting of each machine learning model</i> .....	106

# CHAPTER 1 INTRODUCTION

## 1.1 Problem Statement

The maintenance of transportation infrastructures (e.g., bridges and pavements) has become a major concern for safety and economic loss with many factors such as increasing traffic volumes, deterioration. In particular, deterioration (e.g., external and internal damages) of infrastructures can significantly impact infrastructure service life and often require extensive repairs or replacements. For improving service life, a structure survey of its condition is required for transportation agencies not only to ensure roadway safety but also to provide proper time for appropriate preservation and rehabilitation treatments.

In practical field tests, however, the acquired signals often deviate significantly from the clean and controlled signals obtained in laboratory settings. Field signals exhibit inherent complexity attributed to various environmental factors, including ambient noise, road surface conditions, and slope angles (Baggens & Ryden, 2015; Carino, 2001). These external interferences, ranging from varying road surface conditions to fluctuations in slope angles, exert a substantial influence on the signals obtained during field evaluations (Kang et al., 2021; Yoon et al., 2022). These interferences introduce complexities that can distort the raw data, potentially leading to inaccurate estimations and erroneous assessments of structural conditions (Carino, 2001). The daunting challenge lies in mitigating these interferences, as their presence significantly compromises the reliability of data collected during field evaluations (Gucunski et al., 2008; Liu et al., 2023). Besides, the material of field is varied combined with inhomogeneous medium (IHM), which affect testing results. Despite the progress in understanding WRV in both homogeneous medium (HM) and IHM, notable research gaps persist. While extensive studies in HM have focused on the behavior of electromagnetic waves in isotropic environments, particularly in scattering phenomena, there remains a lack of comprehensive exploration of wave response variation (WRV) patterns, especially in situations where media properties may exhibit variations. Moreover, the understanding of WRV in IHM, such as concrete, remains limited due to challenges posed by random aggregate size and distribution. The

complexities introduced by factors like material properties, wave frequency, crack size and shape, and the surrounding environment hinder the development of analytical solutions.

## **1.2 Objectives and Approach**

The primary goal is to evaluate, identify, and predict structural damage in infrastructure using various types of machine learning (ML) techniques and diverse study cases and input data. The initial objective is to enhance impact-echo (IE) signal post-processing through an ML-based signal extraction approach. Improvements in IE delamination results are achieved by utilizing IE signal images to identify under or overrated data points due to field test environmental conditions. The efficiency of ML is further investigated with varying image resolutions and numbers of images to understand the impact of input parameters on ML computation time and model accuracy. Chapter 3 concentrates on IE laboratory test results with an artificial delamination slab, employing convolutional neural network (CNN) to identify damages and presenting a study on ML efficiency. Chapter 4 outlines the enhancement of IE field test results using CNN to improve the accuracy of signal extraction. The second objective is to explore WRV patterns in HM and IHM with random aggregates and random distribution, considering eight different internal crack geometries. Chapter 5 provides a description of the developed finite element (FE) model with infinite elements for both HM and IHM. The study also explores into the prediction of internal crack geometry using WRV. The third objective is to predict structural damage using structural performance data collected from the bridge weigh-in-motion (BWIM) signal. FE simulations are carried out with the kinematic contact enforcement (KCE) method to simulate the effect of a moving vehicle load. Chapter 6 provides a detailed account of the development of KCE, outlines the structural features, and presents a comparison between four types of ML algorithms. By conducting these studies, structural damage can be assessed and predicted based on either the IE signal or the structural performance, providing insights and timelines for better preventive structural maintenance.

## **CHAPTER 2 BACKGROUND**

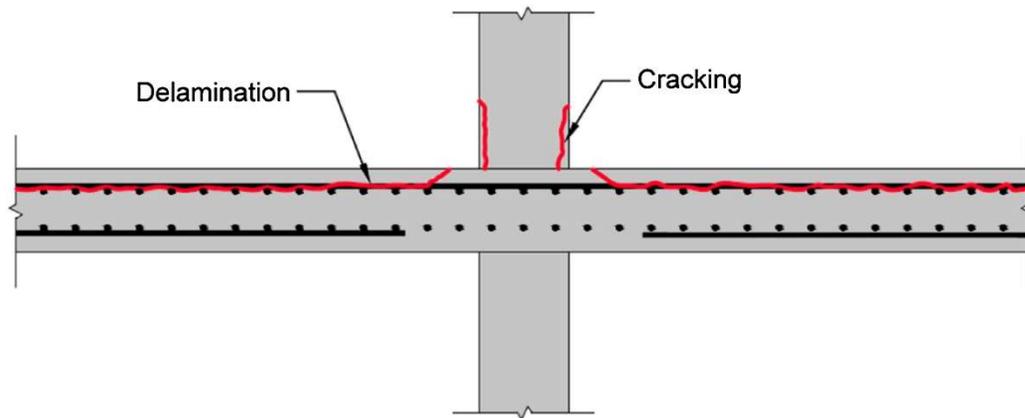
### **2.1 Overview**

This chapter furnishes the fundamental background essential for understanding the studies conducted in this dissertation. It encompasses the types of bridge defects that are the focus of our research, non-destructive testing (NDT) technology, and the machine learning algorithms applied in the dissertation.

### **2.2 Bridge Defects**

#### **2.2.1 *Delamination and cracks***

Delamination and cracks are structural issues that commonly occur in various materials, including concrete and composites. Delamination refers to the separation or detachment of layers within a material, often manifesting as the peeling or flaking of one layer from another. This phenomenon weakens the structural integrity, compromises load-bearing capacity, and can be caused by factors such as moisture ingress, freeze-thaw cycles, or inadequate bonding during construction. On the other hand, cracks are linear fissures or fractures that develop in a material due to stress, temperature fluctuations, or structural movements. Delamination and cracks share a relationship as delamination can contribute to the initiation and propagation of cracks. For instance, the presence of delamination weakens the material, making it more susceptible to crack formation under external forces or environmental conditions. In turn, the development of cracks may exacerbate delamination by providing pathways for further separation between layers.



**Figure 2.1. The delamination and cracks. Delamination and cracks are interconnected, with delamination weakening the material and contributing to crack initiation, while cracks can exacerbate delamination by providing pathways for further separation between layers (Reilly et al., 2014).**

### **2.2.2 Structural Damage**

Structural damage in bridges encompasses a range of issues that compromise the integrity and safety of the infrastructure. This can include phenomena such as delamination, where layers of the bridge material separate, and the formation of cracks due to stress, temperature fluctuations, or other environmental factors. Additionally, structural damage may result from the wear and deterioration of materials over time, exposure to harsh weather conditions, or inadequate construction practices. Identifying and addressing these problems is essential to prevent further deterioration, ensure the structural stability of the bridge, and ultimately guarantee the safety of both pedestrians and vehicular traffic. Various NDT technologies and ML algorithms are often employed to assess and predict the extent of structural damage, aiding in timely and effective maintenance and repair strategies for bridge infrastructure.

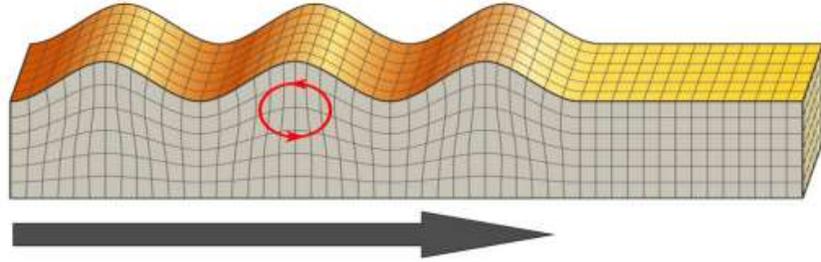


**Figure 2.2.** *The example of structural damage* (Peris-Sayol et al., 2017)

## 2.3 Nondestructive Testing

### 2.3.1 *Wave propagation*

Wave propagation refers to the transmission and spread of waves through a medium or space. Waves, which can be of various types such as mechanical, electromagnetic, or acoustic, carry energy and information as they travel. The behavior of wave propagation is governed by the properties of the medium through which the waves travel, including its density, elasticity, and other material characteristics. Waves can exhibit phenomena such as reflection, refraction, diffraction, and interference, depending on the nature of the medium and the specific conditions of the propagation. Understanding wave propagation is crucial in diverse fields, from physics and engineering to telecommunications and seismology, where it plays a fundamental role in describing how disturbances, signals, or energy propagate through different mediums over time and space.



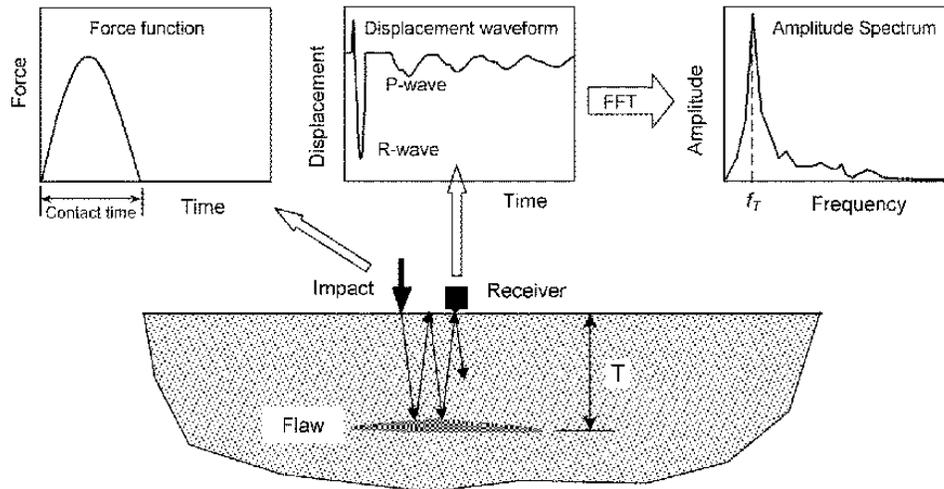
**Figure 2.3. The ground surface motion of Rayleigh waves** (Hishamuddin et al., 2021).

### 2.3.2 *Wave attenuation*

Wave attenuation refers to the gradual decrease in the intensity or amplitude of a wave as it propagates through a medium. This phenomenon occurs due to the dissipation of energy through various mechanisms, such as absorption, scattering, or conversion into other forms of energy. In the context of different wave types, including acoustic, electromagnetic, or seismic waves, attenuation manifests as a reduction in the wave's strength over distance. The extent of attenuation is influenced by the properties of the medium through which the wave travels, such as its composition, density, and viscosity.

### 2.3.3 *Impact-echo*

IE is an NDT technique used to assess the condition of structures by analyzing the acoustic response generated when a mechanical impact, typically in the form of a transient force or stress wave, is applied to the material. This method relies on the detection and analysis of echoes produced by internal structural features, such as delaminations, voids, or other anomalies, as the impact-induced waves interact with these features and reflect back to the surface. By analyzing the time delay and characteristics of these echoes, practitioners can infer valuable information about the structural integrity and identify potential defects within the material. Impact echo testing is widely employed in civil engineering for assessing concrete structures, providing insights into the internal condition without the need for invasive measures.

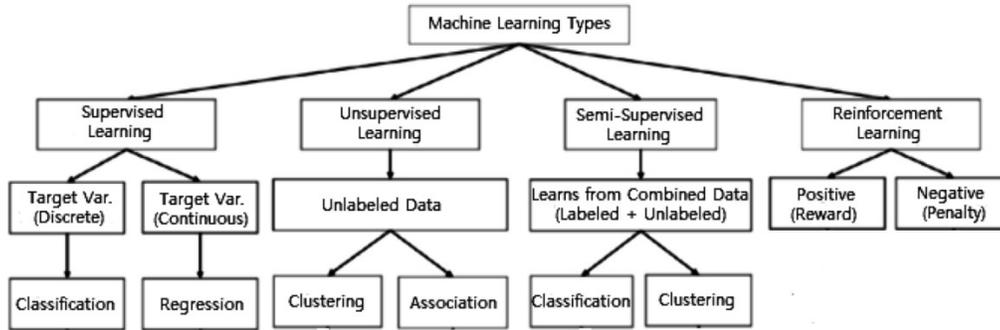


*Figure 2.4. The IE testing configuration (Carino, 2001).*

## 2.4 Signal Identification & Damage Prediction

### 2.4.1 *Machine learning*

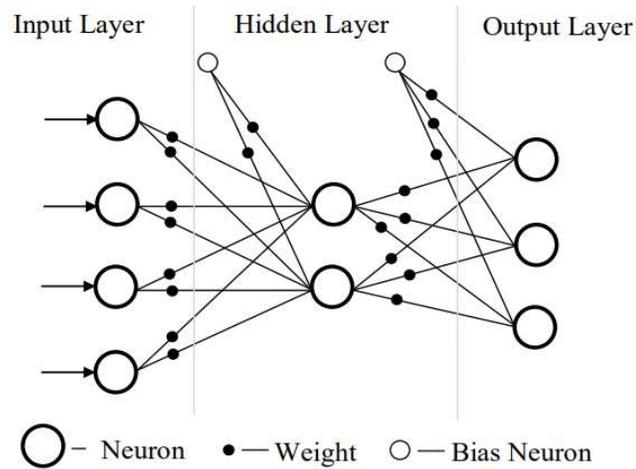
ML is a branch of artificial intelligence (AI) that focuses on developing algorithms and statistical models enabling computer systems to perform tasks without explicit programming. It involves the use of data-driven techniques, where computers learn patterns and make predictions or decisions based on input data. Machine learning encompasses various approaches, including supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, algorithms are trained on labeled datasets to make predictions or classifications, while unsupervised learning involves extracting patterns from unlabeled data.



*Figure 2.5. The dataflow of ML (Sarker, 2021)*

### 2.4.2 Artificial neural network

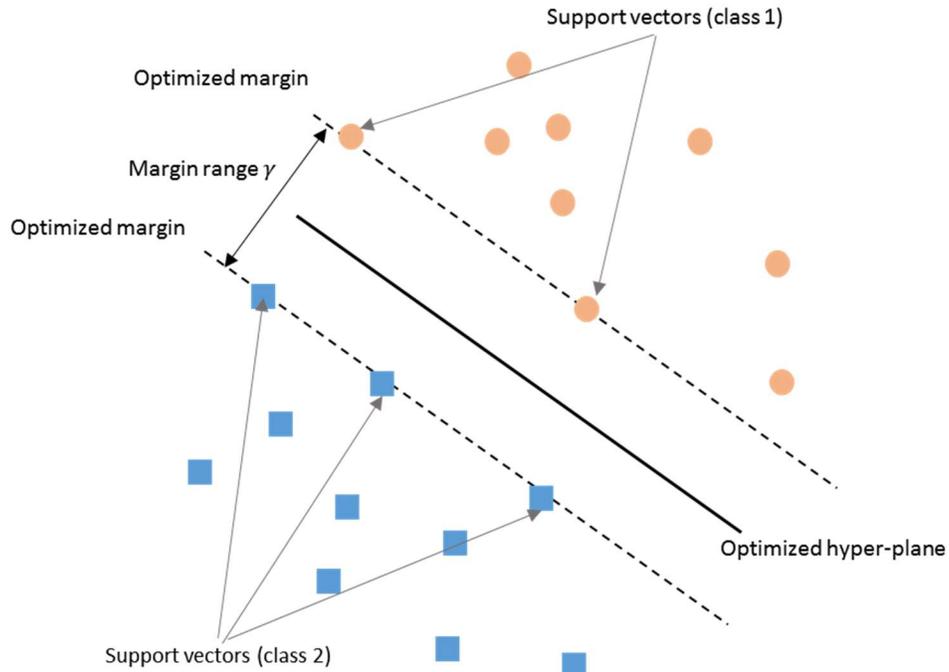
Artificial neural network (ANN) is a computational model inspired by the structure and functioning of the human brain's neural networks. Comprising interconnected nodes, or artificial neurons, organized into layers, ANNs process information through a series of weighted connections that adjust during training. Input data is fed into the input layer, undergoes transformations through hidden layers using learned weights, and produces an output in the final layer. The strength of ANNs lies in their ability to learn complex patterns and relationships from data, making them valuable for tasks like classification, regression, and pattern recognition.



**Figure 2.6.** *The structure of ANN* (Kulkarni et al., 2017).

### 2.4.3 *Support vector machine*

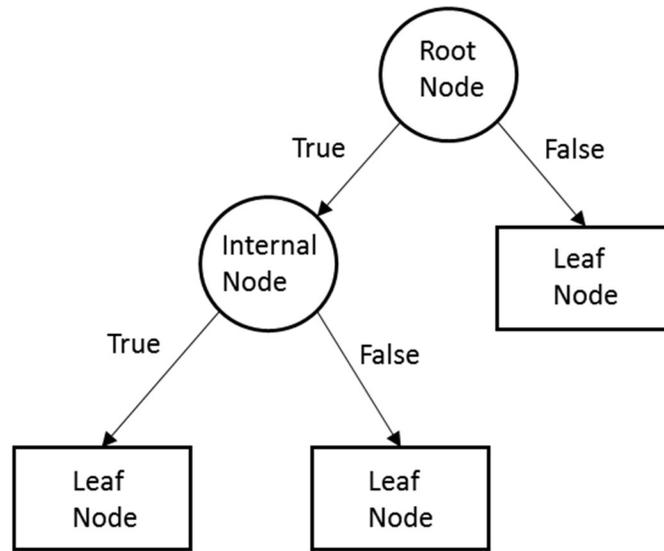
Support vector machine (SVM) is an algorithm used to classify or regress for features. Similar to the regression curve, the border called hyperplane divides two different classes of data. SVM is used to find the optimized hyperplane with the largest margin range. The goal is to find an optimized hyper-plane that has a maximum margin range, the distance between any point in space  $x$  to hyper-plane. The optimized hyper-plane, which has the largest margin range, defines the best classification border of data, as shown in Figure 2.7.



**Figure 2.7.** *The concept of SVM. The distance between features and hyper-plane can calculate the optimized hyper-plane with the maximum margin range. The optimized hyper-plane has a maximum margin range, which defines the best classification border of features. Therefore, the optimized hyper-plane can classify features with high accuracy.*

#### 2.4.4 Decision tree

Decision tree (DT) is an optimized regression method presenting by tree structure where each node displays one attribute, each branch displays one decision, and each leaf indicates one outcome. The algorithm makes the decision in each node depends on the threshold, as known as information entropy. Based on the attribute of features and threshold calculated from entropy, the features will be classified into different branches and leaves in different classes to give the optimized classification: the lower the entropy value, the more similar the features in the dataset.



*Figure 2.8. The simple binary decision tree model. Each node displays one attribute, each branch displays one decision based on the entropy threshold calculated in nodes, and each leaf indicates one output.*

#### 2.4.5 *Extreme gradient boosting*

XGBoost is known as extreme gradient boosting, which combined the benefits of tree boosting and tree assemble. Tree boosting, known as gradient boosting, is used to find the best answer with the lowest error by using the loss function (e.g., mean square error). The tree assembles mean considering the results from all decision trees to vote (or sum) the output with the highest score (e.g., weight)—the simple model of XGBoost with tree ensemble concept as shown in Figure 2.9 (T. Chen & Guestrin, 2016). Similar to a decision tree, in the tree ensemble, each tree has its own threshold (e.g., entropy) in nodes to classify data, and the final prediction is based on the sum of predictions from each tree. In XGBoost, the first and second-order derivative of Taylor polynomial is used for finding the best prediction results.

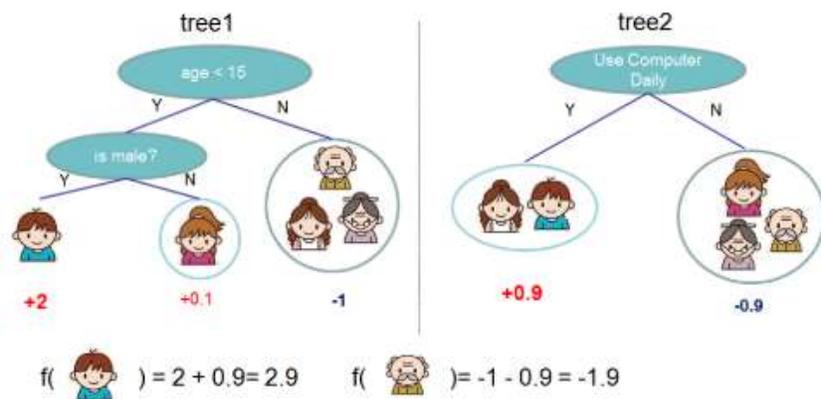


Figure 2.9. XGBoost model with tree ensemble (T. Chen & Guestrin, 2016).

### 2.4.6 Convolution neural network

Convolutional neural network (CNN) is a specialized type of artificial neural network designed for processing and analyzing structured grid data, such as images. CNNs excel at image recognition tasks by leveraging convolutional layers that apply filters to input data, capturing spatial hierarchies and features. These layers are followed by pooling layers, which reduce spatial dimensions while retaining essential information. CNNs are characterized by their ability to automatically learn hierarchical representations of features from raw data, enabling effective pattern recognition. This architecture is widely used in computer vision applications, including image classification, object detection, and facial recognition.

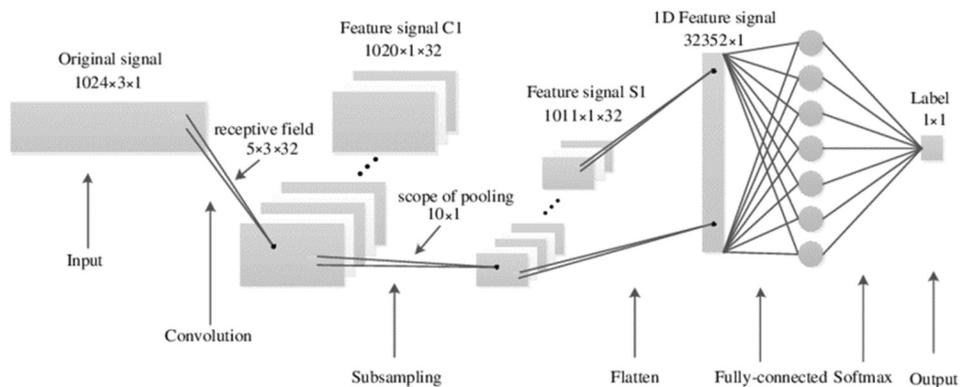


Figure 2.10. The structure of CNN (Hou & Li, 2020)

### 2.4.7 Accumulated local effects

Accumulated local effects (ALE) in machine learning refers to a method used for model interpretation and understanding the impact of input features on a model's predictions. ALE plots visualize the average effect of a single feature by accumulating the individual effects at different values of that feature while keeping other features constant. It helps to uncover non-linear relationships and interactions within the data, providing insights into how changes in a specific feature influence the model's output.

### 2.4.8 Generative adversarial network

Generative adversarial network (GAN) consists of two neural networks, a generator and a discriminator, engaged in a adversarial training process. The generator aims to produce realistic data samples from random noise, while the discriminator attempts to differentiate between real and generated samples. The two networks are trained iteratively, with the generator trying to improve its ability to create realistic data, and the discriminator refining its ability to distinguish between real and generated data. This adversarial interplay results in the generator creating increasingly authentic-looking data.

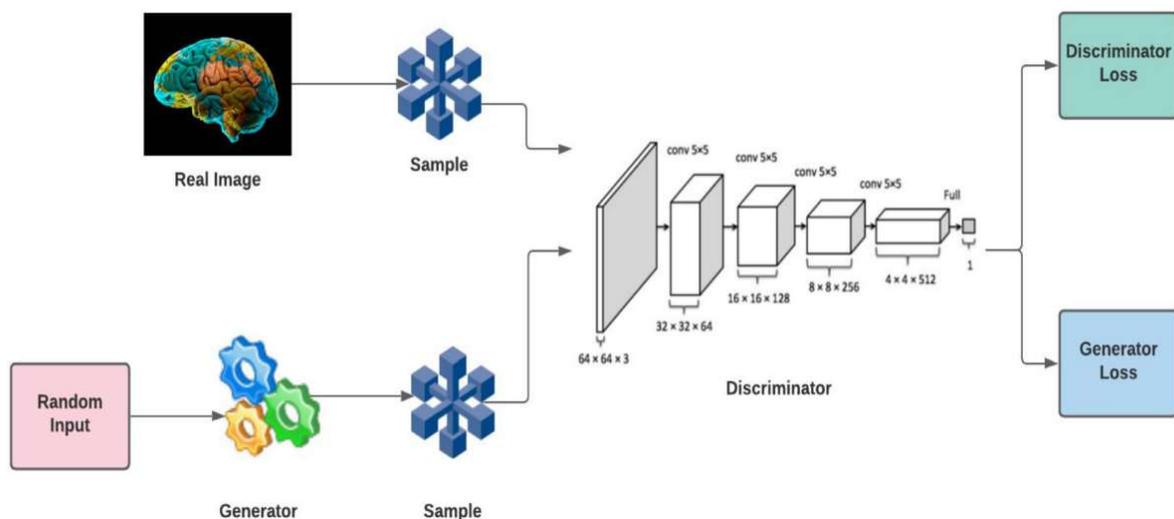


Figure 2.11. The structure of GANs (Aggarwal et al., 2021)

### 2.4.9 Confusion matrix

A confusion matrix serves as a crucial visualization tool in ML, essential for evaluating the performance of a classification algorithm. Its significance arises when the model's predictions are juxtaposed with the actual outcomes, offering a comprehensive breakdown of the model's accuracy and its capability to accurately classify different classes. This matrix provides detailed insights, enabling a thorough assessment of the model's predictive accuracy and its effectiveness in classifying distinct categories.

**Table 2.1. The example of a confusion matrix**

	Actual Positive (1)	Actual Negative (0)
Predicted Positive (1)	True Positive (TP)	False Positive (FP)
Predicted Negative (0)	False Negative (FN)	True Negative (TN)

True positive (TP) is the model correctly predicted positive outcome while true negative (TN) is the model correctly predicted negative outcome; false positive (FP) is the model incorrectly predicted positive outcome, and false negative (FN) is the model incorrectly predicted negative outcome. The model accuracy can be expressed by:

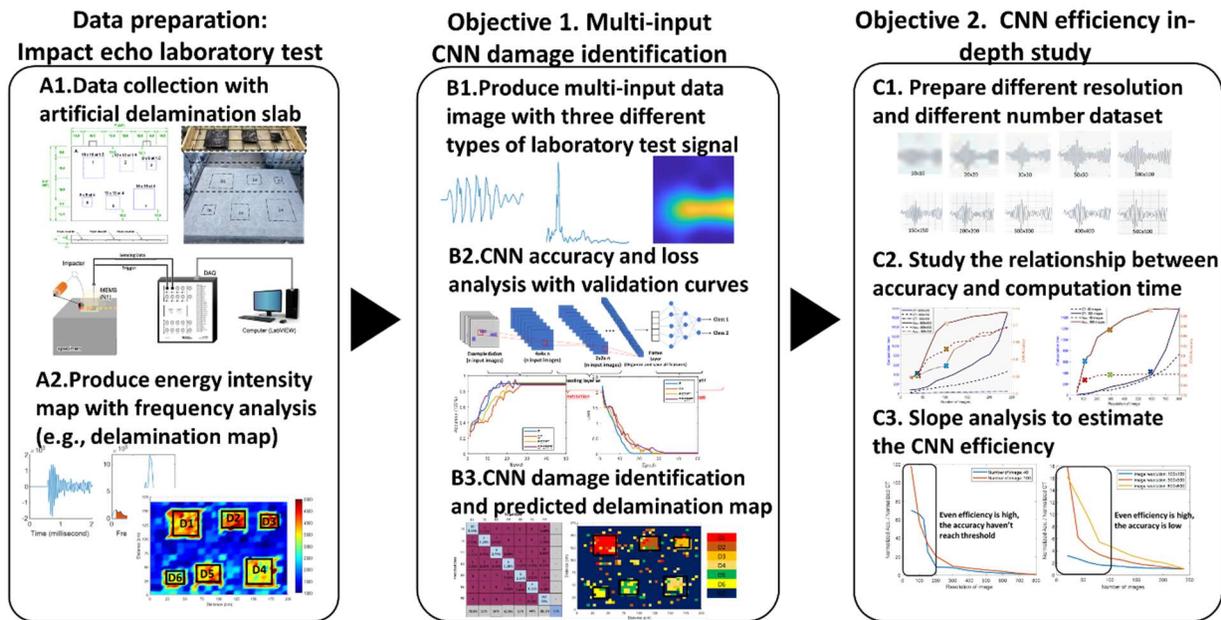
$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2.1)$$

## CHAPTER 3 IE DAMAGE IDENTIFICATION & CNN EFFICIENCY

### 3.1 Chapter overview

This chapter encompasses two distinct objectives. The first objective revolves around exploring the capabilities of CNN in identifying damage within impact echo signals of varying types, including time-domain signals, frequency-domain signals, and STFT representations. The study examines the training, validation, and testing accuracy of the damage identification model, aiming to discern the network's proficiency in accurately categorizing different forms of structural impairment within these diverse signal formats. The second objective is centered on evaluating the efficiency of ML models, specifically focusing on both accuracy and computation time. This aspect explores the delicate balance between precision and computational speed, emphasizing the significance of achieving accurate results within a reasonable timeframe. The choice of employing an image-based CNN over a signal-based approach is rooted in the capacity of image data to convey more intricate spatial and local patterns. When examining the initial peak of the impact-echo signal for delamination mapping, a consistent amplitude range emerges due to a fixed impact power within our system. This waveform pattern consistently features a pronounced peak succeeded by attenuation. By harnessing convolution to extract more nuanced spatial details, the image-based CNN demonstrates superior capability in handling these impact-echo signals, adept at capturing essential features for accurate analysis. By exploring the efficiency metrics, the study demonstrates the practical applicability of the developed AI techniques, elucidating the trade-offs between accuracy and computational resources. The research flowchart as shown in Figure 3.1. The dataset utilized in this study originates from conducted laboratory tests on artificial delamination slabs. The comprehensive data collection process involves the creation of energy intensity maps through frequency analysis. In pursuit of our first objective, leveraging this laboratory data, the realm of multi-input CNN is applied. Here, the power of advanced ML techniques, specifically the multi-input CNN model, is harnessed to enhance the precision of damage identification. Through this approach, comprehensive and precise delamination maps, providing insights into the structural integrity of the materials under examination are created. Simultaneously, our second objective involves a

thorough investigation of the efficiency of CNN. This exploration transcends mere accuracy, aiming to pinpoint the delicate balance between accuracy and computation time. The study rigorously dissects the relationship between the complexity of the input data, represented by variables such as image resolution and the number of images, and the computational demands incurred. Unraveling this intricate interplay provides invaluable insights into optimizing CNN models for real-time assessments, ensuring that both accuracy and efficiency are at the forefront of structural evaluation methodologies.



**Figure 3.1.** The flowchart of Chapter 3. The research pursues two key objectives: firstly, employing the collected data, it explores the realm of multi-input CNN to enhance damage identification precision. This innovative approach involves harnessing the power of advanced ML techniques, specifically the multi-input CNN model, to generate detailed and accurate delamination maps, illuminating the structural integrity of the materials studied. Concurrently, the study focuses into the nuanced interplay between precision and computation time within CNN efficiency.

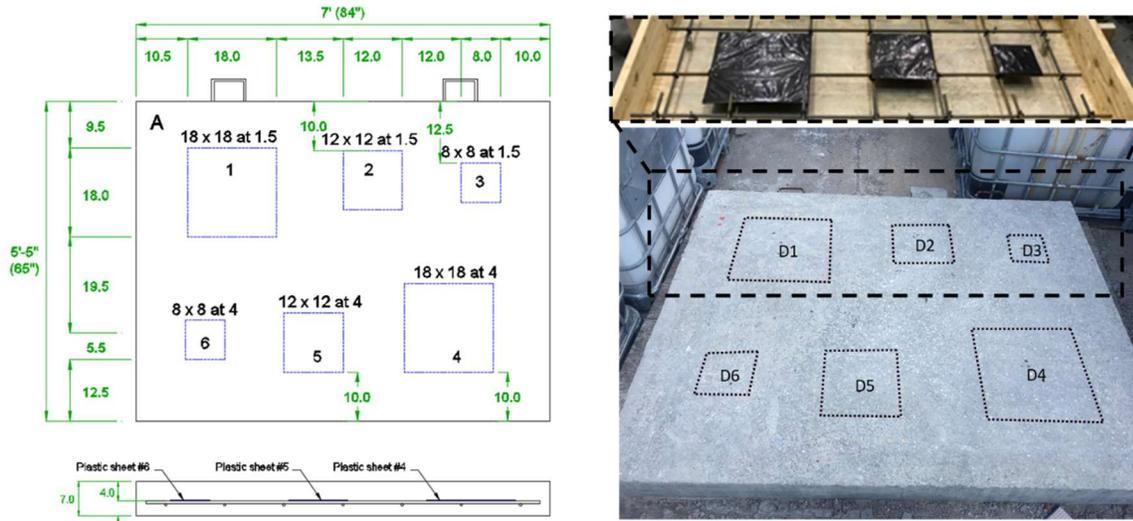
## 3.2 Data preparation

### 3.2.1 Impact echo laboratory test data

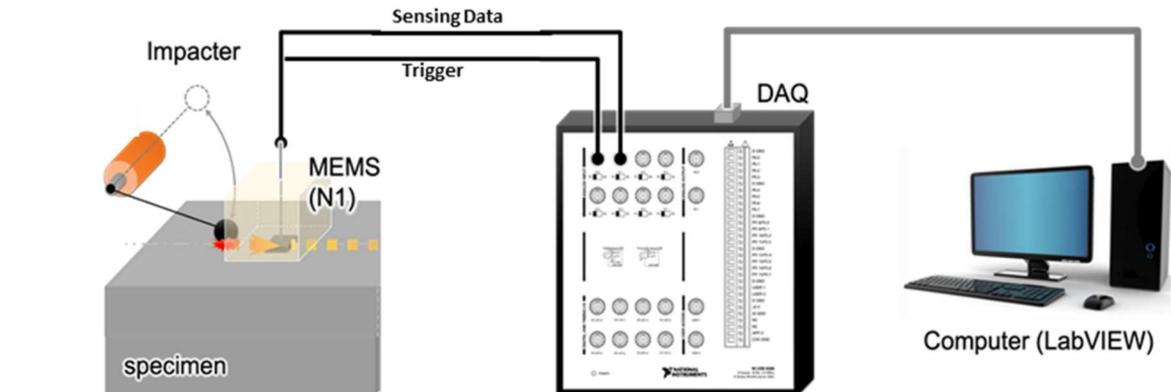
The testing slab dimensions measure 2.1 meters by 1.7 meters, incorporating six strategically placed artificial delamination of varying sizes and depths. These delaminations, mimicking real-world scenarios, serve as crucial indicators of structural integrity. The experiment utilizes three distinct thin plastic plate sizes: 46 cm × 46 cm, 30 cm × 30 cm, and 20 cm × 20 cm, positioned at depths of 4 cm and 10 cm within the slab shown in Figure 3.2. In our previous study (Kang et al., 2022), a standard impact diameter of 12 millimeters, providing consistency in our testing methodology using non-contact micro-electromechanical system (MEMS) are established. Specifically, delamination D1 to D3 are positioned at a depth of 4 cm, while delamination D4 to D6 are located at a depth of 10 cm. The slab's dimensions result in a total of 1280 testing points arranged in a grid of 32 by 40, with a 5-centimeter spacing between each testing point. This comprehensive configuration is visually represented in Figure 3.3 illustrating the meticulous placement of delamination and plastic plates across the testing slab. The impact echo algorithm is detailly described in previous study (Kang et al., 2022), a 1 kHz to 5 kHz frequency range is considered as vibration mode on our targeted size of delamination by increasing the frequency range. Thus, the energy intensity (EI) can be expressed as:

$$EI_n(f) = \int_{f_l}^{f_h} \Psi_n(f) df = \sum_{i=1}^N \Psi_n(f_i), \quad (f_1 = f_l, f_N = f_h), \quad (3.1)$$

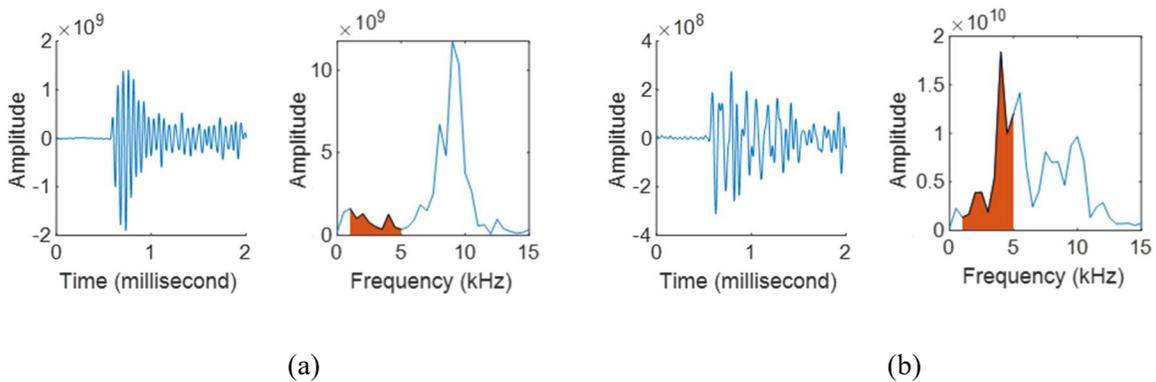
where  $EI_n(f)$  is energy intensity of  $n^{\text{th}}$  MEMS sensor;  $\Psi_n(f)$  is the spectral energy density of  $n^{\text{th}}$  MEMS sensor,  $\Psi_n(f) = |X_n(f)|^2$ , with a magnitude of the frequency component  $X_n(f)$ ;  $f_l$  and  $f_h$  are the low and high-boundary frequency (i.e., 1 kHz and 5 kHz in our system). The obtained EI related to flexural vibration mode is processed to create a two-dimensional (2-D) scanning image, or 2-D colormap, presenting delamination. All-post-processing is performed with MATLAB mathematical computing software. The testing signal samples are illustrated in Figure 3.4 and the severity of the damage is identified based on the energy of the target frequencies ranges. Typically, higher energy levels indicate a higher likelihood of delamination, whereas lower energy levels correspond to non-delamination cases. The objective of this test is to gather data, prepare inputs for the CNN model, and validate the predictions made by the CNN model.



**Figure 3.2.** The artificial delamination slabs, with three different size and two deployed depths cited from (Kang et al., 2022). There are a total of 1280 testing points on the slab, arranged in a grid of 32 by 40, with a spacing of 5 centimeters.



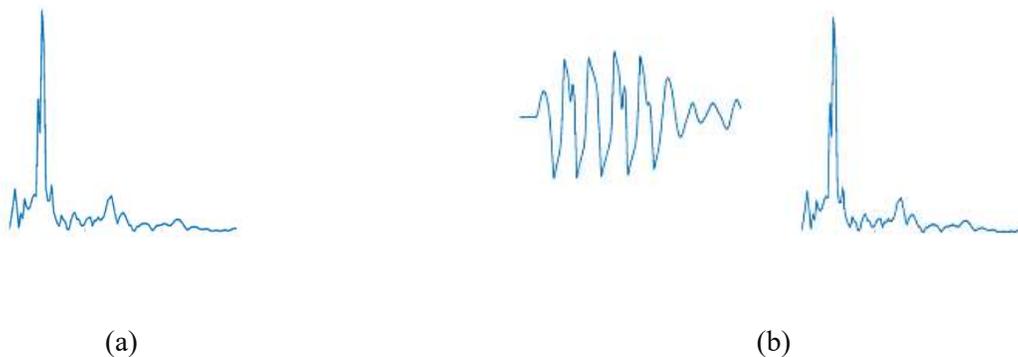
**Figure 3.3.** The testing configuration includes the use of non-contact MEMS sensors. The data acquisition system is developed using National Instruments (NI), and LabVIEW software is utilized to save the data.

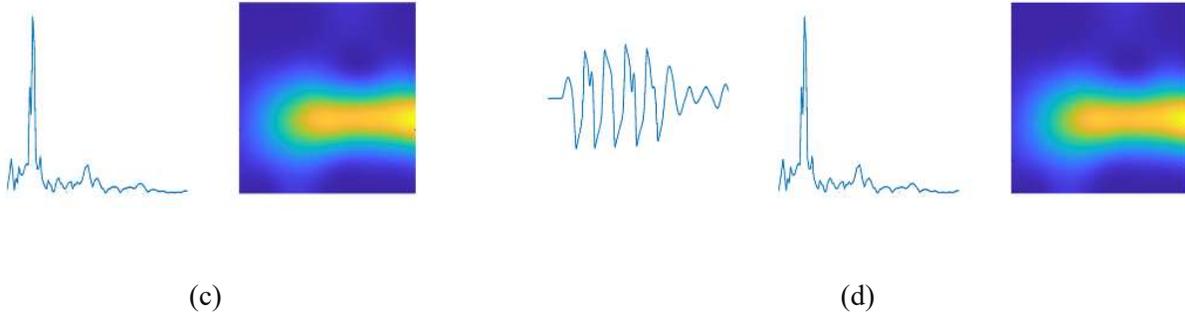


**Figure 3.4. Impact echo signal samples reveal delamination. Higher energy (1kHz to 5kHz) indicates severe delamination, while lower energy signifies intact structures. Shown in (a) non-delamination and (b) delamination cases.**

### 3.2.2 Multi-input data of damage identification model

In this study, four distinct CNN models, each utilizing different types of input data as shown in Figure 3.5 is investigated. The first model referred to as F relies on frequency-domain signals to provide insights into the frequency components of the data. The second model referred to as TF combines time-domain signals with frequency-domain signals, capturing a broader range of information for a more comprehensive analysis. The third model referred to as FSTFT incorporates frequency-domain signals with STFT, enhancing the model's ability to analyze time-varying frequencies and offering a more detailed understanding of signal dynamics. Lastly, the fourth model referred to as TFSTFT utilizes an image composed of time-domain signals, frequency-domain signals, and STFT, integrating these elements into a unified image format. By considering multiple data representations, the model gains a nuanced perspective on the input, enriching the analysis. The input image samples are shown in Figure 3.5. This study comprises seven classifications (D1 to D6 and non-delamination). For model training, 70% of the samples will be utilized, while the remaining 30% will be reserved for testing the damage identification model.





**Figure 3.5.** Examples of input images include: (a) frequency-domain signal input ( $F$ ), (b) multi-input with time-domain and frequency-domain signals ( $SF$ ), (c) multi-input with frequency-domain signal and STFT ( $FSTFT$ ), and (d) multi-input with time-domain signal, frequency-domain signal, and STFT ( $TFSTFT$ ).

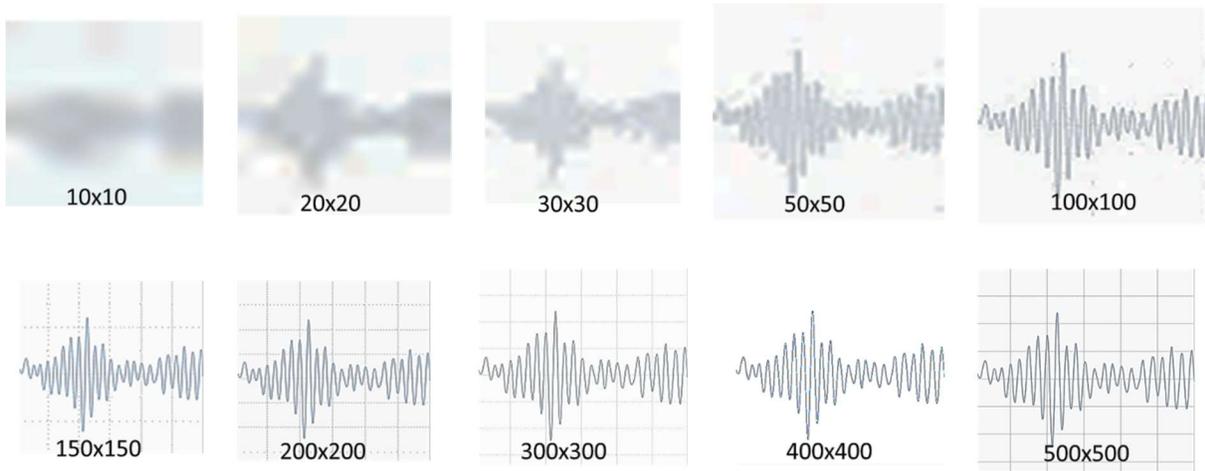
**Table 3.1.** The total number of samples correct from laboratory test.

Classification	D1	D2	D3	D4	D5	D6	ND
Number of samples	64	36	25	64	36	25	1030

### 3.2.3 Graphical-CNN efficiency study data

In addition to our primary goal of advancing the structural performance prediction model in this chapter, embarking on a parallel quest to explore the efficiency of AI, in particular, graphical ML such as CNN, in achieving a delicate balance between accuracy and computation time. Within this pursuit, two important factors come under scrutiny: image resolution and database or training size. These factors are crucial for optimizing ML models to achieve a balanced relationship between accuracy and computation time. Traditionally, lower-resolution images tend to yield quicker computation time but at the expense of accuracy, while higher-resolution images promise enhanced accuracy but demand longer computation time. However, the relationship between accuracy and computation time isn't a straightforward linear inverse correlation; in fact, achieving even a marginal 1% increase in accuracy may entail a substantial computation time increase of over 30 minutes. To explore deeper into this complex interplay and provide valuable insights for improving NDT results, understanding CNN efficiency becomes instrumental in guiding

decisions regarding image resolution and database size selection. To illustrate this, an example in Figure 3.6, which showcases varying image resolutions is present. Ten different resolutions ranging from 10 by 10 to 500 by 500 pixels, encompassing a total of 3000 images is considered. Each of the ten resolution cases consists of 300 images, enabling a comprehensive exploration of the impact of resolution on the trade-off between accuracy and computation time. Additionally, our study looks into training database size considerations, examining the effects of 40-image and 100-image databases across the same range of resolutions from 100 by 100 to 800 by 800 pixels. Through this multifaceted investigation, the valuable guidance for making informed decisions regarding image resolution and database size is provided, ultimately optimizing impact echo results while managing computational efficiency effectively.



**Figure 3.6.** *The image example of resolution study. Ten different resolutions ranging from 10 by 10 to 500 by 500 pixels, encompassing a total of 3000 images is considered. Each of the ten resolution cases consists of 300 images, enabling a comprehensive exploration of the impact of resolution on the trade-off between accuracy and computation time.*

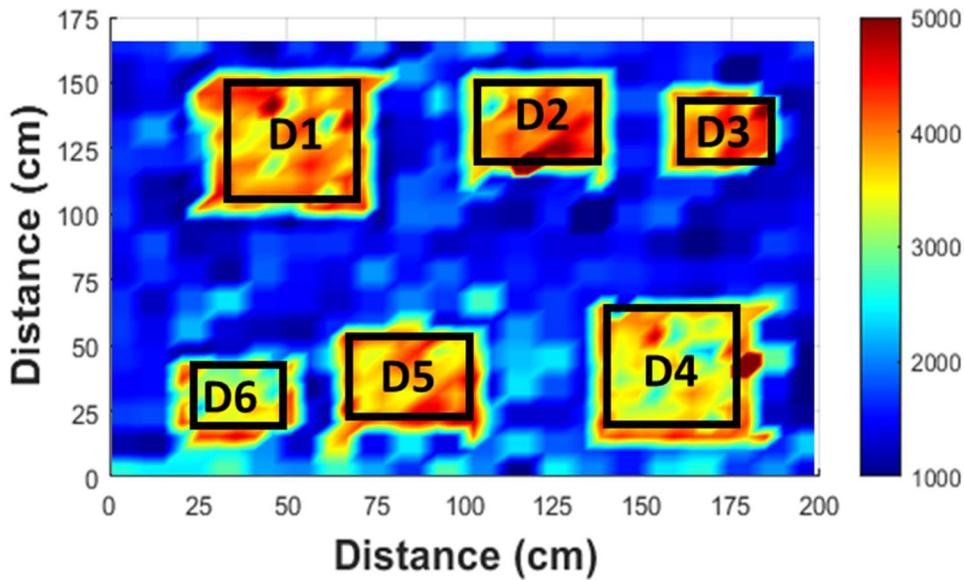
$$AI \text{ efficiency} = \frac{\text{Normalized Accuracy}}{\text{Normalized computation time.}} \quad (3.2)$$

### 3.3 Result and discussion

In this section, a comprehensive analysis that encompasses three key findings are unveiled. Firstly, the impact echo laboratory test results demonstrate its capability to identify variations in depth or size of delamination. Secondly, our exploration into CNN damage identification yields groundbreaking results.

### 3.3.1 *Impact echo laboratory testing result*

The IE testing results, illustrated in Figure 3.7, provide valuable insights into the delamination damage of the artificially damaged concrete slab. Significantly, areas exhibiting higher energy levels within this frequency spectrum are indicative of a higher likelihood of delamination. Upon closer examination, the locations labeled D1 to D6 exhibit notably higher energy levels compared to non-delaminated (ND) cases. The color bar represents the accumulated EI in the low-frequency range between 1 kHz to 5 kHz. Higher EI values are depicted in red, indicating highly delaminated areas, while lower EI values appear in blue, representing areas with minimal damage. A particularly intriguing observation emerges when comparing areas of varying depths. The top three regions (D1-D3), corresponding to shallow delaminations, display significantly higher energy levels compared to the bottom three regions (D4-D6) housing deeper delaminations. This disparity underscores the impact of depth on the behavior of mechanical waves. As the waves travel to deeper layers, the mechanical waves tend to disperse their energy. Consequently, this phenomenon causes a decrease in energy levels, thereby revealing a distinct pattern influenced by the depth of delamination. In essence, this nuanced analysis not only provides a clear understanding of the presence of delamination but also unravels the intricate dynamics of energy dispersion concerning the depth of the affected layers.



*Figure 3.7. The laboratory impact echo test was conducted on a slab featuring artificial delaminations. The color bar represents the accumulated EI in the low-frequency range between 1 kHz to 5 kHz. Higher EI values are depicted in red, indicating highly delaminated areas, while lower EI values appear in blue, representing areas with minimal damage. A noteworthy observation from the test results is the way energy diffusion varies with the depth of delamination. The upper three regions (D1-D3), indicative of shallow delaminations, exhibit markedly higher energy levels than the lower three regions (D4-D6) housing deeper delaminations.*

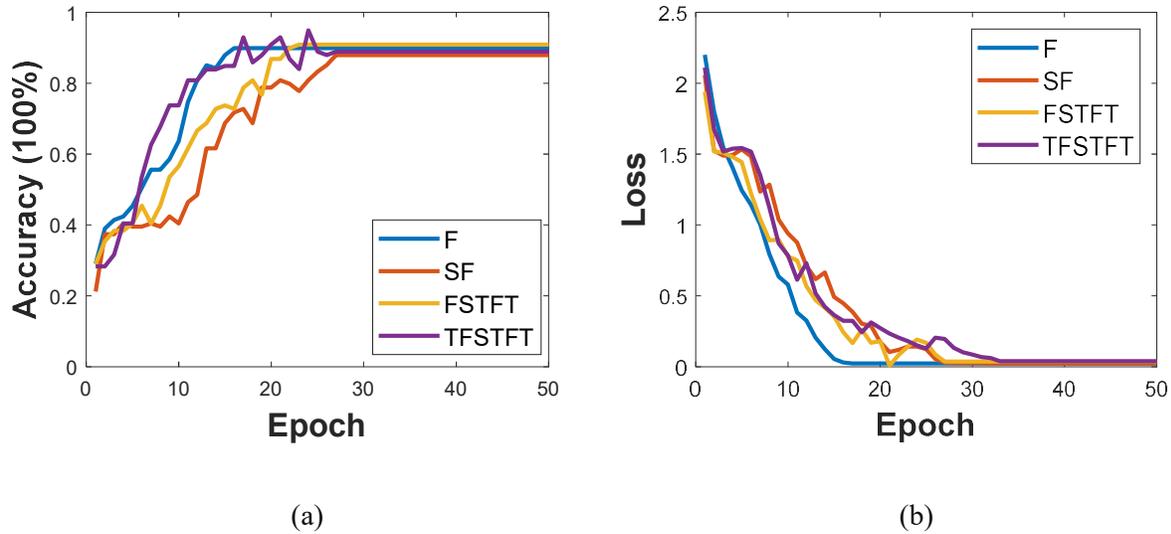
### 3.3.2 Damage identification result

In evaluating the performance of ML models, three distinct metrics come into play: training accuracy, validation accuracy, and test accuracy. Training accuracy serves as a crucial measure, depicting the model's ability to comprehend and interpret input data, discerning the intricate features in the data alongside their corresponding labels. Generally, a training accuracy exceeding 90% is anticipated, although this figure can vary based on the complexity of the input data. Once the training accuracy meets a predefined threshold, typically indicating a comprehensive understanding of the training dataset, the focus shifts to validation accuracy. Validation accuracy acts as a critical safeguard against overfitting, a scenario where the model learns the training data too well, capturing even insignificant features. Its role is to assess the model's capacity to generalize its knowledge to unseen data samples. This metric essentially tests the model's self-

evaluation ability, ensuring it can identify samples it has never encountered during training. Ideally, the validation accuracy should closely align with the subsequent test accuracy, reflecting the model's robustness and its competence in dealing with unfamiliar data. In this section, the accuracy metrics in the context of four different input datasets are explored. Through rigorous examination, not only discuss the nuances of training, validation, and test accuracy but also study into the intricacies of model predictions. A key output of this analysis is the predicted delamination map, which visually represents the model's insights into the structural condition. This map acts as a tangible representation of the model's learning, providing a valuable visual aid for understanding the distribution and severity of delaminations within the material. By considering these accuracy metrics alongside the corresponding delamination map, a comprehensive assessment of the ML model's performance and predictive capabilities is achieved.

The comparative analysis of four distinct input types is detailed in Figure 3.8. Notably, the final training accuracy for all inputs converges close to the 90% mark (as illustrated in Figure 3.8 (a)). However, it's intriguing to observe that the frequency-domain model (denoted as F) remarkably reaches its peak accuracy faster, typically within approximately 12 epochs, which means that the model remarkably reaches its peak accuracy faster. In other words, after iterating through the entire training dataset about 12 times, the model achieves optimal performance, demonstrating efficiency in learning and adaptation. This swift convergence suggests that the data in this category might be comparatively simpler, facilitating quicker model training. In contrast, the other three models require a more extended training period, around 25 epochs, to achieve their maximum accuracy. This discrepancy hints at the potential richness of information when using combined images, offering intricate details that demand a more thorough training process for the model to grasp. Figure 3.8 (b) further illustrates the training loss curves during the iterative training process. Consistently, F model demonstrates a rapid decline in loss, indicating a more efficient training process. Conversely, the other three models exhibit longer training times, signifying a more intricate learning curve. Upon achieving commendable training accuracies, the focus shifts to evaluating the validation curves, providing a crucial measure of the model's ability to generalize its knowledge to unseen data. These insights

into both training and validation processes serve as pivotal indicators of the model's robustness and predictive capabilities.

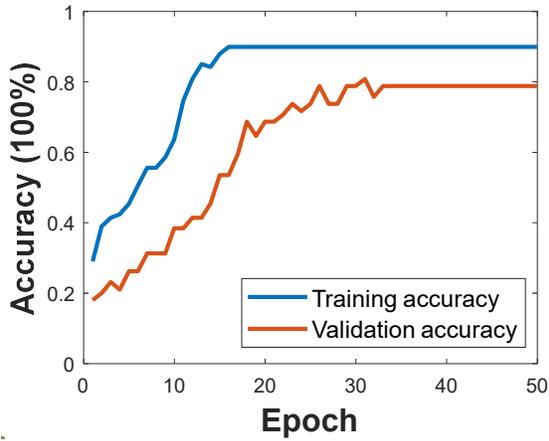


**Figure 3.8.** *The ML training accuracy and loss curves of four different models: (a) The training accuracy, and (b) training loss. The training accuracy and loss patterns for the four distinct models are evident. Notably, F model stands out with its rapid training process compared to the other three models, swiftly reaching maximum accuracy and minimal loss. This observation underscores that the data within category F is more easily learned by the model, indicating its simplicity in training.*

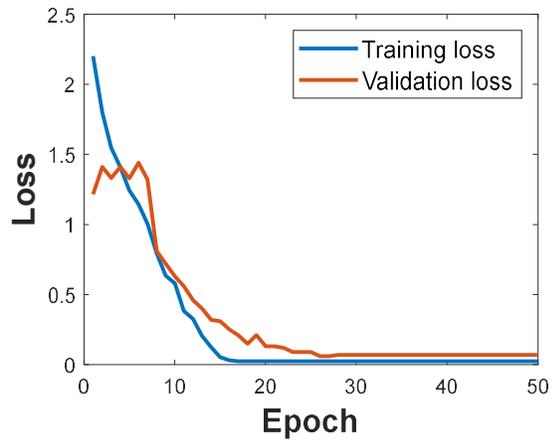
The validation accuracy and loss for various models are detailed in Figure 3.9 to Figure 3.12. An intriguing observation across all cases is the consistent trend: validation accuracy is lower than training accuracy, while validation loss surpasses training loss. This phenomenon is both reasonable and expected. During validation, the model undergoes a self-assessment process. It encounters a subset of input data, pretending to be unfamiliar with the corresponding answers for validation samples. This intentional lack of knowledge mimics real-world scenarios where the model encounters previously unseen data. Hence, the lower accuracy and higher loss in validation results can be attributed to the model's effort to generalize its learned patterns to unfamiliar samples, a crucial aspect in ensuring the model's robustness in practical applications. Validation accuracy serves as a criterion to discern whether an ML model is well-trained or susceptible to

overfitting. The validation accuracy curves of the four models consistently increase with training epochs, indicating a lack of overfitting. The incremental trends in accuracy curves are similar across all models, with the F model and TFSTFT model achieving maximum accuracy in fewer epochs.

This observation can be attributed to the complexity of the data and the impact of extraneous feature details. In cases where data is overly complex with unnecessary feature details, more training epochs are typically required. Among the four different models, the F model (Figure 3.9 (b)) exhibits the lowest validation loss, indicating that training with a single-frequency input image is easier and faster for CNN compared to other multi-input models. The F model reaches its minimum validation loss by epoch 30, outpacing the other three models. The validation and training curve for the SF model (Figure 3.10 (b)) reflects a larger gap between the two losses, suggesting lower model performance in learning SF data. SF data tends to provide excessive detail, requiring more epochs to reach the minimum loss value, and the minimum validation loss is the highest among the models. Conversely, the FSTFT and TFSTFT models show more similar trends to the F model and SF model, indicating that the data furnishes sufficient information for the models to understand features and enhances their ability to identify unlabeled or unclassified data (Figure 3.11 (b) and Figure 3.12 (b)). In addition, a notable trend is observed when comparing the training and validation curves. The validation curves consistently require more epochs to reach their peak values. This discrepancy underscores a fundamental principle in ML: the model's ability to generalize learned patterns to new, unseen data.

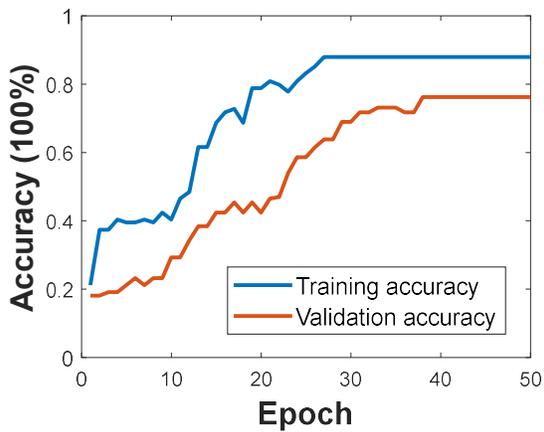


(a)

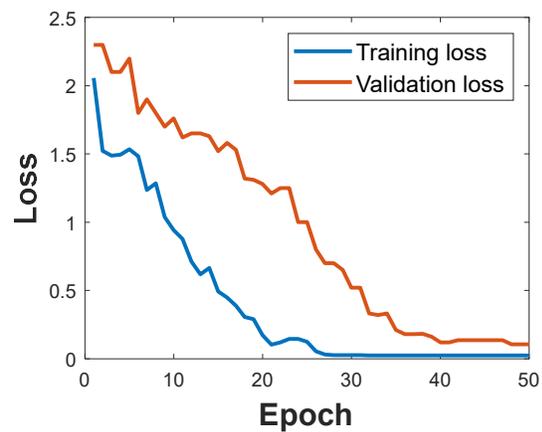


(b)

**Figure 3.9. Validation and training curves for F model: (a) accuracy curves, and (b) loss curves. The validation curves require more epochs to reach their peak values, indicating the model's adaptability in unpredictable environments with unknown data.**

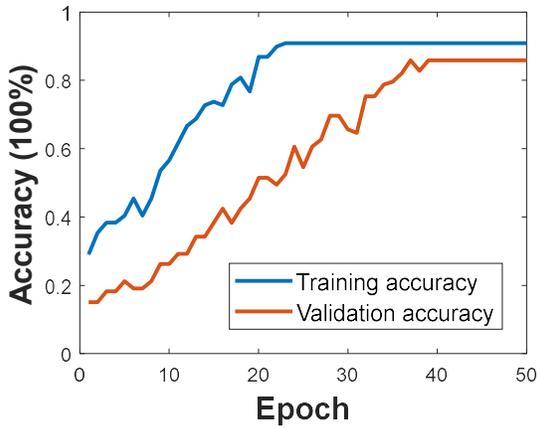


(a)

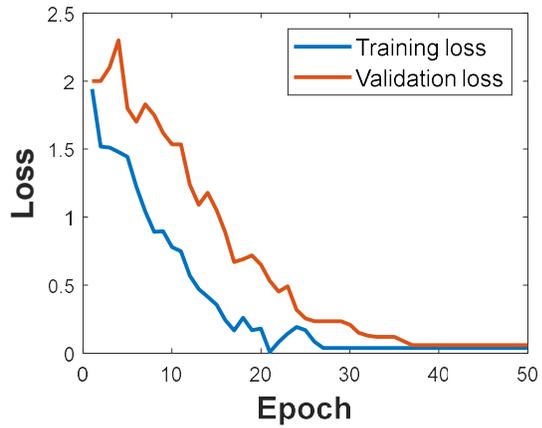


(b)

**Figure 3.10. Validation and training curves for SF model: (a) accuracy curves, and (b) loss curves. Compared with Figure 3.9, the validation loss takes more epochs to reach its minimum value, indicating the SF model has higher complexity compared to model F.**

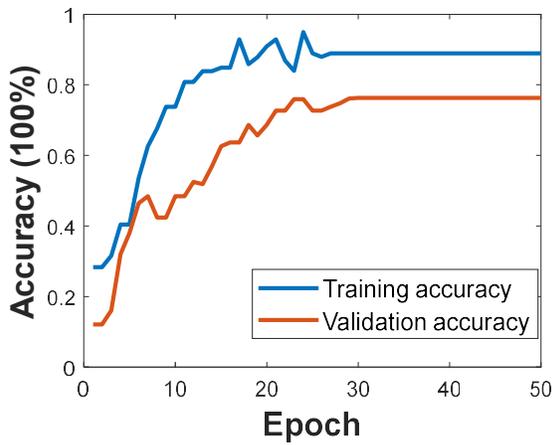


(a)

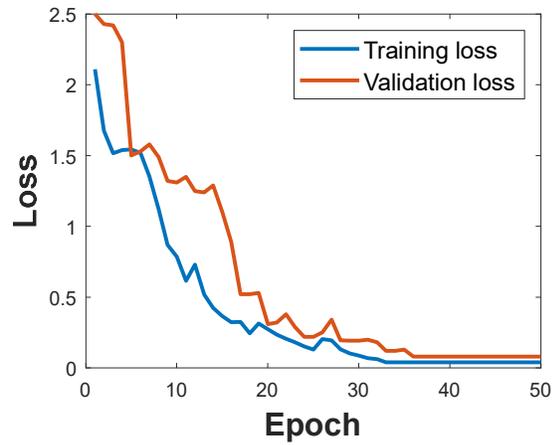


(b)

**Figure 3.11. Validation and training curves for FSTFT model: (a) accuracy curves, and (b) loss curves. The validation accuracy and loss are meet to threshold value around 40 epochs.**



(a)



(b)

**Figure 3.12. Validation and training curves for TFSTFT model: (a) accuracy curves, and (b) loss curves. The validation accuracy meets to threshold value around 30 epochs.**

The testing accuracy serves as a vital metric, gauging the model's proficiency in identifying unknown samples, which were not included in the training process. It's important to note that the testing sample constitutes 30% of the total laboratory testing data, ensuring a diverse and representative subset. The results are comprehensively depicted in the confusion matrix, showcased in Figure 3.13. Comparing the four

models, it becomes evident that the FSTFT model outperforms others, achieving the highest accuracy rate at 87%. In contrast, F model demonstrates an accuracy of 78%, while the remaining models maintain a solid 76% accuracy. These results are further detailed in Table 3.2, offering a nuanced understanding of the models' proficiency in damage identification. Remarkably, the FSTFT model exhibits exceptional accuracy in identifying delaminations across all depths (D1-D6). Additionally, the TFSTFT model displays commendable performance in recognizing D3 and D4, while the F model excels in identifying D4 and D5. This detailed breakdown underscores the specific strengths of each model in pinpointing different delamination levels, providing valuable insights for practical applications. Overall, these results underscore the nuanced capacities of the models and emphasize their potential in real-world scenarios where accurate damage identification is crucial.

		Target Class							
		D1	D2	D3	D4	D5	D6	ND	
Predicted Class	D1	12 3.09%	2 0.51%	0	0	0	0	5 1.69%	-
	D2	2 0.77%	5 1.28%	2 0.51%	0	0	0	8 2.06%	-
	D3	2 0.51%	1 0.25%	3 0.77%	1 0.25%	0	0	13 3.35%	-
	D4	1 0.25%	1 0.25%	1 0.33%	5 1.28%	1 0.25%	1 0.25%	17 4.38%	-
	D5	0	0	0	1 0.25%	5 1.28%	2 0.51%	5 1.28%	-
	D6	0	0	0	1 0.25%	3 1.01%	4 1.03%	17 4.38%	-
	ND	0	0	0	0	0	2 0.51%	266 68%	-
			70.5%	55%	50%	62.5%	55%	44%	80.3%

(a)

		Target Class							
		D1	D2	D3	D4	D5	D6	ND	
Predicted Class	D1	14 3.6%	2 0.51%	0	0	0	0	5 1.28%	-
	D2	2 1.03%	5 1.28%	1 0.25%	1 0.25%	0	0	11 2.83%	-
	D3	1 0.25%	1 0.25%	3 0.77%	2 0.51%	1 0.25%	2 0.51%	19 4.89%	-
	D4	0	1 0.25%	2 0.51%	4 1.03%	1 0.25%	1 0.25%	16 4.12%	-
	D5	0	0	0	1 0.25%	4 1.03%	2 0.51%	8 2.31%	-
	D6	0	0	0	0	3 0.77%	4 1.03%	12 3.09%	-
	ND	0	0	0	0	0	0	258 66.2%	-
			82.3%	55%	50%	50%	44%	44%	78%

(b)

		Target Class							
		D1	D2	D3	D4	D5	D6	ND	
Predicted Class	D1	15 2.83%	1 0.25%	0	0	0	0	3 0.77%	-
	D2	1 0.25%	6 1.53%	1 0.25%	0	0	0	4 1.03%	-
	D3	1 0.25%	1 0.25%	4 1.03%	1 0.25%	0	0	6 1.54%	-
	D4	0	1 0.25%	1 0.25%	5 1.28%	2 0.51%	0	10 2.57%	-
	D5	0	0	0	1 0.25%	5 1.03%	1 0.51%	7 1.8%	-
	D6	0	0	0	1 0.25%	2 0.51%	6 1.53%	8 2.06%	-
	ND	0	0	0	0	0	2 0.51%	294 75%	-
			88%	66%	66%	62.5%	55%	66%	88.8%

(c)

		Target Class							
		D1	D2	D3	D4	D5	D6	ND	
Predicted Class	D1	13 3.35%	1 0.25%	0	0	0	0	8 2.06%	-
	D2	2 0.51%	5 1.28%	0	0	0	0	9 2.31%	-
	D3	2 0.51%	1 0.25%	4 1.03%	1 0.25%	0	0	9 2.31%	-
	D4	0	2 0.51%	1 0.25%	5 1.28%	1 0.25%	2 0.51%	10 2.57%	-
	D5	0	0	1 0.25%	2 0.51%	4 0.77%	1 0.25%	15 3.86%	-
	D6	0	0	0	0	2 0.51%	4 1.03%	14 3.6%	-
	ND	0	0	0	0	0	2 0.51%	266 68.5%	-
			76.4%	55%	66%	62.5%	44%	44%	80%

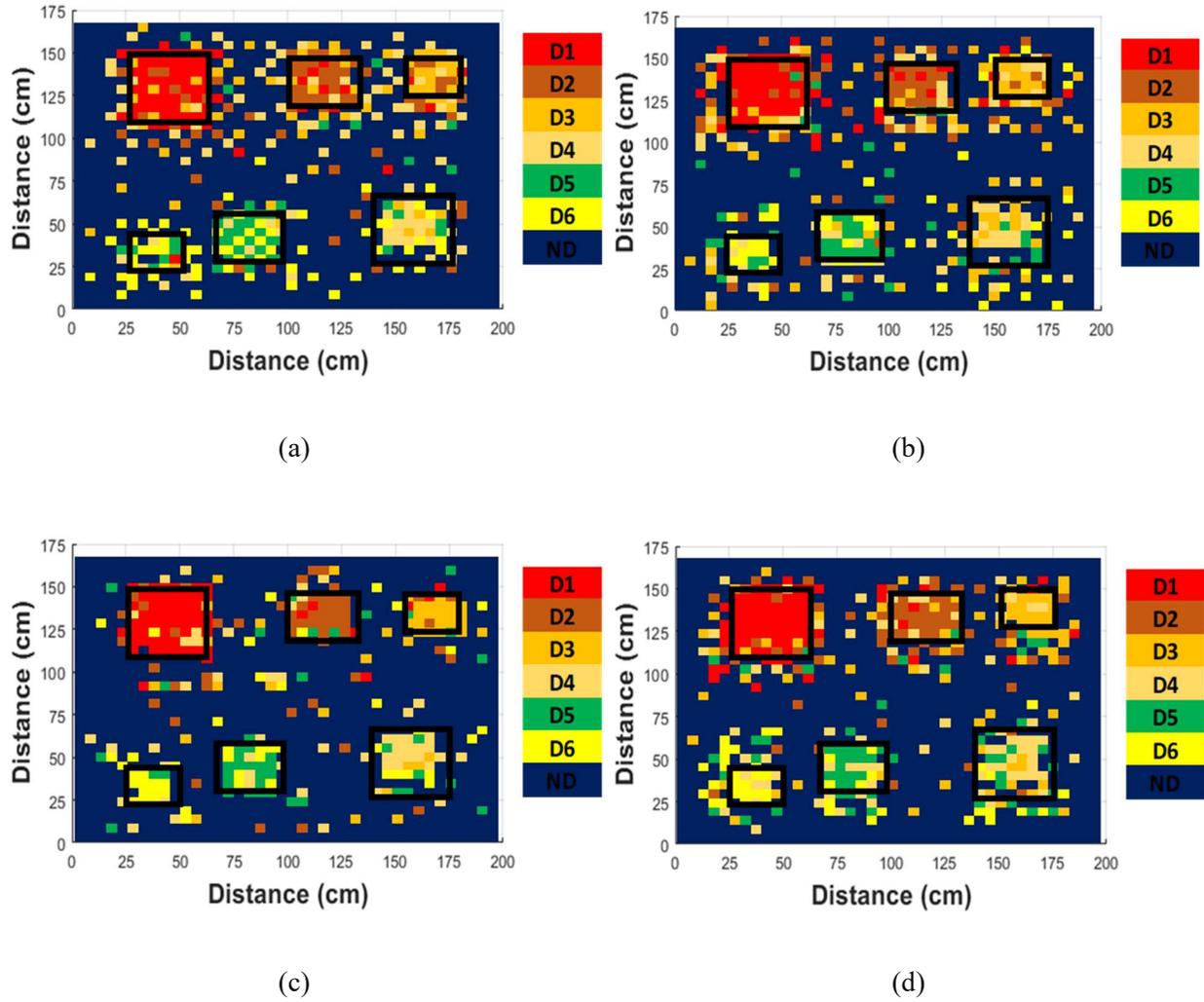
(d)

Figure 3.13. The confusion matrix of testing accuracy: (a) F model, (b) SF model, (c) FSTFT model, and (d) TFSTFT model. Overall, the FSTFT has the highest testing accuracy to show a better ability to identify delamination.

**Table 3.2. The accuracy of each damage of four models. Unit (%)**

	D1	D2	D3	D4	D5	D6	ND	Total
F	70.5	55	50	62.5	55	44	80.3	78
SF	82.3	55	50	50	44	44	78	76
FSTFT	88	66	66	62.5	55	66	88.8	87
TFSTFT	76.4	55	66	62.5	44	44	80	76

The damage prediction maps, as displayed in Figure 3.14, provide a comprehensive overview of the models' predictive capabilities to detect different damage sizes and depths. In the training process, 70% of the samples were utilized, while the remaining 30% were dedicated to testing. Each testing sample was examined, with known sample locations serving as reference points to obtain the predicted results from the CNN. These predictions were then used to construct the damage maps. Upon analysis, it becomes evident that the FSTFT model shown in Figure 3.14 (c) outperforms the others, demonstrating fewer incorrect predictions. This superior performance can be attributed to the validation and testing accuracy metrics. The FSTFT model boasts an impressive 87% test accuracy, while the other three models achieve around 75% accuracy. This disparity is pivotal; it indicates that the FSTFT model provides crucial features that enable the model to distinguish between damaged and undamaged areas more accurately. The nuanced insights from these maps offer valuable implications. Not only do they highlight the FSTFT model's superior predictive abilities, but they also underscore the importance of selecting inputs with higher accuracy rates.

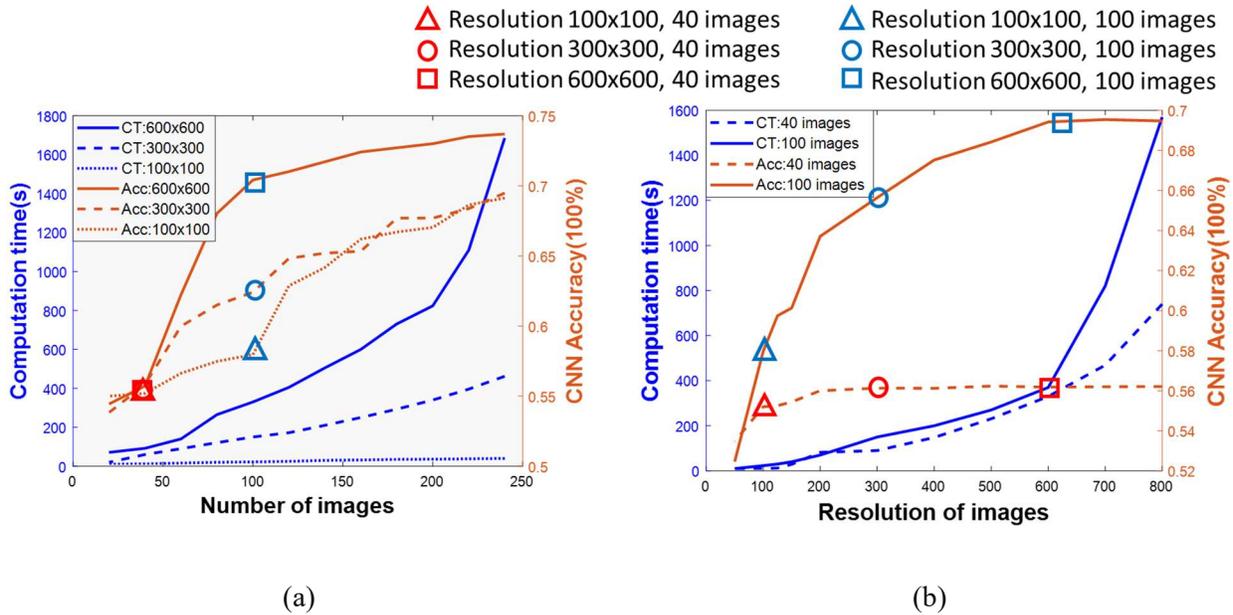


**Figure 3.14. The CNN prediction delamination map: (a) *F* model, (b) *SF* model, (c) *FSTFT* model, and (d) *TFSTFT* model. The results presented in (c) demonstrate the most accurate delamination map among all cases.**

### 3.3.3 CNN efficiency analysis result

During the damage identification process, the computation time for each training iteration is recorded. The relationship between computation time, accuracy, image resolution, and the number of training images is illustrated in Figure 3.15. The triangles, circles, and rectangles marked with the same color in both Figure 3.15 (a) and Figure 3.15 (b) represent CNN accuracy values obtained from the corresponding testing models. In Figure 3.15 (a), it becomes evident that higher image resolutions lead to increased accuracy. For

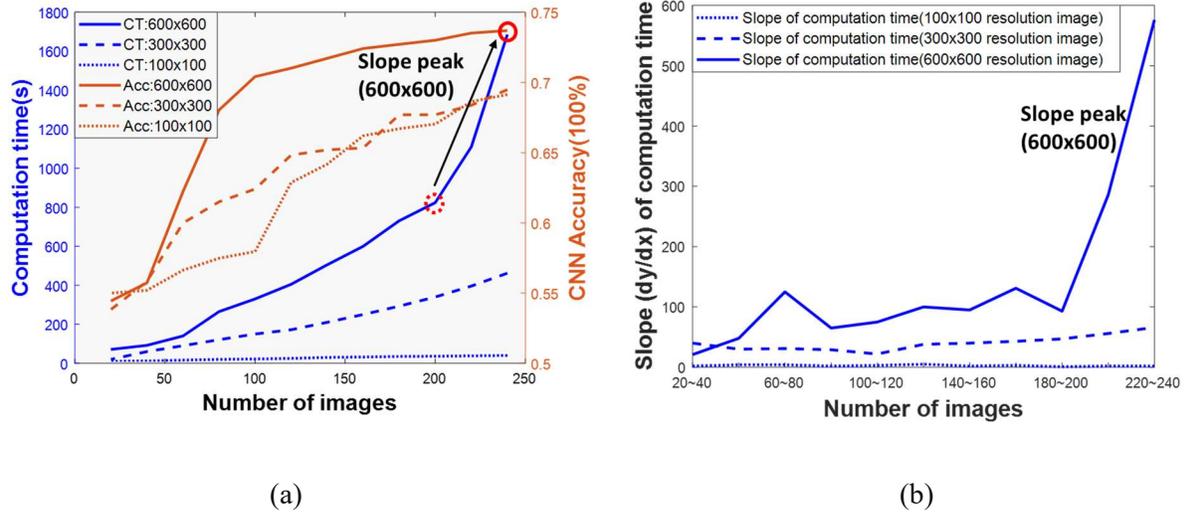
instance, with 100 images, a 600x600 resolution image attains 72% accuracy (blue rectangular mark), while a 100x100 resolution image achieves only 58% accuracy (blue triangle mark). However, this improvement in accuracy comes at a cost. The computation time exponentially increases. For the cases, the 600x600 resolution image requires 300 seconds, whereas the 100x100 resolution image demands only 50 seconds for processing. Similarly, with 240 images, the 600x600 resolution image achieves 74% accuracy, while a 100x100 resolution image attains 70% accuracy. However, the computation time drastically differs by 1600 seconds for the former versus 80 seconds for the latter. This comparison reveals that although a 4% accuracy difference exists between these two cases, the higher accuracy demands significantly more time for processing. Figure 3.15 (b) echoes these trends. When comparing models using 40 images and 100 images with an 800x800 resolution, the 100-image model attains close to 70% accuracy with a 1600-second computation time. In contrast, the 40-image model achieves 58% accuracy with a 700-second computation time. These phenomena highlight a critical relationship: image resolution and the number of training samples dramatically affect both computation time and accuracy, especially after reaching a certain threshold. Notably, resolutions larger than 600x600 lead to exponentially increased computation time. To strike a balance between accuracy and processing time, a comprehensive analysis is imperative. Determining this equilibrium point will be pivotal in optimizing the model's performance for practical applications.



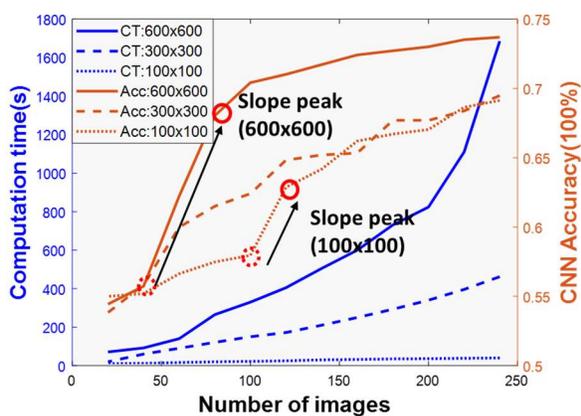
**Figure 3.15. The computation time and accuracy curves: (a) different number of images study, and (b) different image resolution study. The accuracy and computation time are significantly influenced by the number of images and the image resolution. However, beyond a specific threshold, such as 600x600 resolution, the computation time exponentially increases for a mere 2% improvement in accuracy.**

To find out the threshold, the changing difference between each increment is calculated, for example, analyze the difference between the computation time with 200x200 resolution image and the 100x100 resolution image, to identify how the resolution affect computation time. The number of image study with computation time and accuracy are shown in Figure 3.16 and Figure 3.17. Examining Figure 3.16 (b), a significant slope peak occurs with the 600x600 resolution image within the range of 200 to 240 images. This observation clarifies that for a 600x600 resolution image, utilizing more than 200 images results in an exponential increase in computation time, as evident in Figure 3.16 (a). Hence, the threshold for the number of images, concerning computation time, for the 600x600 resolution image is set at 200 images. In the analysis of accuracy slopes depicted in Figure 3.17 (b), two distinct slope peaks are evident, corresponding to the cases of 600x600 and 100x100 resolution images. Figure 3.17 (a) illustrates this phenomenon through the increase in accuracy, notably with the 600x600 resolution image when transitioning from 40 to 60 images, and with the 100x100 resolution image when increasing from 100 to 120 images. Consequently,

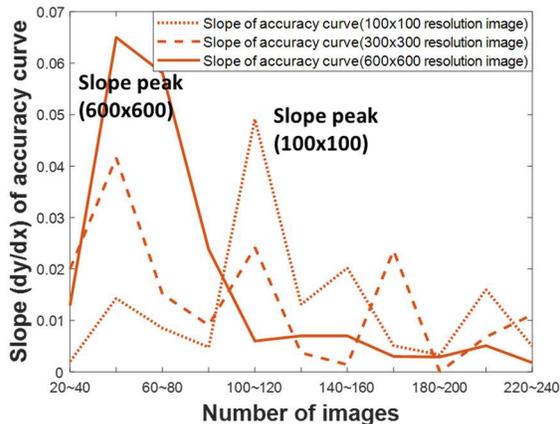
the number of images threshold, concerning accuracy, is set at 60 images for the 600x600 resolution and 120 images for the 100x100 resolution image. These thresholds represent crucial points where the balance between computational efficiency and accuracy is effectively maintained.



**Figure 3.16.** The slope analysis of computation time with number of image study: (a) the curves of computation time and accuracy, and (b) the slope analysis of computation time. For 600x600 resolution image, the number of image threshold for computation time is 200 images.



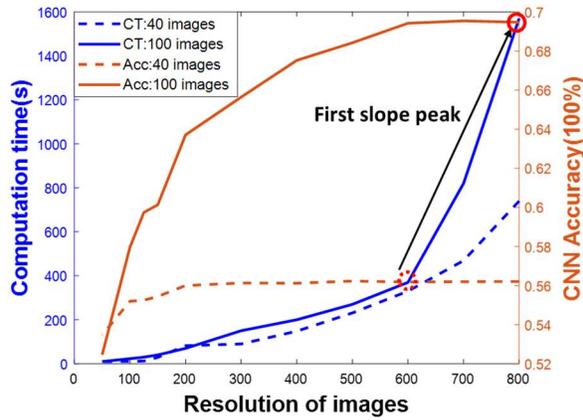
(a)



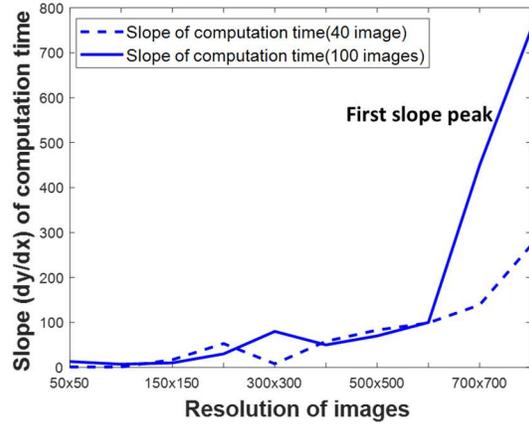
(b)

**Figure 3.17. The slope analysis of accuracy with number of image study: (a) the curves of computation time and accuracy, and (b) the slope analysis of accuracy. For 600x600 resolution image, the number of image threshold for accuracy is 60 images, while for 100x100 resolution image the threshold is 120 images.**

The analysis of slopes pertaining to resolution is presented in Figure 3.18 and Figure 3.19, considering inputs of 40 images and 100 images, respectively. In Figure 3.18 (b), a discernible slope occurs when the image resolution transitions from 600x600 to 700x700, as illustrated in Figure 3.18 (a). This delineates the threshold for computation time, signifying that a resolution of 600x600 provides a more efficient processing time. Examining the accuracy slopes in the resolution study, depicted in Figure 3.19 (b), two prominent peaks indicate significant accuracy enhancements. These observations are detailed in Figure 3.19 (a). The initial boost occurs when the resolution increases from 50x50 to 100x100, and the subsequent boost is observed from 150x150 to 200x200. Notably, resolutions exceeding 200x200 yield marginal increases in accuracy. This nuanced analysis underscores the critical impact of resolution on accuracy, emphasizing specific points where resolution adjustments significantly affect computational efficiency and model performance.

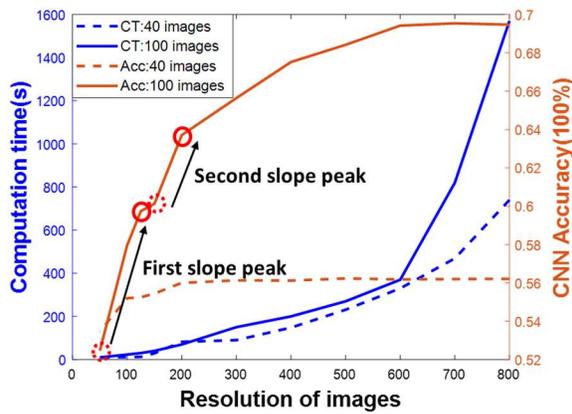


(a)

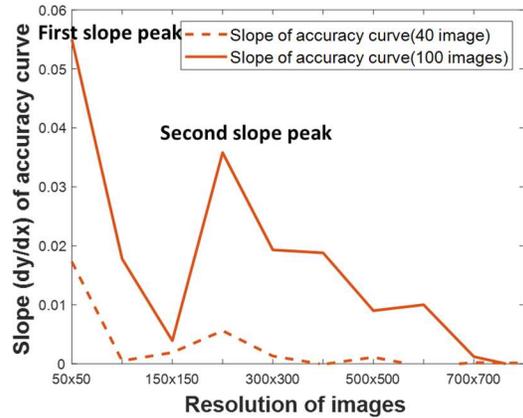


(b)

**Figure 3.18.** The slope analysis of computation time with image resolution study: (a) the curves of computation time and accuracy, and (b) the slope analysis of computation time. For 100 images input, the image resolution threshold for computation time is 600x600.



(a)

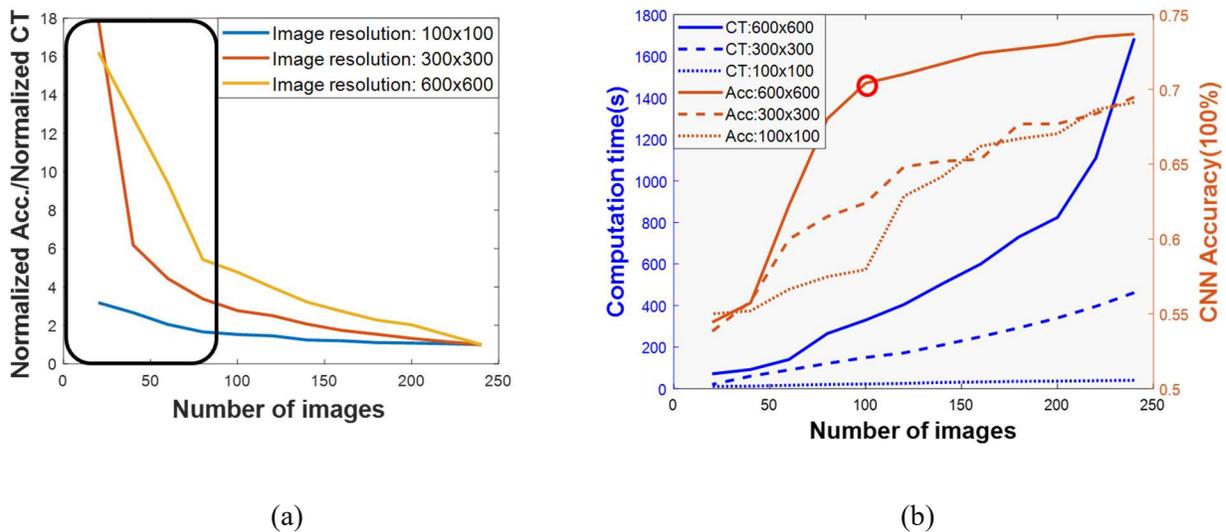


(b)

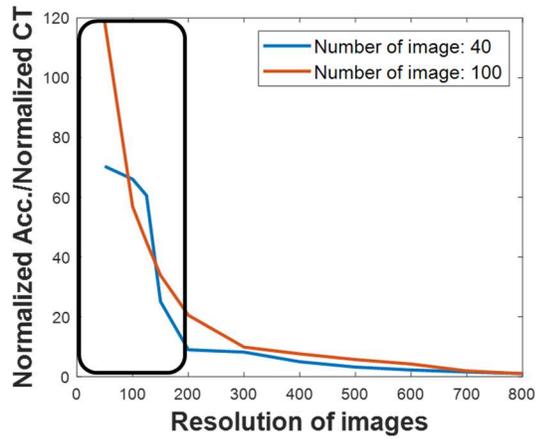
**Figure 3.19.** The slope analysis of accuracy with image resolution study: (a) the curves of computation time and accuracy, and (b) the slope analysis of accuracy. For 100 images input, the image resolution threshold for accuracy is 200x200.

The efficiency curves calculated using Eq. 3.2 are depicted in Figure 3.20 and Figure 3.21. Figure 3.20 (a) reveals that efficiency values are higher when the number of images is below 100. However, these values must be evaluated in conjunction with the accuracy and computation time thresholds discussed in Figure

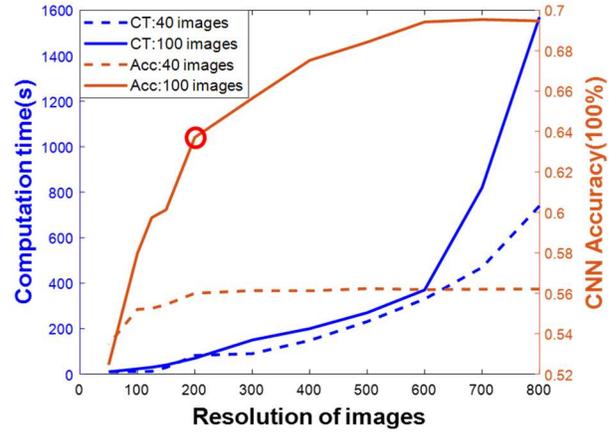
3.16 to Figure 3.19. For instance, the input featuring a 600x600 resolution with 20 images exhibits the highest efficiency. Despite this, the accuracy only reaches 56%, falling short of the accuracy threshold of 72% (the red mark in Figure 3.20 (b)). Therefore, even though the efficiency is high, it cannot be directly utilized for CNN model input setting. Similar trends are evident in Figure 3.21 (a), where high efficiency is observed with a 50x50 resolution image, but the accuracy is only 53% (as shown in Figure 3.21 (b)). Taking both accuracy and computation time into account, the optimal setting identified in this study is a 200x200 resolution with 100 images input (the red mark in Figure 3.21 (b)). This configuration achieves high accuracy while maintaining a rapid computation time, striking a balance that ensures both precision and efficiency in the CNN model.



**Figure 3.20. The CNN efficiency study with image resolution: (a) The efficiency curve of number of image study; (b) the curves of computation time and accuracy. The analysis indicates that efficiency values tend to be higher with fewer than 100 images. However, these values need to be carefully considered alongside accuracy and computation time thresholds.**



(a)



(b)

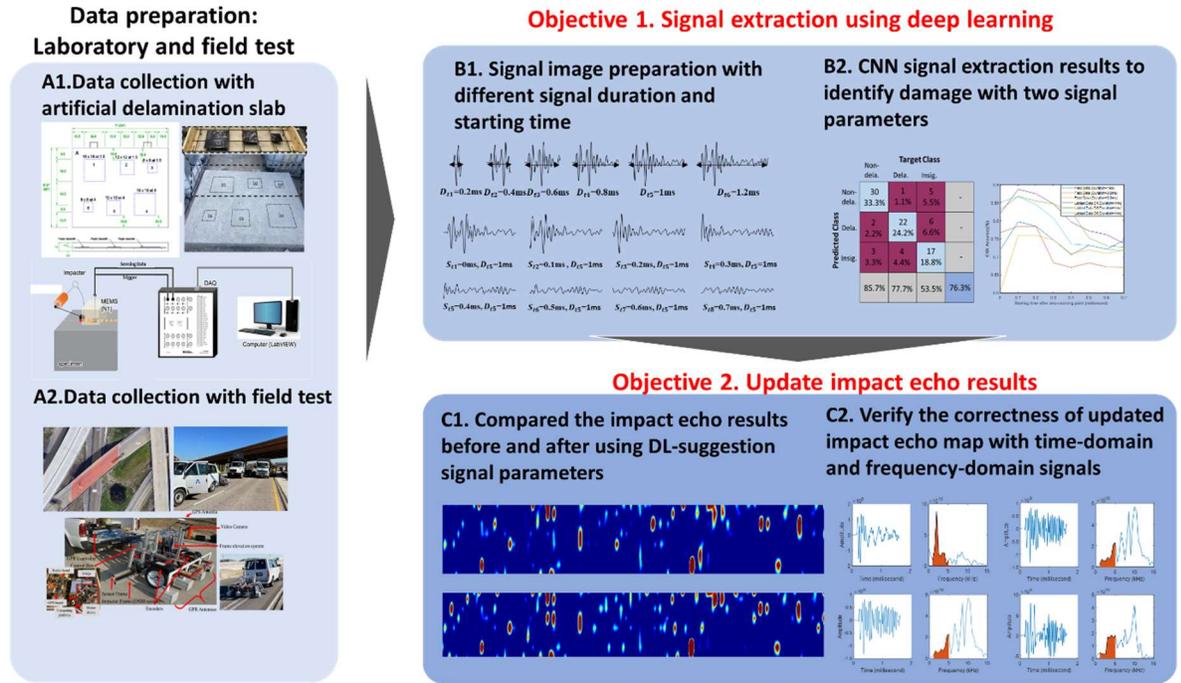
Figure 3.21. The CNN efficiency study with number of images: (a) The efficiency curve of image resolution study. (b) the curves of computation time and accuracy. A high efficiency is noted with a 50x50 resolution image, but the accuracy remains at 53%.

## CHAPTER 4 IE SIGNAL EXTRACTION STUDY

### 4.1 Chapter overview

Given the existing difficulties in signal extraction from advanced impact echo data in the field test, arising from diverse environmental conditions or multiple impact sources, the use of ML emerges as a potential solution. This study primarily employs image-based CNN to analyze various signal types. Additionally, a comparison is conducted with data-based ANN to evaluate the features usage of signal identification.

In order to extract signals effectively from massive data collect, this advanced model plays an important role in categorizing three distinct signal types: delamination signals, non-delamination signals, and noise or insignificant signals, as illustrated in the flowchart in Figure 4.1. In the first objective, our exploration involves two comprehensive signal studies: exploring different signal durations ( $D_{ti}$ ) and examining various signal starting times ( $S_{ti}$ ). In the second objective, the CNN-recommended signal parameters of  $D_{ti}$  and  $S_{ti}$  are applied to update the impact echo results, leading to a remarkable enhancement in map accuracy. It's noteworthy that the data utilized for training the CNN and ANN models is sourced directly from rigorous field tests. The integrity of the impact echo system used for data collection has been previously verified and published (Kang et al., 2022).



**Figure 4.1.** The flowchart of Chapter 4. Objective 1 focuses on signal extraction through deep learning, aiming to determine the signal parameters such as time duration and starting time, crucial for accurate damage identification. In Objective 2, these identified parameters are utilized to update the impact echo results.

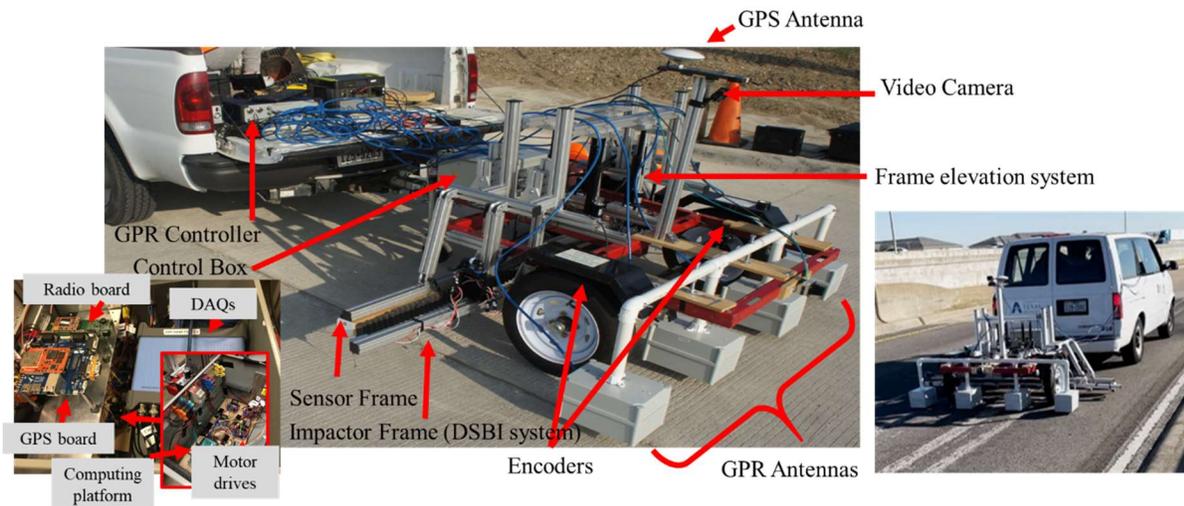
**4.2 Field test environment**

The bridge field tests, and system configuration are illustrated in Figure 4.2 and Figure 4.3. The field test faced significant challenges due to environmental interferences and machine-related factors affecting signal clarity, including issues such as potholes and the slope of the scanning path, as depicted in Figure 4.4. These surface irregularities significantly impact the accuracy of impact echo data analysis, influencing the quality and reliability of the results obtained. Potholes and surface damage can disrupt the impactor’s contact with the bridge, leading to variations in the generated mechanical waves; The impactor’s ability to generate mechanical waves, crucial for the echo signal, is directly influenced by the ground surface condition. If the surface is uneven or damaged, the impactor may not strike the ground effectively, leading to lower energy signals.

These low-energy signals pose a challenge during impact echo data analysis, as they might be overlooked, resulting in inaccurate delamination map outcomes. Furthermore, since our system includes nine motor drivers to generate waves, and each motor has a 20-millisecond delay, signals generated by other motors may be observed by each sensor. This is a challenge because signals from distant impacts could potentially affect the data from the nearest impact (Kang et al., 2022). These undesirable signals, despite having lower amplitude, remain discernible in the data. To address these challenges arising from natural environmental interference and the multi-impact signals during field tests, our study explored signal extraction techniques enhanced through ML methods, as discussed in the subsequent sections.



**Figure 4.2.** Field testing photo - (a) scanned area of the bridge, and (b) system and safety control.

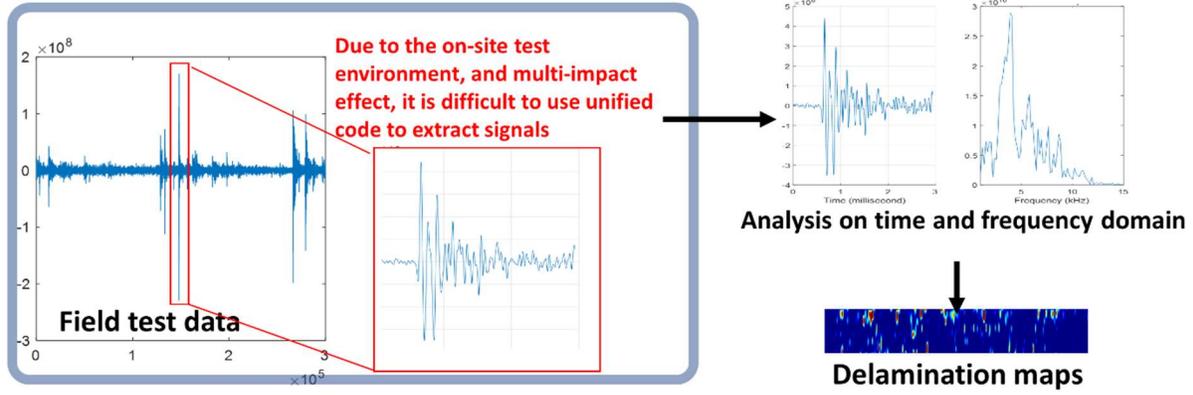


**Figure 4.3.** The ACES scanning system encompasses impact-echo and GPR technologies cited from (Kang et al., 2022).



***Figure 4.4. The surface condition highlights prevalent issues such as potholes and damaged surfaces. Potholes and surface damage can disrupt the impactor's contact with the bridge, leading to variations in the generated mechanical waves.***

In Chapter 4, the goal is to enhance signal extraction leveraging ML aimed at overcoming prevalent challenges in impact echo testing, such as environmental interferences and machine-related factors affecting signal clarity. The example signals as shown in Figure 4.5. Analyzing extensive scanning datasets from field tests covering entire bridges, often surpassing 30 million samples and susceptible to environmental variations, poses a challenge in accurately identifying target signals. Factors like delayed impact and diminished mechanical wave amplitudes add complexity, making the use of uniform mathematical methods difficult. The precision of delamination identification relies on accurate signal extraction. To address these challenges, two powerful DL algorithms are employed: data-based ANN and image-based CNN. These methods help to classify signals, distinguishing delamination, non-delamination, and insignificant signals, aiming to enhance signal extraction accuracy from real-world impact echo testing data.



**Figure 4.5.** The impact echo algorithm aims to generate a delamination map, but it encounters a significant challenge: processing over 30 million sample data points within a single scan time and the effect of multi-impact. Identifying the precise target peaks using traditional mathematical algorithms becomes daunting due to this vast dataset. To address this challenge, DL algorithms are employed to accurately identify signals.

### 4.3 Signal extraction ML algorithm

#### 4.3.1 ANN

ANN consists of interconnected nodes or neurons organized in layers, including an input layer, one or more hidden layers, and an output layer. These networks are trained using the backpropagation (BP) algorithm, a supervised learning technique tailored for neural networks. The BP updating error equation can be expressed by:

$$\delta_i = (Out_i)(1 - Out_i)w_{i,i-1}\delta_i \quad (4.1)$$

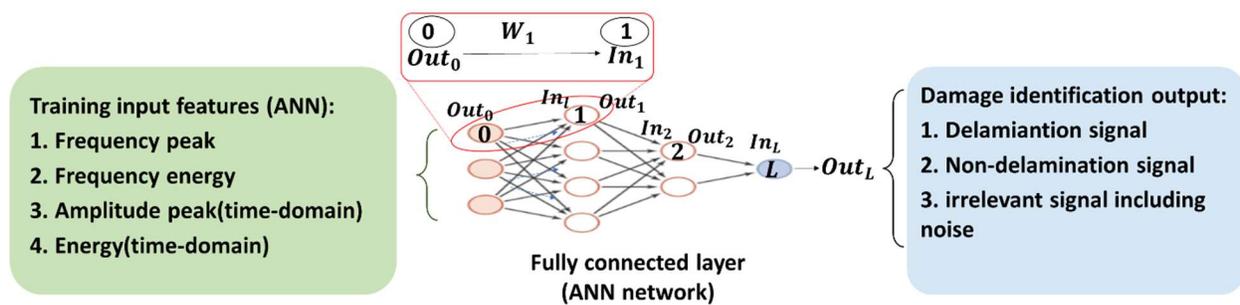
$$\delta_1 = Out_1(1 - Out_1)w_{2,1}\delta_2 \quad (4.2)$$

$$\delta_2 = Out_2(1 - Out_2)w_{L,2}\delta_L \quad (4.3)$$

where  $w$  represents data weights,  $\delta_i$  signifies the error value in neuron  $i$ ,  $L$  is the last layer neuron and  $Out_i$  represents the output data from neuron  $i$  as shown in Figure 4.6. During training, BP calculates the error

between the predicted output and the actual target values. It then adjusts the weights and biases of the network, working backward from the output layer to the input layer, in order to minimize this error. This iterative process continues until the network's predictions closely align with the desired outcomes.

Figure 4.6 as example, the output from neuron 0 is used to input for next layer neurons (e.g., neuron 1), and the error  $\delta$  from neuron 1 is updated from output of neuron 1, data weight, and error from neuron 2 as described in Eq. 4.3. As a result, ANNs become adept at recognizing intricate patterns, enabling accurate predictions and classifications. The inputs for neuron 0 consist of features derived from signals described in Eq. 4.4 to Eq. 4.7. The updated calculated weight from neuron 0 serves as input for the next layer neuron, such as neuron 1. The final output of the last neuron encompasses three types of damage identification.



**Figure 4.6. Schematic representation of the ANN architecture. In the training process, BP computes the error by comparing predicted and actual outputs. It then iteratively adjusts the network's weights and biases from the output layer to the input layer to minimize this error. The input features consist of frequency and time-domain signals, with the output focused on damage identification.**

The time-domain peak serves as a direct representation of the impactor energy and the energy of mechanical waves in the impact echo signal. A higher amplitude indicates correct impactor hitting conditions, although sometimes environmental interferences and machine-related factors can lead to lower amplitudes. These time-domain peaks have a direct impact on the frequency-domain energy, affecting the accumulated energy as well. To enhance the granularity of features used for training and testing the ANN model, four distinct feature values are introduced to characterize the impact echo signal: 1) the maximum peak in the time-

domain, 2) the maximum peak in the frequency-domain, 3) the accumulated energy between 1kHz to 5kHz, and 4) the accumulated energy between 7kHz to 11kHz, as expressed in Eq. 4.4 to Eq. 4.7:

$$Feature\ 1 = \max(Amp(t)_{x,y}) \quad (4.4)$$

$$Feature\ 2 = \max(Amp(f)_{x,y}) \quad (4.5)$$

$$Feature\ 3 = \sum_{1kHz}^{5kHz} E(f) \quad (4.6)$$

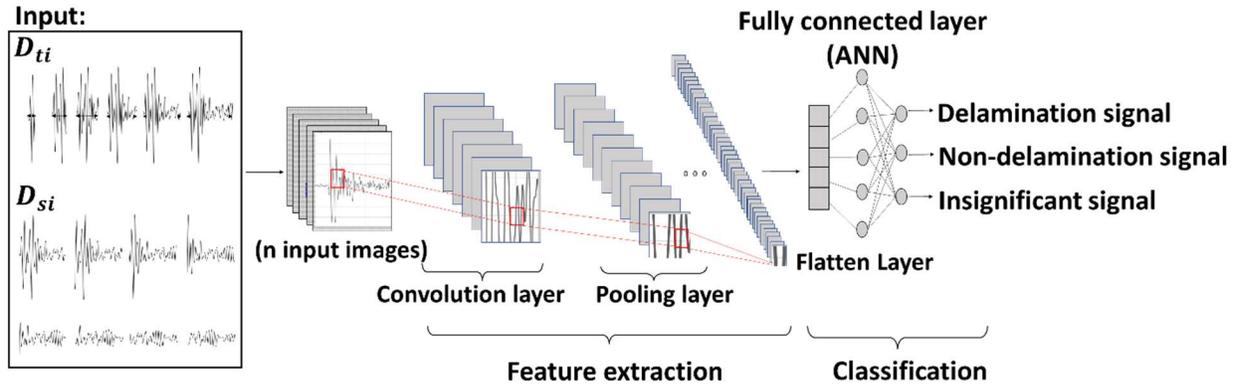
$$Feature\ 4 = \sum_{7kHz}^{11kHz} E(f) \quad (4.7)$$

where  $Amp(t)_{x,y}$  and  $Amp(f)_{x,y}$  are amplitude of time-domain (t) and frequency-domain(f) with time and amplitude value, separately.

### 4.3.2 CNN

The CNN model is leveraged to process time-domain signals accompanied by damage labels, distinguishing between delamination, non-delamination, and insignificant signals. Two parameters associated with various signal windows, namely starting time signal ( $S_{ti}$ ) and different duration signal ( $D_{ti}$ ) of the input signal, are used to identify damage labels. Afterward, the CNN output, which furnishes the  $S_{ti}$  and  $D_{ti}$  values associated with the highest accuracy in damage identification, will be implemented to enhance the extraction algorithm, thereby improving the final delamination map. The CNN's operation entails intricate layers, including convolutional layers and pooling algorithms. These layers examine input images, utilizing

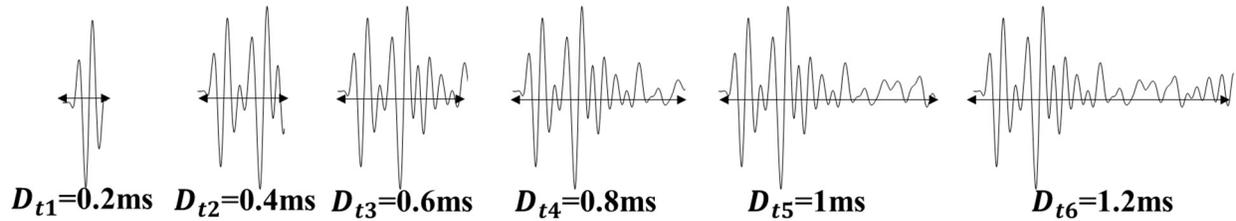
convolution with smaller elements to deconstruct the image into more manageable feature images, as depicted in Figure 4.7. It is crucial to highlight that our input data originates from both field and laboratory tests, encompassing real-world scenarios. This approach ensures the CNN model's practical applicability.



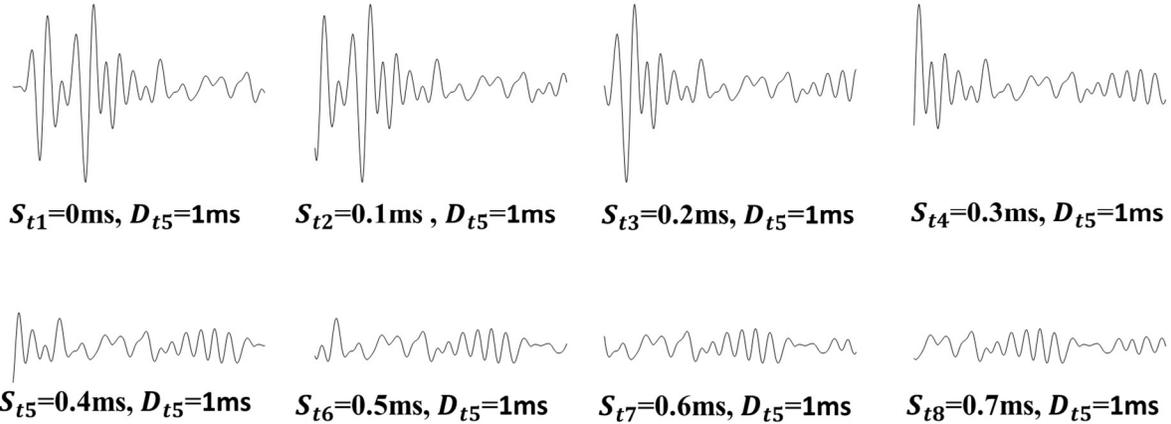
**Figure 4.7.** The CNN model comprises a convolution layer for feature extraction and a pooling layer to reduce dimensions, retaining essential features. Extracted features are connected to a flatten layer and a neural network for training and testing purposes.

The cornerstone of our methodology is the construction of a comprehensive database, tailored to signal parameters, encompassing variations in the starting time and duration of signals. This parameter-driven database serves as the bedrock for our efforts to differentiate between significant signals, which are instrumental in producing precise impact echo results, and insignificant signals that might introduce noise or interfere with accurate assessments. The sample of different  $S_{ti}$  and  $D_{ti}$  in time-domain as shown in Figure 4.8 and Figure 4.9. The signal duration study depicted in Figure 4.8, a range of 10 different durations spanning from 0.2 milliseconds to 2 milliseconds (e.g.,  $D_{t1}=0.2$  millisecond,  $D_{t2}=0.4$  millisecond and  $D_{ti} = i \times 0.2$  millisecond) are examined. Each duration represents a distinct temporal window within which the signal is analyzed. Furthermore, for the  $S_{ti}$  study, eight different starting time scenarios (e.g.,  $S_{t1} = 0$  millisecond after zero-crossing point,  $S_{t2}=0.1$  millisecond after zero-crossing point and  $S_{ti} = (i - 1) \times 0.1$  millisecond after zero-crossing point) are investigated. The initial case of  $S_{t1}$  commenced precisely at the zero-crossing point, aligning with the moment just before the acquisition of the structural signal. For

instance, the initiation of  $S_{t2}$  occurred 0.1ms after the zero-crossing point, and  $S_{t3}$  subsequently followed with a delay of 0.2ms delay, as exemplified in Figure 4.9. In  $D_{ti}$  study, a total of 3000 images were analyzed. This encompassed 10 distinct cases, with each  $D_{ti}$  variant comprising a dataset of 300 images. Similarly, in the starting time study, a substantial dataset comprising 2400 images was employed. This comprehensive dataset included eight different  $S_{ti}$  scenarios, with each scenario comprising a collection of 300 images. This extensive dataset facilitated an in-depth exploration of the effects of varying starting times on our signal analysis.



**Figure 4.8.** The CNN inputs for signal extraction, examining various durations ( $D_{ti}$ ) of time-domain impact echo signals. A total of 10 different durations, ranging from 0.2 milliseconds to 2 milliseconds, are analyzed. Each dataset (e.g.,  $D_{t1}$  dataset) comprises 300 images obtained from field tests.



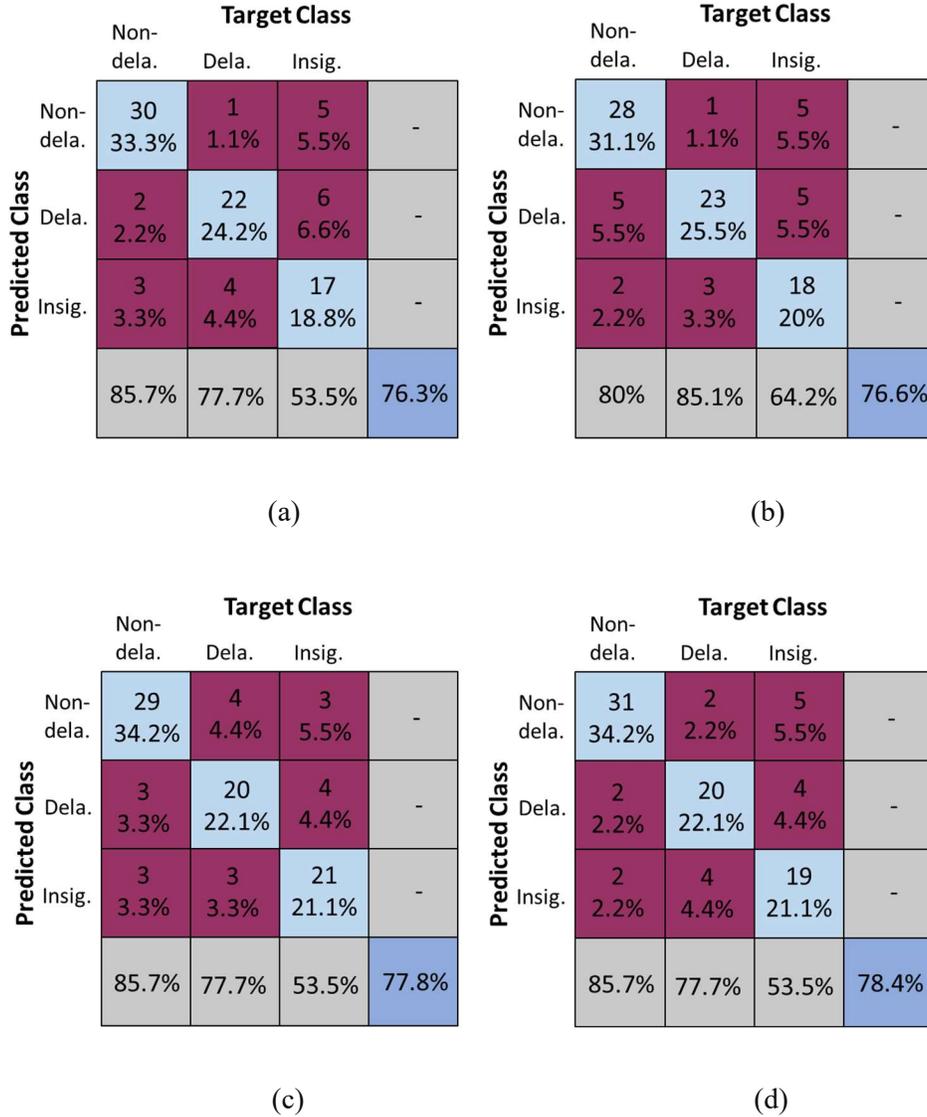
**Figure 4.9.** The study explores CNN inputs for signal extraction, varying the starting time ( $S_{ti}$ ) after the zero-crossing point. A total of eight different signals are considered, ranging from the immediate zero-crossing point to 0.7 milliseconds after it.

#### 4.4 Results and discussions

This section provides results from two aspects. Firstly, the precise classification of signals into delamination, non-delamination, and insignificant signal categories utilizing time-domain signals. This classification is conducted across diverse  $D_{ti}$  and  $S_{ti}$  after the zero-crossing point. Secondly, the impact echo delamination maps are refined using results obtained from our CNN-based signal extraction process. The enhancements are achieved by incorporating optimized values for  $D_{ti}$  and  $S_{ti}$ , ensuring a more accurate and detailed delamination mapping process.

#### 4.4.1 *CNN Signal extraction result*

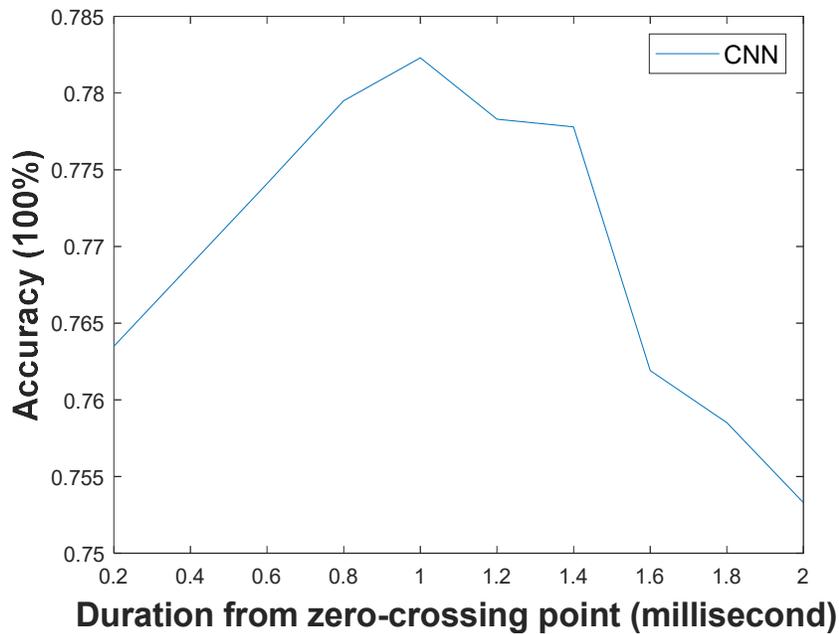
The CNN's signal extraction testing accuracy of  $D_{t1}=0.2$  millisecond case is depicted in Figure 4.10 (a), where it is essential to note that there were 210 images utilized for training and 90 images for testing. As an example, Figure 4.10 (a) show the 33.3% accuracy for non-delamination class that is calculated by dividing the 30 of correctly predicted sample number over the total 90 testing images, resulting in 33.3%. The figure also presents a value of 85.7% at the bottom, derived from the correctly predicted samples within each class (30 out of 35), accounting for 85.7%. The overall accuracy of the CNN model is represented by 76.3%, calculated as the sum of the correct predictions for each class (33.3% for non-delamination, 24.2% for delamination, and 18.8% for insignificant signal). Compared to result in Figure 4.10 (a), Figure 4.10 (b) represents the signal identification result for the case with  $D_{t2}=0.4$  milliseconds model, showing a slightly higher accuracy (76.6%). However, its ability to identify non-delamination has a slightly lower accuracy at 80%. Figure 4.10 (d) exhibits a higher model accuracy (78.4%) with  $D_{t5}=1$  millisecond model.



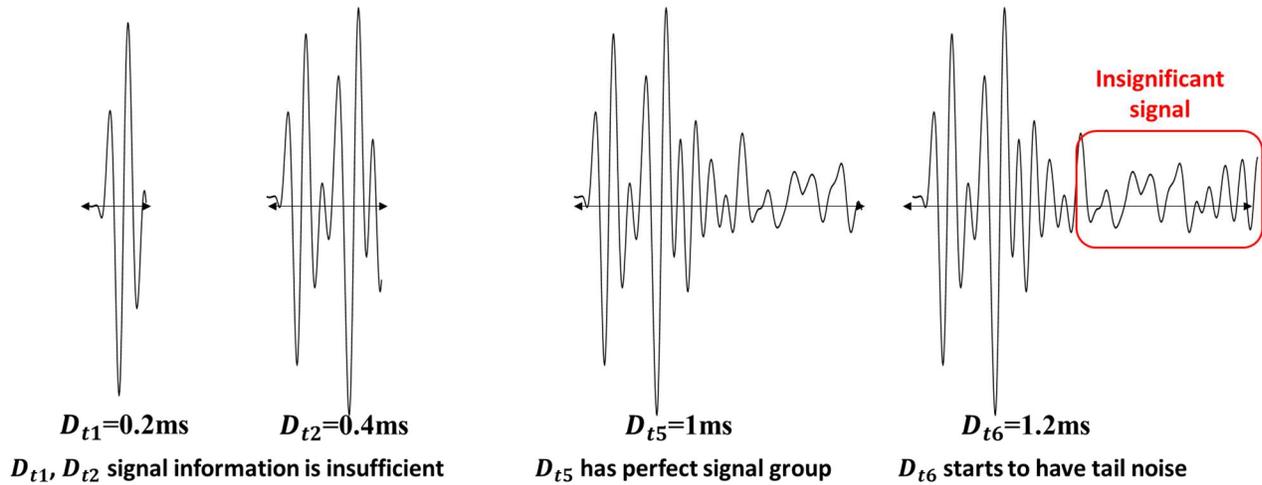
**Figure 4.10.** The confusion matrix illustrates the identification of different duration signals, categorizing them into non-delamination, delamination, and insignificant signal classes. (a)  $D_{t1} = 0.2ms$ , (b)  $D_{t2} = 0.4ms$ , (c)  $D_{t4} = 0.8ms$ , and (d)  $D_{t5} = 1ms$ .

In-depth analysis of testing accuracy across different  $D_{ti}$  presents distinctive outcomes, as illustrated in Figure 4.11. Particularly, when  $D_{t5} = 1ms$ , a notably higher prediction accuracy is achieved. This highest accuracy can be attributed to the focused 1ms window, efficiently capturing crucial impact signals indicative of structural interior damage. However, as the duration extends to 1.2ms, insignificant tail noise in the red box begins to be considered as shown in Figure 4.12, leading to a slight decline in CNN

classification accuracy. Further increasing the duration to 1.6ms results in a significant drop in accuracy, and at  $D_{t10} = 2\text{ms}$ , the accuracy diminishes further, encompassing a broader spectrum of insignificant signals such as tail resonance or noise. This broader inclusion leads to a notable decrease in accuracy, highlighting the critical importance of selecting an optimal signal duration for accurate classification. In conclusion of the duration study, it is evident that signals with a duration of 1ms contain sufficient information, minimizing the impact of tail noise or insignificant signals, enabling effective damage identification.

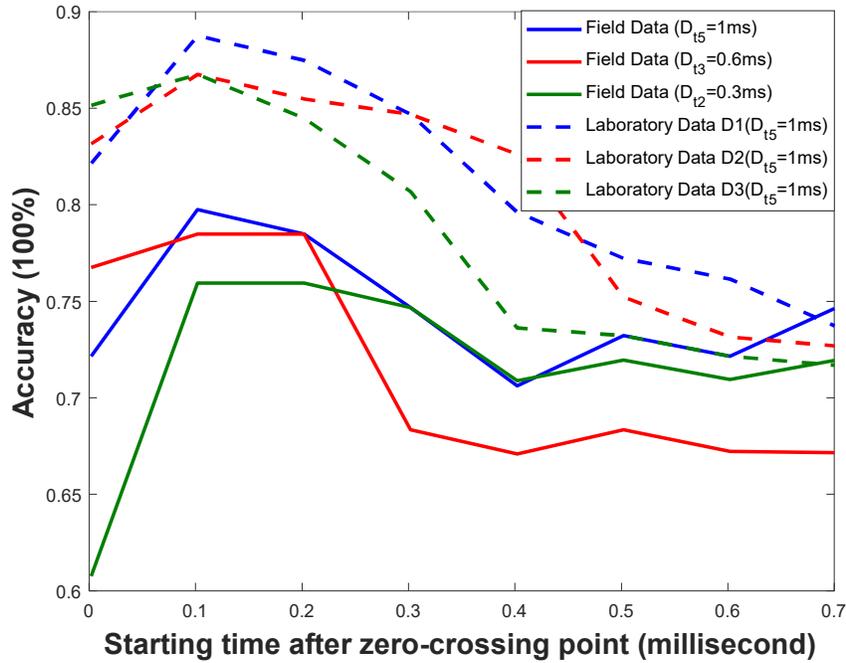


**Figure 4.11.** *The outcomes of the varied signal duration study are evident. Specifically, when duration is 1ms, a significantly higher prediction accuracy is achieved. This is because within the 1ms window, crucial impact signals reflecting structural interior damage are effectively captured. In contrast, when duration=2ms is considered, the accuracy drops significantly. This decrease is attributed to the broader duration including a multitude of insignificant signals, such as tail resonance or noise, resulting in decreased accuracy.*



**Figure 4.12.** The interpretation of signal extraction accuracy with varying duration cases reveals distinct patterns.  $D_{t1}$  and  $D_{t2}$ , characterized by limited signal information, result in lower CNN identification accuracy. In contrast,  $D_{t5}$  exhibits a precise signal group, leading to higher CNN accuracy. However,  $D_{t6}$  includes more signals, including tail insignificant signals, which adversely affect accuracy.

The outcomes from various  $S_{ti}$ , for three different  $D_{ti}$  are depicted in Figure 4.13. A consistent trend is observed across all six cases: the highest accuracy is achieved when  $S_{t2}=0.1\text{ms}$  or  $S_{t3}=0.2\text{ms}$ . This result can be attributed to the exclusion of the front part surface wave, which significantly affects signal behavior. The surface wave occurs exclusively during the first half cycle of the wave, after which the subsequent signal consists entirely of mechanical echoes originating from the combination of surface wave and the P-wave reflections, either from delamination or the solid depth. Conversely, when  $S_{t6}=0.5\text{ms}$ , the accuracy decreases as the signal primarily encompasses tail noise and insignificant signal. Additionally, the accuracy observed in laboratory test data mirrors these trends: the highest accuracy in damage identification occurs when  $S_{t2}=0.1\text{ms}$  or  $S_{t3}=0.2\text{ms}$ . It's worth noting that the laboratory test data, being clearer with distinct target peaks and signals free from noise, validates the accuracy and value of the information derived from the field test data. Despite being trained with field test data, the CNN model's correct identification of laboratory test data underscores the authenticity and importance of the field test findings.

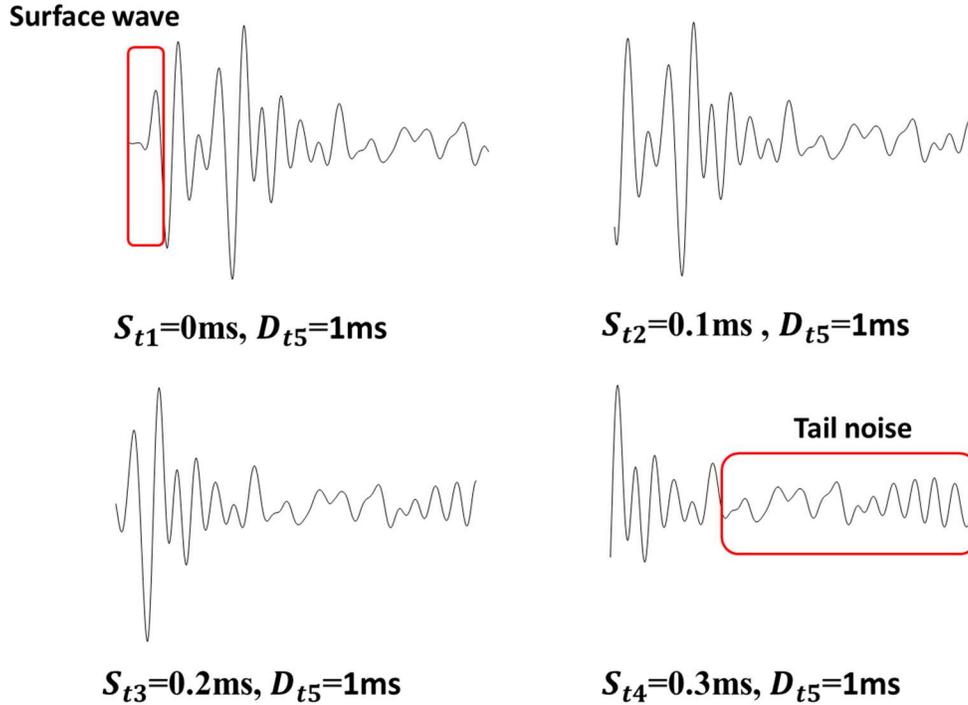


**Figure 4.13.** The outcomes of various signal starting times  $S_{ti}$  across three different durations, including lab test delamination data, exhibit a consistent pattern. In all six scenarios, the highest accuracy is achieved when  $S_{t2}=0.1m$  or  $S_{t3}=0.2ms$ . This heightened accuracy can be attributed to the exclusion of the initial surface wave, a component that significantly influences signal behavior.

The detailed explanation for the CNN accuracy findings can be elucidated using Figure 4.14. In the case where  $S_{t1}=0ms$  and  $D_{t1}=1ms$ , the initial half of the signal primarily comprises a pure surface wave, as indicated in the top red box of Figure 4.14. However, this surface wave does not contribute significantly to the impact echo flexural mode. Across all signals, the surface wave provides similar information, making it a relatively insignificant feature in CNN training. It's noteworthy that the P-wave, which is utilized to identify delamination, occurs after the pure surface wave.

Conversely, when  $S_{t2} = 0.1ms$  and  $S_{t3} = 0.2ms$ , the pure surface wave is eliminated, focusing the CNN on the target signal that is combination of impact echo signal and surface wave while incorporating more tail noise. These signals can be effectively identified by CNN due to the richer information from input data. However, when  $S_{t4}=0.3ms$ , the absence of the target signal results in a decrease in accuracy, particularly

when the input data contains a higher proportion of tail noise, as indicated in  $S_{t4}$  result located in the bottom red box of Figure 4.14.

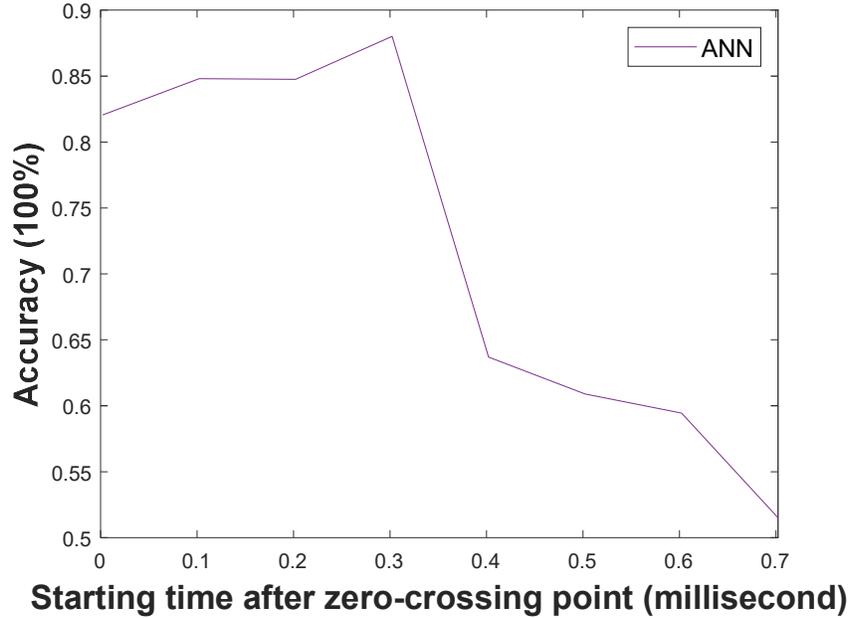


**Figure 4.14.** The explanation of CNN accuracy shown in Figure 4.13. At  $S_{t1}=0\text{ms}$ , both surface wave and target wave are present, but the surface wave offers similar information across all signals, making it less significant for CNN training. When  $S_{t2}=0.1\text{ms}$  and  $S_{t3}=0.2\text{ms}$ , the elimination of the surface wave focuses CNN on the target signal while introducing more tail noise. These signals are effectively identified due to their richer information. However, at  $S_{t4}=0.3\text{ms}$ , the exclusion of the target signal results in decreased accuracy.

The results of ANN using four features processing from Eq. 4.4 to Eq. 4.7 as shown in Figure 4.15..

In this ANN study, the  $D_{t5}=1\text{ms}$  is fixed while varying  $S_{ti}$ . This is because the ANN feature nearly reaches its maximum amplitude when the duration reaches the first peak. Therefore, increasing the duration does not significantly impact the features of the ANN. The observed trend aligns with the findings from the starting time study (Figure 4.14), indicating that  $S_{t2}=0.1\text{ms}$  to  $S_{t4}=0.3\text{ms}$  yield the highest CNN accuracy. In the case of ANN, a sharp decline in accuracy occurs at  $S_{t5}=0.4\text{ms}$  due to the signal being impacted by tail-end noise, mirroring the observations in Figure 4.14. Overall, the

analysis of various starting times reveals that  $S_{t2}=0.1\text{ms}$  and  $S_{t3}=0.2\text{ms}$  offer crucial information for accurate damage identification using impact echo signals. This valuable insight serves as a foundation for enhancing delamination results, a topic explored in the subsequent section.



**Figure 4.15.** The ANN accuracy with four features from Eq. 4.4 to Eq. 4.7. The observed pattern correlates with the results from the starting time study in Figure 4.14, showing that  $S_{t2}=0.1\text{ms}$  to  $S_{t4}=0.3\text{ms}$  provides the highest CNN accuracy. Conversely, for ANN, accuracy sharply decreases at  $S_{t5}=0.4\text{ms}$  due to the signal being influenced by tail-end noise, consistent with the observations in Figure 4.14.

#### 4.4.2 Delamination results from field test

Due to the environmental interferences and machine-related factors in the field test, often obscure delamination signals, leading to challenges in accurate detection. Signals may be either too faint, leading to their dismissal, or excessively amplified, creating false impressions of damage severity. To address this issue, the time-domain signals with varying durations and starting times to mitigate these interferences are systematically examined. In this section, further investigation is performed with a detailed discussion and comparison of the field test delamination results before and after implementing CNN-recommended signal

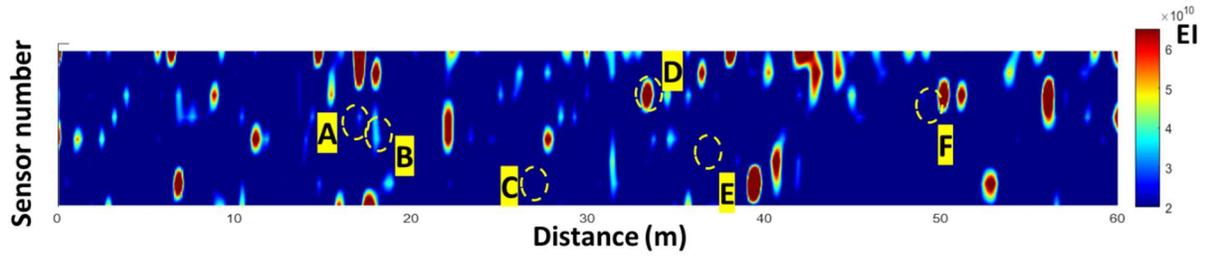
parameters, specifically focusing on starting time and signal duration adjustments. Marked differences are carefully analyzed, and both time-domain and frequency-domain signals are scrutinized to assess the reasonableness of the generated delamination maps.

The delamination maps before and after applying CNN-recommended signal parameters are depicted in Figure 4.16. A careful comparison between Figure 4.16 (a) and Figure 4.16 (b) reveals distinct differences: certain points exhibit altered colors, with some points becoming visible and others disappearing upon the application of CNN-recommended signal parameters. Take Region A, for instance, in Figure 4.16 (a), the accumulated energy between 1kHz to 5kHz is low, suggesting a non-delamination area. However, when plotted using CNN-recommended signal parameters ( $D_{t5}=1\text{ms}$  with  $S_{t2}=0.1\text{ms}$ ) in Figure 4.17 (a), a clear low-frequency peak indicative of delamination becomes visible. This phenomenon can be explained by the original signal in Figure 4.16 (a) having an extended duration, encompassing excessive high-frequency noise that diminishes the influence of low-frequency energy.

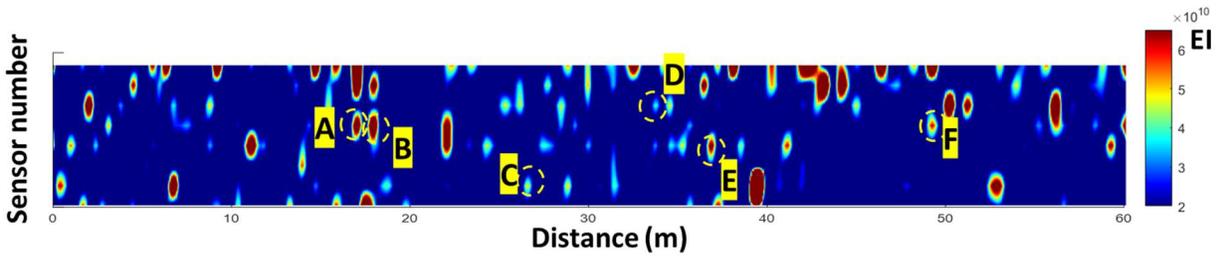
Consider Region C as another example: Figure 4.16 (a) indicates a non-delamination case, whereas Figure 4.16 (b) suggests slight damage in that area. The signal, as depicted in Figure 4.17 (c), showcases a distinct lower-frequency peak for delamination, although the energy level is not notably high. This lower amplitude signal could result from the impactor failing to strike the surface accurately at the correct angle, possibly due to surface conditions like potholes where the impactor may become stuck and vibrate.

Region D offers a valuable insight: Figure 4.16 (a) depicts a large delamination area, while Figure 4.16 (b) indicates minor damage. The signal at Region D, shown in Figure 4.17 (d), demonstrates a peak in a high-frequency region, indicating a highly non-delamination case. This example emphasizes that when the impactor strikes the surface accurately with sufficient energy, the reflected mechanical waves also carry higher energy compared to inaccurate hits. However, higher energy alone is not a sufficient criterion to distinguish between delamination and non-delamination; the specific location of frequency peaks is equally crucial. By analyzing these cases, it becomes evident that the CNN-recommended signal parameters used

to generate Figure 4.16 (b) effectively enhance the delamination maps while mitigating the impact of environmental factors in field tests.

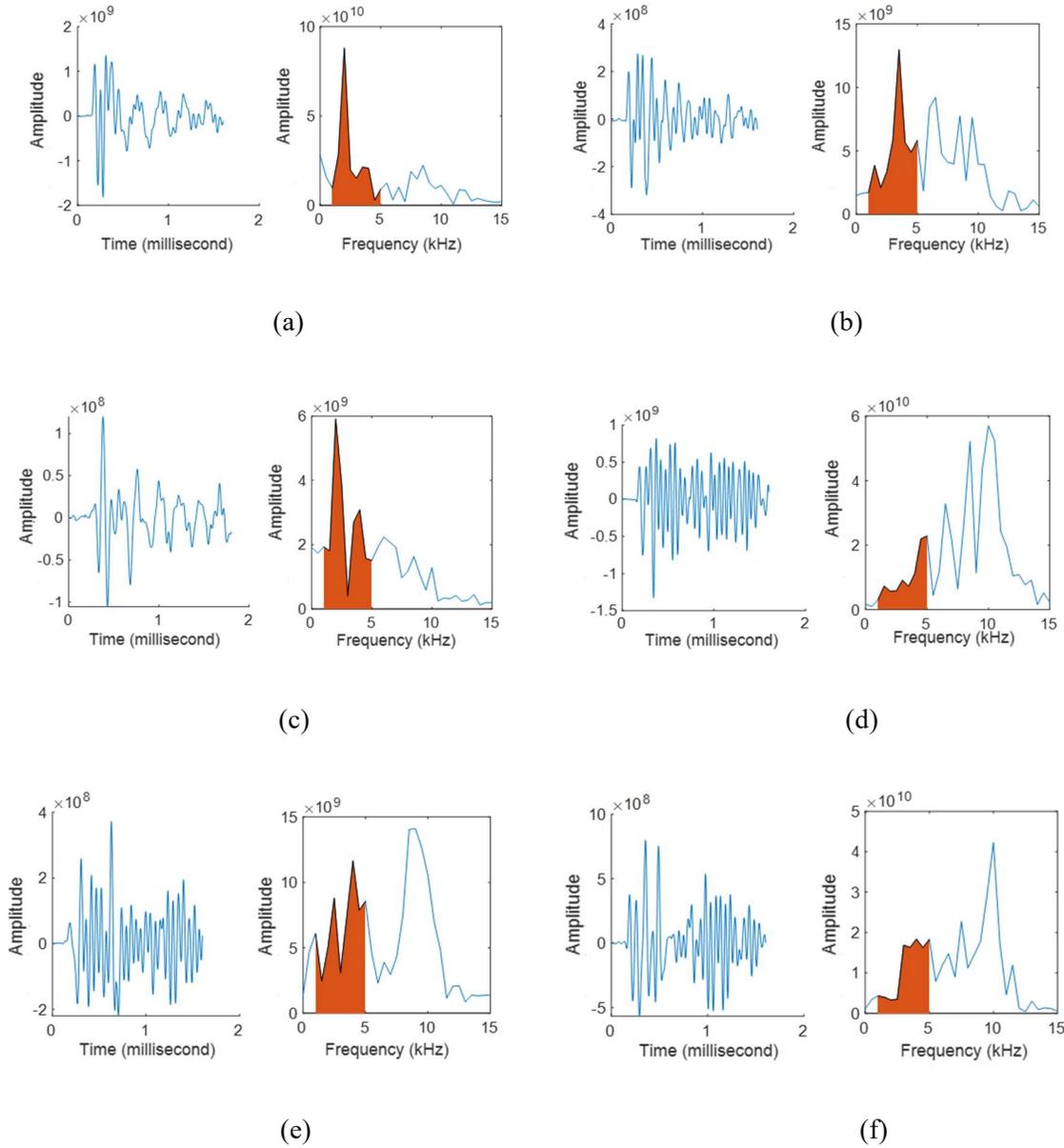


(a)



(b)

**Figure 4.16.** *The delamination maps of scanning path 1. (a) Original algorithm, and (b) CNN-suggestion signal parameter from signal extraction study (e.g.,  $D_{t5}=1ms$  and  $S_{t2}=0.1ms$ ). Consider point C as another example: Figure 4.16 (a) indicates a non-delamination case, whereas Figure 4.16 (b) suggests slight damage in that area. This lower energy could result from the impactor failing to strike the surface accurately at the correct angle, possibly due to surface conditions like potholes where the impactor may become stuck and vibrate.*



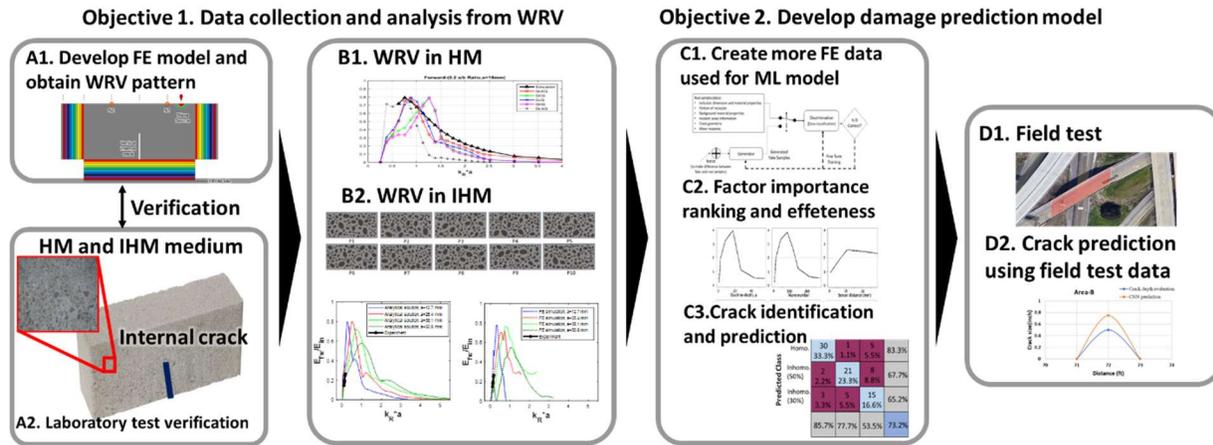
**Figure 4.17.** The verification of Figure 4.16. The samples in Figure 4.16 depict the time-domain and frequency-domain signals for different cases: (a) Region A, (b) Region B, (c) Region C, (d) Region D, (e) Region E, and (f) Region F. Consider sample C as an illustration; despite the relatively low energy, a distinct low-frequency peak indicates a delamination case. Utilizing the CNN-recommended signal parameters depicted in Figure 4.16 (b) results in accurate identification of the delamination in this sample.

## **CHAPTER 5 INTERNAL CRACK IDENTIFICATION IN IHM MEDIUM**

### **5.1 Chapter overview**

The primary goal of this study is to advance the understanding of a WRV is the wave response ratio of scattering waves to incident waves according to specific wave frequency and crack geometry in particular, an internal vertical crack in IHM. Initially, the focus shifted to establishing the mathematical relationship between wave response in HM and analyzing WRV patterns. Given the limited research on WRV in IHM like concrete, the developed FE model from Kang (2021) is applied and conducts laboratory tests with IHM to understand these unique patterns (Kang et al., 2021).

The flowchart in Figure 5.1 outlines the research steps. First, the FE model was applied and verified through laboratory tests with artificial crack specimens. Second, the WRV patterns between HM and IHM were analyzed, considering the effects of aggregate size and distribution. Third, GANs and CNN were employed to develop an internal crack prediction model, which was tested with field test data.



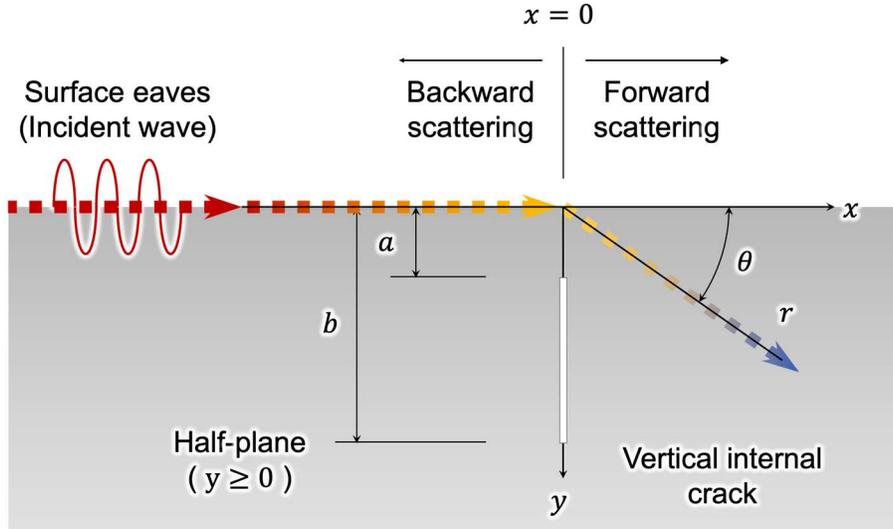
**Figure 5.1. Overview of research steps.** In this study, a sequential research approach was conducted, divided into several tasks. The process began with a comprehensive review of wave response theory, establishing a foundation. Subsequently, the focus shifted to the mathematical relationship between wave response in HM, and WRV patterns were analyzed. Due to limited research on WRV in IHM like concrete, a FE model was developed, and IHM laboratory tests were conducted, confirming the FE model's validity with varied crack geometries. Machine learning techniques were applied to enhance the database and predict crack depth based on understood WRV patterns. Additionally, exploration of GANs was undertaken to generate more WRV data. The study culminated in a field test, evaluating the accuracy of machine learning in identifying damages.

## 5.2 Wave response variation

### 5.2.1 Mathematical formulation with the relation between the wave scattering and various material-crack properties in an elastic medium

The investigation of the analytical solution of WRV in HM using various numerical integration methods is studied (Kang et al., 2020). A vertical internal crack geometry is shown in Figure 5.2. The  $r$  and  $\theta$  are components of the polar coordinate, “a” represents the distance between the surface and the top of the crack tip referred to as depth-to-crack; “b” is crack size. Assuming that the surface wave direction is aligned with the x-axis and the crack is in a vertical direction, the wave split into two parts when it encounters the crack, depending on the wave direction. The wave that propagates forward is referred to as forward scattering, while the reflected wave is called backward scattering. In theory, if the wave energy is not dispersed in the

medium, the sum of the amplitudes of forward and backward scattering should be equal to that of the incident wave (Kang et al., 2021).



**Figure 5.2.** The wave scattering mechanism involves the incident wave energy splitting into backward and forward scattering upon encountering an internal crack. In this context, "r" and "θ" are polar coordinate components, "a" represents the depth-to-crack (distance between the surface and the top of the crack tip), and "b" signifies the crack size.

The displacement potentials are defined in the scattered field and meet the radiation condition of the elliptic boundary-value problem proposed by Wickham (Wickham, 1977).

$$\varphi = A_{\pm} e^{\pm i k_R x - (k_R^2 - k_L^2)^{1/2} y} + A(\theta) (k_L r)^{-\frac{1}{2}} e^{i k_L r} + o((k_L r)^{-1/2}) \quad (5.1)$$

where A constant,  $k_R$ ,  $k_L$ , and  $k_T$  are the wavenumbers of Rayleigh, longitudinal, and transverse waves, respectively. The variable  $r$  and  $\theta$  represent the components of polar coordinates  $(r, \theta)$ . The  $\pm$  signs denote forward direction (+), and backward direction (-). The formula for the constant A is expressed as per Gregory (Gregory, 1975):

$$A_{\pm} = \frac{(2k_R^2 - k_T^2)^2 k_R}{2k_T^2 S} \int_a^b dy \{ \pm d_y(y) P(y) + i d_x(y) Q(y) \} \quad (5.2)$$

where

$$P(y) = 2k_R \left\{ e^{-(k_R^2 - k_L^2)^{1/2} y} (k_R^2 - k_T^2)^{1/2} - e^{-(k_R^2 - k_T^2)^{1/2} y} (k_R^2 - k_L^2)^{1/2} \right\}, \quad (5.3)$$

$$Q(y) = e^{-(k_R^2 - k_L^2)^{1/2} y} \frac{(2k_R^2 - 2k_L^2 + k_T^2)(k_R^2 - k_T^2)^{1/2}}{(k_R^2 - k_L^2)^{1/2}} - e^{-(k_R^2 - k_T^2)^{1/2} y} (2k_R^2 - k_T^2), \quad (5.4)$$

The integration component in (5.2) are:

$$\int_a^b dy \{ \pm d_y(y) P(y) + i d_x(y) Q(y) \} \quad (5.5)$$

and (5.5) can be expressed with integration formula:

$$\int_{-1}^1 f(x) dx \approx \sum_{i=1}^n f(x_i) w_i \quad (5.6)$$

Eq. 5.5 can be applied with various polynomial equations to solve the integration problem. There are five different methods of Gaussian quadrature were considered, including Legendre quadrature (GLEQ), Gauss-Chebyshev quadrature (GCQ), Gauss-Jacobi quadrature (GJQ), Gauss-Hermite quadrature (GHQ), and Gauss-Laguerre quadrature (GLAQ), as outlined in Table 5.1 (Kang et al., 2020). Additionally, FE model to with different WRV patterns under varying crack geometries are developed, and the verification study

between the analytical solutions and the FE model are conducted. Specifically, for this study, the Gauss-Legendre quadrature method is applied.

Table 5.1. Five different integral methods used in WRV analytical solution

Integral method	Weight function
Gauss-Legendre quadrature	$w_i = \frac{2}{(1-x_i^2)[P'_n(x_i)]^2}$
Gauss-Chebyshev quadrature	$w_i = \sqrt{1-x^2}$ .
Gauss-Jacobi quadrature	$w_i = (1-x)^\alpha(1+x)^\beta$ .
Gauss-Hermite quadrature	$w_i = \frac{2^{n-1}n!\sqrt{\pi}}{n^2[H_{n-1}(x_i)]^2}$
Gauss-Laguerre quadrature	$w_i = \frac{x_i}{(n+1)^2[L_{n+1}(x_i)]^2}$

Thus, the expression of incident and forward wave applied Table 5.1 can be (Achenbach & Brind, 1981; Gregory, 1975; Kang et al., 2020):

$$\begin{aligned}
& u_x^{fs} \\
& = \partial \frac{(2k_R^2 - k_T^2)^2 k_R e^{ik_R x - (k_R^2 - k_T^2)^{\frac{1}{2}} y}}{2k_T^2 \left\{ (4k_R^2 - k_T^4)(k_R^2 - k_L^2)^{1/2}(k_R^2 - k_T^2)^{1/2} - 2k_R^4(2k_R^2 - k_L^2 - k_T^2) \right\}} M(y) / \partial x \\
& + \partial \frac{2ik_R(k_R^2 - k_L^2)^{\frac{1}{2}} e^{ik_R x - (k_R^2 - k_T^2)^{\frac{1}{2}} y}}{2k_T^2 \left\{ (4k_R^2 - k_T^4)(k_R^2 - k_L^2)^{1/2}(k_R^2 - k_T^2)^{1/2} - 2k_R^4(2k_R^2 - k_L^2 - k_T^2) \right\}} M(y) / \partial y
\end{aligned} \tag{5.7}$$

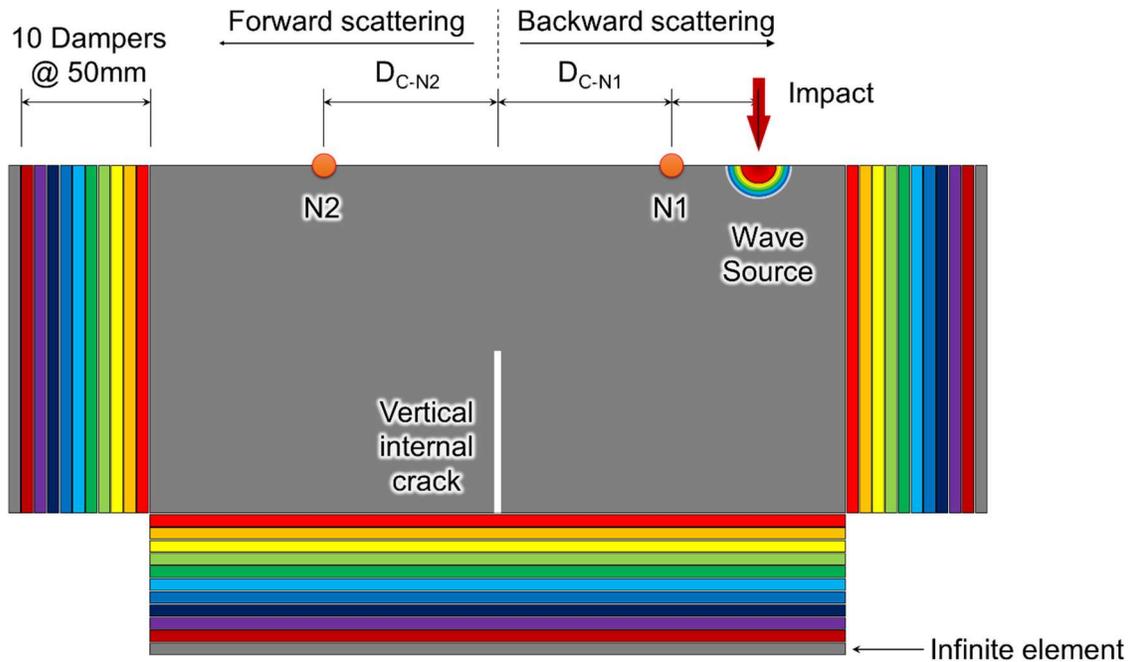
$$M(y) = \int_a^b dy \{ d_y(y)R(y) + id_x(y)Q(y) \} \tag{5.8}$$

$$u_x^{in} = A_R e^{ik_R x} \left\{ e^{-(k_R^2 - k_T^2)^{1/2} y} - \frac{2k_R^2}{(2k_R^2 - k_T^2)} e^{-(k_R^2 - k_L^2)^{1/2} y} \right\} \quad (5.9)$$

where  $u_x^{in}$  is incident surface wave displacement;  $u_x^{fs}$  is forward wave displacement;  $A_R$  is the wavelength of the incident Rayleigh wave.  $K_R$  is the wave-number of a Rayleigh wave. The displacement variation ratio ( $u_x^{fs}/u_x^{in}$ ) are presented by the incident wave frequency and crack geometry variables,  $k_R a$ , representing the energy diffusion before and after encountering the crack.

### 5.2.2 Comparative study of wave response variation in different geometry sizes using an infinite element in homogeneous medium

Figure 5.3 displays a representative example of the FE model utilized in this study, preprocessing through the HyperMesh software, and solved using Abaqus. To mitigate noise and reflected waves from the boundaries, an infinite element and dampers were incorporated (Oh et al., 2012). Ten layers of dampers were implemented, each possessing a distinct damping value. The red damper had the lowest damping value (1000) as the initial damping factor, with an incremental damping factor of 5000 for the subsequent layer. The infinite element was positioned at the last damper (black), as outlined in Table 5.2. The utilization of ten layers with varying damping values ensured that wave responses, such as amplitude or waveform, did not undergo drastic changes. This approach guaranteed the accurate preservation of the wave shape for comprehensive analysis. The test specimen was assumed to possess typical solid material properties ( $\rho=2400 \text{ kg/m}^2$ ,  $E=35 \text{ GPa}$ , and  $\nu=0.2$ ). The model for the ABAQUS solver was constructed using a two-dimensional (2-D) four-node plane strain element (CPE4R) with a 2 mm mesh size for the solid medium group and dampers, except for the infinite element (CINPE4). The input pulse employed was a three-cycle sine wave with frequencies ranging from 10kHz to 200kHz.

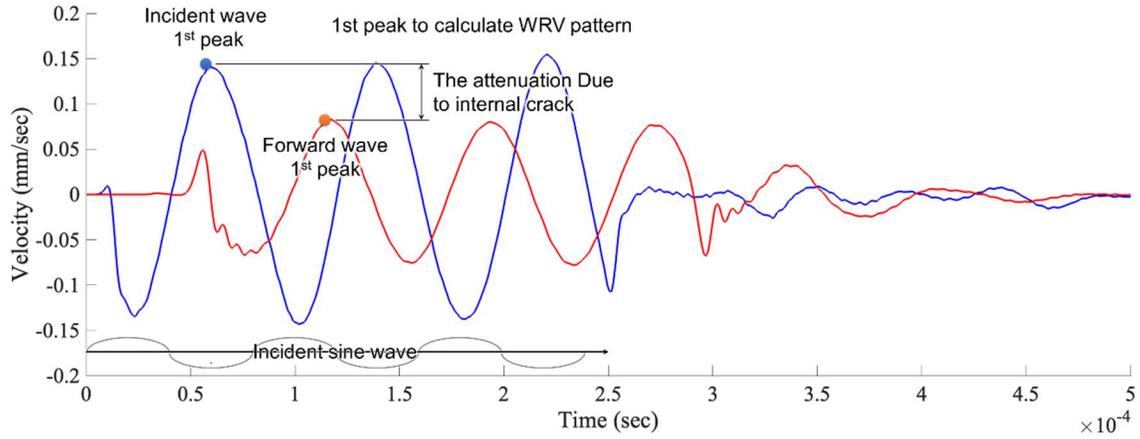


**Figure 5.3.** This FE model example includes listening nodes N1 and N2. N1 captures incident and reflected waves, while N2 captures forward waves. Ten layers of dampers are incorporated to minimize boundary reflections.

**Table 5.2.** The damper coefficient is denoted by Damper 1 for the inner one attached to the specimen, while Damper 10 refers to the outer damper.

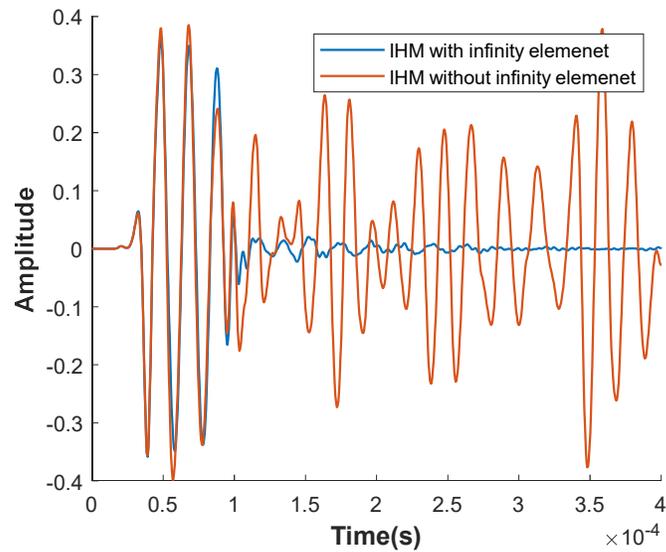
Damper 1	Damper 2	Damper 3	Damper 4	Damper 5	Damper 6	Damper 7	Damper 8	Damper 9	Damper 10
1000	6000	11000	16000	21000	26000	31000	36000	41000	46000

The WRV pattern is computed by considering the incident and forward/backward amplitudes. Figure 5.4 illustrates an example of N1 listening points along the x-axis in Figure 5.3. The wave with a higher amplitude represents the incident wave (blue), whereas the wave with a lower amplitude corresponds to the forward wave (red). The scattering amplitude is reduced due to energy being transferred to backward scattering and diffusing within the medium. To validate the FE model simulation against the analytical solution, the wave response pattern is calculated using the first peak of the wave amplitude, its corresponding wavelength, and crack depth for comparison.



**Figure 5.4.** An example of the incident wave and forward wave in the laboratory test is shown in Figure 5. The incident wave (blue solid line) has a higher amplitude, while the forward wave (red solid line) has a lower amplitude due to the presence of a crack.

The signal reflection validation from the IHM model, applied with and without the infinity element, is shown in Figure 5.5. The model dimensions are the same as the laboratory specimen size. When comparing the two signals, the infinity element signal exhibits less reflection in the tail part, while the initial front peak is not significantly affected by reflection. A similar waveform can be observed in the model without the infinity element, where the front waveform matches the infinity element model, demonstrating that the front signal is not significantly affected by reflection. The WRV is calculated from the first peak of the incident wave and forward scattering wave.



**Figure 5.5.** The reflection validation from FEM. The model with the infinity element (blue) and without the infinity element (orange) experiences less reflection in the front first peak. The WRV is calculated from the first peak of the incident wave and forward scattering wave.

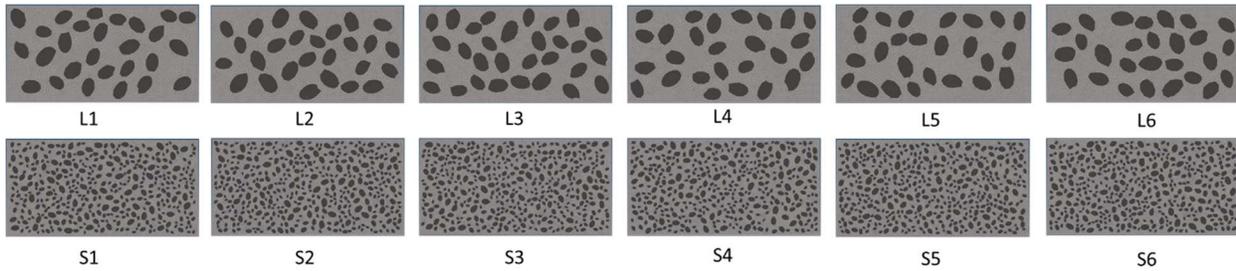
### 5.2.3 Develop the damage prediction model using ML technology

To gather sufficient data for damage identification in the later stages of the study, creating numerous specimens with varying depths of cracks in the laboratory within the given project timeframe can be challenging. To address this limitation, FE models are employed to generate additional data. The IHM design closely resembles that of the HM, with the primary distinction being the use of random aggregates in the IHM specimens. The method for creating random aggregates has been validated in previous research studies (Yun et al., 2020). The study explores different aggregate sizes with the same portion ratio, as depicted in Figure 5.6. The larger aggregates are denoted as L1-L6, and the smaller aggregates are denoted as S1-S6. L1-L6 represent exclusively utilize larger aggregates, while S1-S6 represent only smaller aggregates but with a random distribution. Additionally, ten different aggregate patterns are investigated illustrated in Figure 5.7. These patterns encompass varying numbers and sizes of aggregates, as detailed in

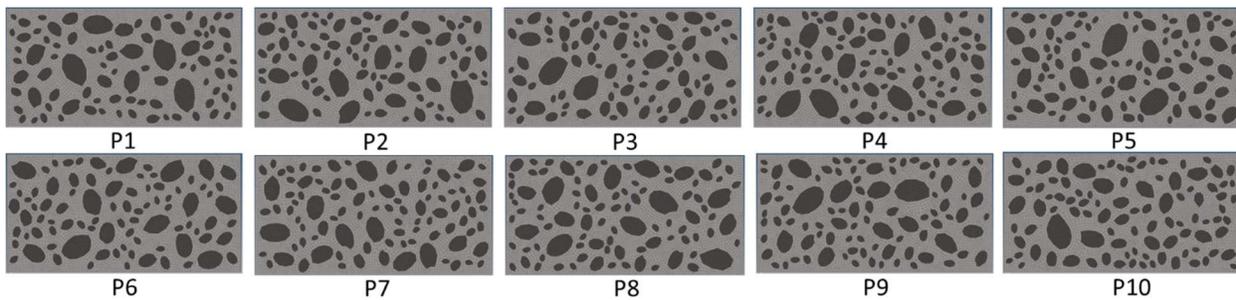
Table 5.3. By comparing the waveforms and WRV patterns, the differences and effects of different aggregate configurations can be studied effectively.

**Table 5.3. The aggregate detail of ten patterns**

Cases	Aggregate portion (%)	Total number $0 < \text{dia.} \leq 4$	Total number $4 < \text{dia.} \leq 7$	Total number $7 < \text{dia.} \leq 10$	Minimum dia.	Maximum dia.
L1	28.12	0	0	100	7.02	9.58
L2	28.54	0	0	100	7.11	9.43
L3	28.26	0	0	100	7.15	9.68
L4	27.91	0	0	100	7.18	9.34
L5	27.27	0	0	100	7.05	9.47
L6	28.66	0	0	100	7.07	9.51
S1	26.51	100	0	0	1.17	3.63
S2	26.22	100	0	0	1.14	3.70
S3	25.91	100	0	0	1.16	3.63
S4	25.93	100	0	0	1.14	3.55
S5	26.18	100	0	0	1.13	3.57
S6	27.37	100	0	0	1.14	3.65
P1	29.27	53	28	7	2.27	11.31
P2	29.84	62	22	9	2.36	11.68
P3	29.22	59	29	7	2.27	11.12
P4	30.44	67	24	8	2.35	11.62
P5	29.59	59	30	6	2.41	11.62
P6	31.81	53	22	13	2.41	11.00
P7	30.04	62	24	11	2.35	10.02
P8	31.59	66	18	11	2.37	10.36
P9	30.39	61	22	10	2.29	11.69
P10	29.74	60	34	5	2.41	11.69



**Figure 5.6. Different aggregate size study. Two study groups were defined: L1-L6, consisting of larger aggregates exclusively, and S1-S6, comprising smaller aggregates with a random distribution. These cases maintained a similar aggregate portion ratio.**

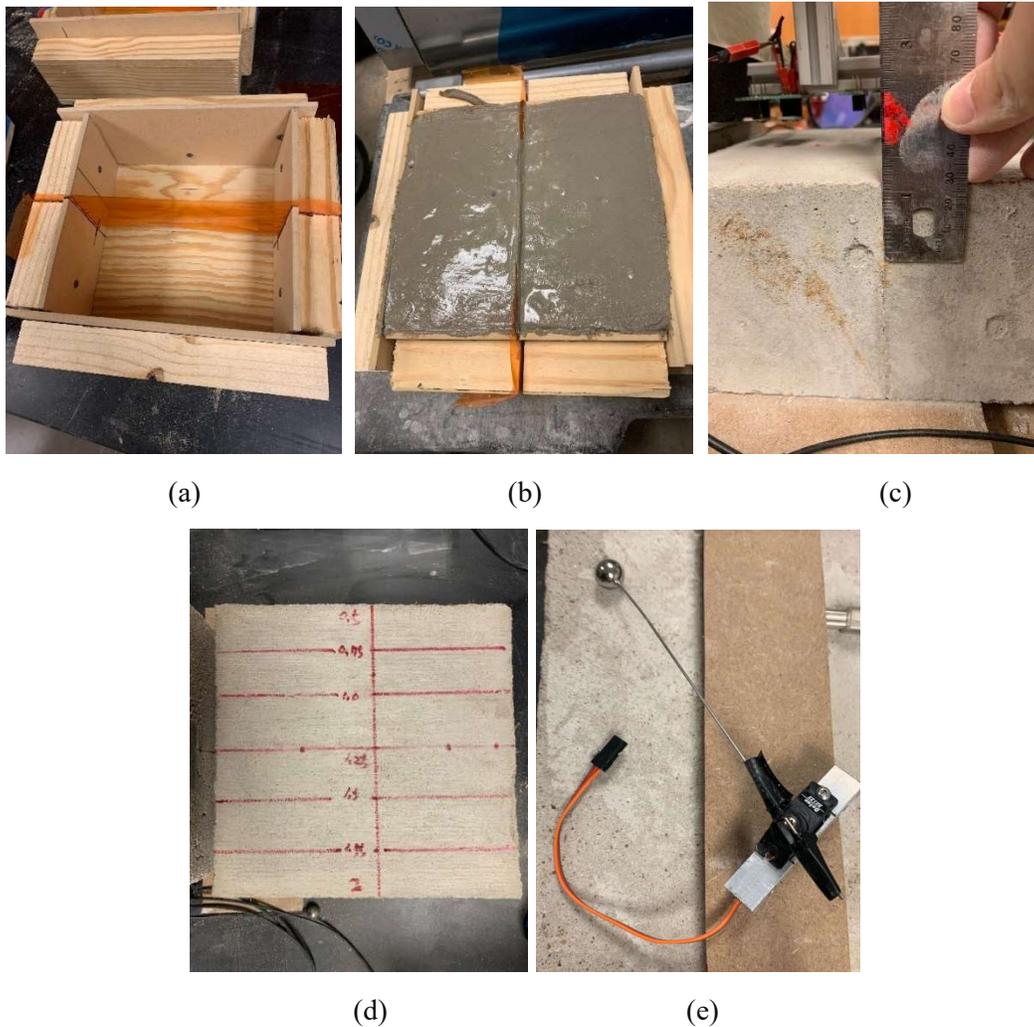


**Figure 5.7. Ten patterns were created using random aggregate sizes and distributions, offering diverse designs for the study.**

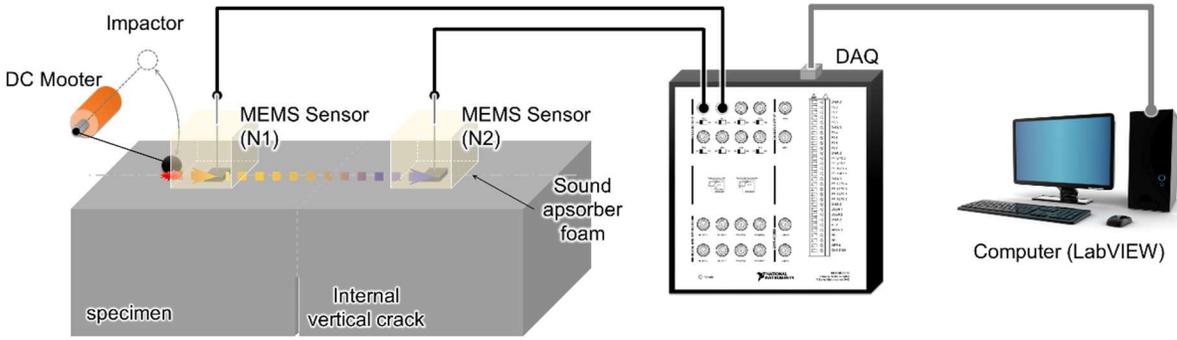
#### 5.2.4 Laboratory test

The study of IHM, particularly the WRV phenomenon, has been a challenging endeavor due to its intricate nature influenced by diverse factors such as material properties, wave frequency, crack size and shape, and environmental conditions. To address these complexities, laboratory tests are used to validate our IHM FE model. In the laboratory, an 8 x 8 x 4-inch specimen container with a water-to-cement (w/c) ratio of 0.5 is utilized. An artificial crack was created using a 0.5-millimeter-thick plastic shim, which was removed after 2 hours of mixing to achieve the desired conditions based on the laboratory environment. Four types of crack specimens were generated: 1) mortar with a constant 1-inch depth-to-crack, 2) mortar with a diagonal depth-to-crack ranging from 0.5 to 2 inches, 3) concrete with a constant 1-inch depth-to-crack, and 4) concrete with a diagonal depth-to-crack ranging from 0.5 to 2 inches. Figure 5.8 illustrates the process of

creating these artificial cracks and the final specimens. Impact tests were conducted using a 0.5-inch impact ball at a frequency of approximately 12 kHz (Kang et al., 2022). The testing setup, depicted in Figure 5.9, utilized two MEMS sensors as receivers and one DC motor to generate the mechanical wave.



**Figure 5.8. Artificial internal crack samples. (a) the specimen container with dimensions 8x8x4 inches; (b) the fresh concrete mixture; (c) a model with a constant depth-to-crack of 1 inch; (d) a top-view of diagonal depth-to-crack ranging from 0.5 inch to 2 inches; and (e) the DC motor used in the testing setup. These components were essential in simulating and studying crack patterns within IHM.**



**Figure 5.9.** *The laboratory test setup involved the use of a DC motor to generate mechanical waves, while MEMS sensors were employed to capture the incident wave, reflected wave, and forward wave.*

### 5.3 Develop the damage prediction model

#### 5.3.1 *Develop the damage prediction model using ML*

To acquire a sufficient dataset, the GANs framework involves training two models simultaneously: a generator that produces synthetic data and a discriminator that assesses the authenticity of the generated data. In this adversarial training, the generator aims to deceive the discriminator, while the discriminator endeavors to distinguish between real and synthetic data (Goodfellow et al., 2014).

The nonlinear modal analysis of WRV in IHM presents challenges due to limited understanding of its characteristics, including random scattering, variations in internal crack size and wavelength, wave energy attenuation due to IHM damping, and environmental factors like temperature. To address these complexities, two machine learning methods were employed in this study. The first method utilized GANs to generate valuable IHM scattering pattern data. The second method focused on developing a damage identification model using CNN, aiming to enhance the accuracy of damage detection and characterization. GANs, a deep learning architecture, consist of two neural networks: a generator network (G) and a discriminator network (D). The generator aims to produce synthetic data ( $x'$ ) that closely resembles real data ( $x$ ), while the discriminator attempts to distinguish between the real and synthetic data generated by

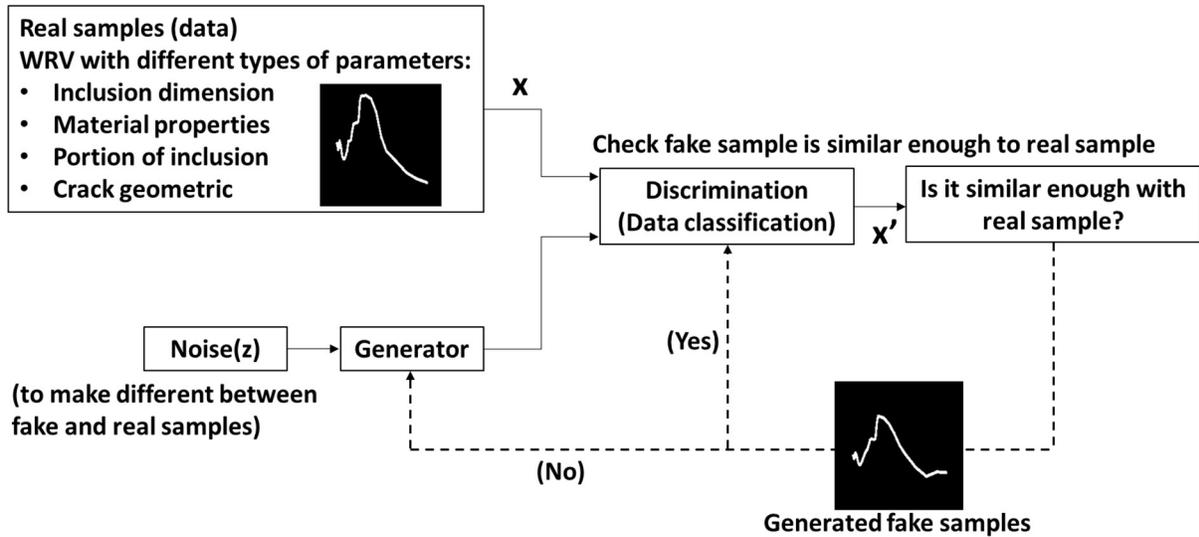
the generator. The training process operates as a minimax game between the generator and discriminator, with the objective function  $J(G, D)$  defined as follows:

$$J(G, D) = \min \max V(D, G) \quad (5.10)$$

where  $V(D, G)$  is the value function defined as:

$$V(D, G) = E_x[\log D(x)] + E_{x'}[\log (1 - D(G(x')))] \quad (5.11)$$

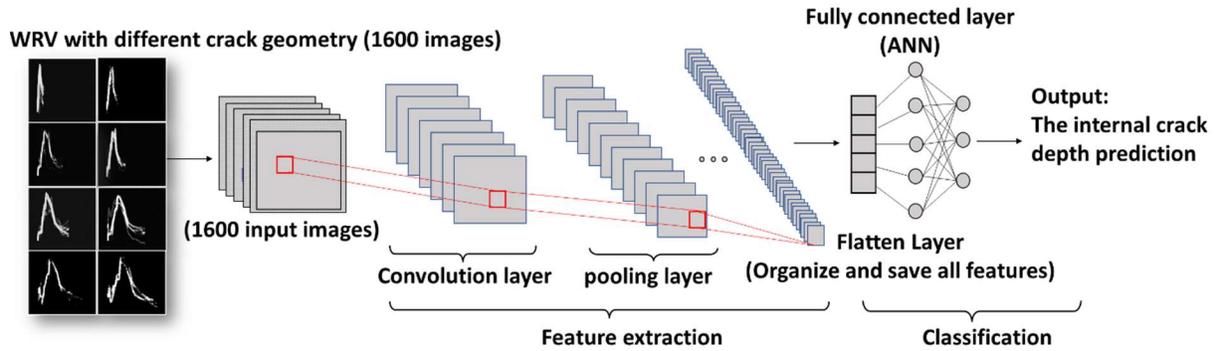
where  $E$  is loss value from real data  $x$  and fake data  $x'$ . The GANs training process involves iteratively training both the generator and discriminator. In each iteration, the discriminator is updated to maximize the value function, while the generator is updated to minimize it. The training continues until the generator produces synthetic data that is indistinguishable from the real data, indicating that the GANs has converged to an optimal solution. A concept of GANs architecture showing in Figure 5.10. A generator network  $G$  and a discriminator network  $D$ , with the generator taking a random noise vector  $z$  as input and producing synthetic data  $x'$ , which is passed through the discriminator to produce a binary classification output. The generator and discriminator are trained in an adversarial manner, where the generator tries to produce synthetic data that resembles the real data, and the discriminator tries to distinguish between the real data and the synthetic data.



**Figure 5.10.** The concept of GANs model. A generator network (G) and a discriminator network (D) operate in tandem, with the generator taking a random noise vector ( $z$ ) as input and generating synthetic data ( $x'$ ). The generator and discriminator are trained adversarially: the generator strives to create synthetic data resembling real data, while the discriminator aims to differentiate between real and synthetic data. The input and generated data in this study is WRV pattern.

After generating new WRV samples using GANs, the CNN algorithm is utilized to identify the WRV pattern and predict internal crack depth. CNN extracts feature from convolution and pooling layers, which are employed to train the model. The neural network processes these features and calculates classification and prediction results, as illustrated in Figure 5.11. The input for the internal crack prediction CNN model comprises WRV curves calculated from wave incident and forward energy, while the output includes internal crack geometry and depth. Furthermore, this study explores the influence of the quantity of synthetic data samples on improving the accuracy of CNN-based crack prediction. Six distinct CNN models were implemented using varying numbers of synthetic samples. The initial model was trained on 40 images featuring patterns P1 to P5 in the IHM dataset, while another 40 images encompassing patterns P6 to P10 were used for training. The subsequent five models incorporated additional synthetic samples for both training and testing, building upon the first CNN model as outlined in Table 5.4. The limitation of CNN lies in its lack of interpretability, a challenge that can be partially addressed by visualizing the impact of features on the response. This visualization helps reveal patterns of effects at local levels for individual features'

values. Accumulated local effect (ALE) plots serve as a valuable tool for visually assessing the influence of each feature, taking into account the effects of other features.



**Figure 5.11.** The CNN model comprises a convolution layer for feature extraction and a pooling layer to reduce dimensions, retaining essential features. Extracted features are connected to a flatten layer and a neural network for training and testing purposes.

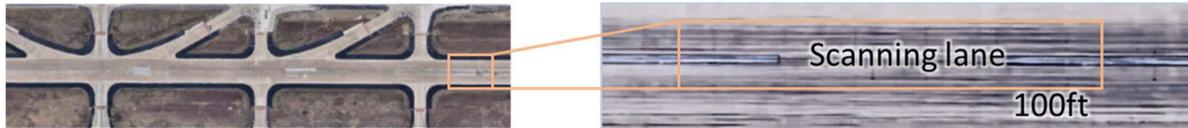
**Table 5.4.** The training sample and testing sample of CNN with different number of GANs samples.

Model No.	Real train samples (from FEM)	Fake train samples (from GANs)	Total train samples	Test samples
1	40	0	40	40
2	40	40	80	80
3	40	120	160	160
4	40	200	200	240
5	40	400	440	440
6	40	1080	1120	480

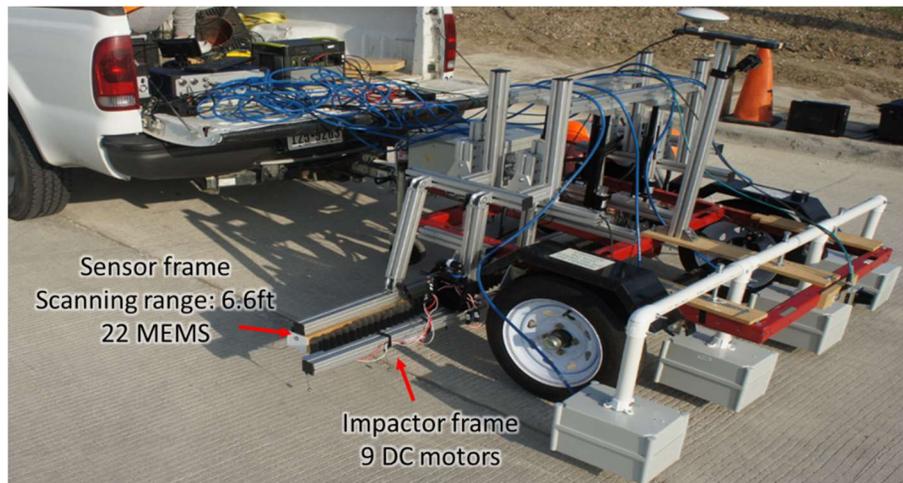
### 5.3.2 Field test verification

The field test aims to validate the CNN model trained on FEM simulation results of both HM and IHM. It provides a real-world context, allowing us to understand how realistic internal crack patterns differ from ideal FEM results. Our focus is on developing a damage identification model using data from this test, analyzing WRV differences between HM and IHM with varying crack geometry (Kang et al., 2022).

Conducted at an airport runway, Figure 5.12 and Figure 5.13 depict the scanning location and system setup. The collected data is processed using our model, enabling a comprehensive analysis to assess its accuracy in identifying damage and defects in the runway.



**Figure 5.12.** *The field test is conducted on an airport runway, where the bottom layer consists of a concrete slab with potential internal cracks. This scenario raises concerns about the presence of hidden defects within the structure.*



**Figure 5.13.** *The ACES system configuration cited from (Kang et al., 2022).*

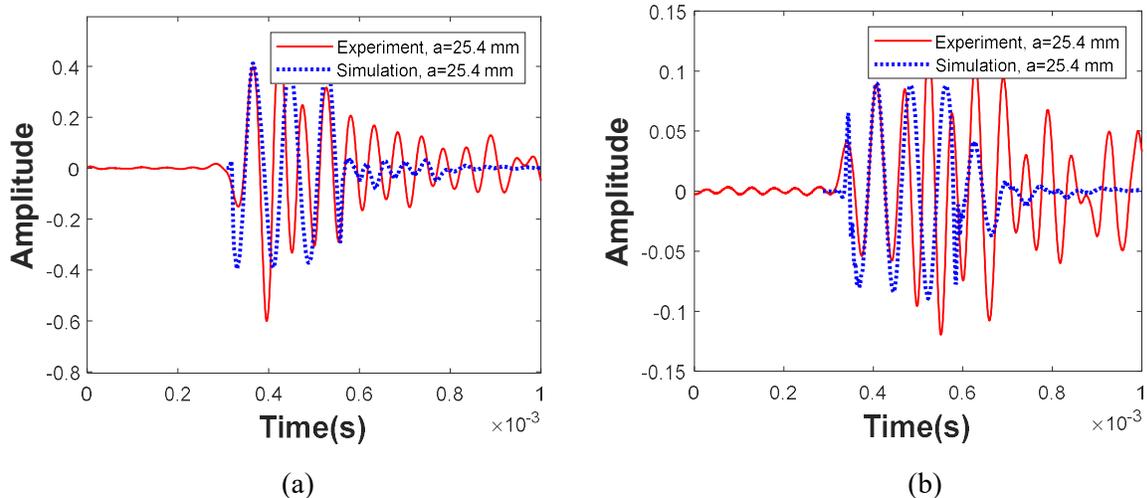
## 5.4 Result and discussion

This section is divided into three parts that discuss the results obtained. The first part presents the results of the WRV pattern in HM, which includes the experimental findings, FE simulation, and analytical solution. The second part deals with the WRV pattern in IHM, for which only the experimental and FE simulation comparisons are available since there is currently no analytical solution. Additionally, the study examines the WRV pattern in IHM with different aggregate sizes with a random distribution. The third part explores ML applications, specifically the approach used to enhance the database with GNAs and the use of CNN

model for damage prediction.

#### 5.4.1 *WRV study in homogenous medium*

The analytical solution offers a clear signal devoid of laboratory measurement interferences, making it challenging to generate a diverse set of crack specimens to establish a comprehensive database for ML damage prediction. Consequently, an efficient approach is adopted, wherein a FE model is utilized to generate this database. The validation of the FE model is conducted through laboratory tests and analytical solutions. In Figure 5.14, a comparison between the FE simulation and experimental data is presented. As the WRV is computed from the first peak ratio of the incident wave and the forward wave, it is crucial to validate the ratio between experimental and FE data. The experimental data exhibits reflections in the tail part, which do not impact the WRV value. Figure 5.14 (a) illustrates the incident wave, while Figure 5.14 (b) depicts the forward wave response. For the FE simulation, the impulse frequency was determined based on the impact frequency (12 kHz) and the impact size. In Figure 5.14 (a), the experimental results (represented by the red solid line) display more noise and wave reflection from the bottom of the specimen, whereas the simulation signal (indicated by the blue dotted line) demonstrates a significant impulse. The disparity between the experimental and simulation signals in Figure 5.14 (a) arises because the FE simulation incorporated infinite elements, gradually decreasing the reflected signal and resulting in a cleaner waveform. The WRV value ratio between the experiment and FE simulation is consistent, approximately around 0.15, affirming the suitability of the FE model for generating subsequent WRV data.

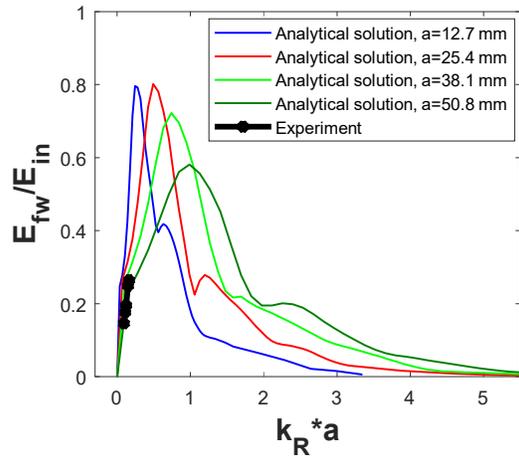


**Figure 5.14.** The comparison between experiment and FE simulation. (a) incident wave, and (b) forward wave.

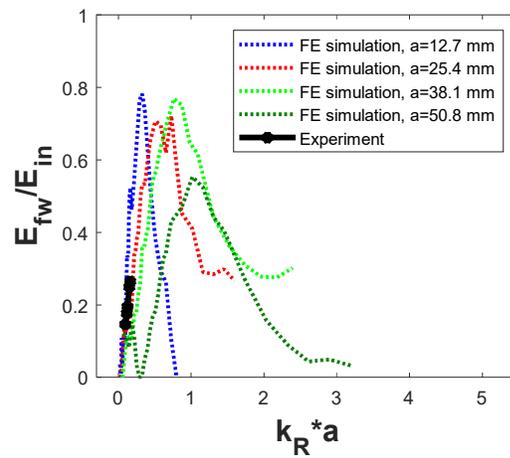
Figure 5.15 presents a comparison of WRV patterns between FE simulation and experimental data, along with a comparison between the analytical solution and the FE model. On the y-axis, the normalized forward wave is depicted, representing the ratio of the first peak by dividing the forward wave  $E_{fw}$  amplitude by the incident wave amplitude  $E_{in}$ . The x-axis represents the product of crack depth, "a" and wavenumber,  $k_R$ . In Figure 5.15 (a) and Figure 5.15 (b), the WRV patterns from analytical, FE model, and experimental data are displayed. However, the experimental data only covers a limited range of the x-axis due to the fixed testing impact frequency (12 kHz) and the specimen "a" ranging from 0.5 inches to 2 inches. To incorporate cases with deeper cracks, enlarging the specimen by increasing its length and width becomes necessary. However, this approach is less economical and involves a slower process of experimental preparation. Moreover, the reflected signal from the boundary may interfere with the target signal, posing challenges in current WRV experiments.

Figure 5.15 (a) and Figure 5.15 (b) illustrate the WRV patterns for different crack depths. The shallower crack case ( $a = 12.7$  mm) exhibits a peak at a lower impulse frequency, while the deeper crack ( $a = 50.8$  mm) shows the WRV peak at a higher frequency. This variation occurs because different impulse frequencies have varying abilities to propagate through deeper mediums, influenced by their wave speed and

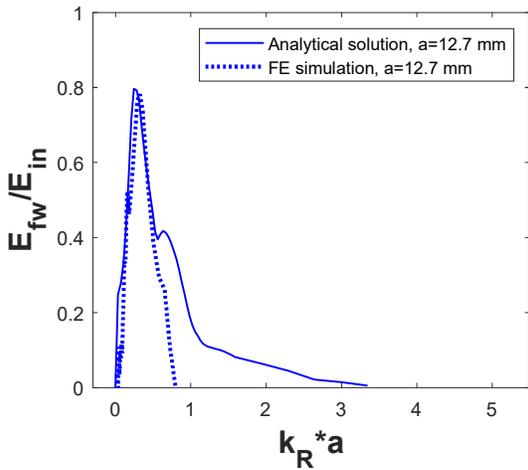
wavelength. Figure 5.15 (c) to Figure 5.15 (f) compare the analytical solution and FE simulation for different crack depths. All these figures demonstrate similar WRV trends, including the peak occurring within a specific frequency range and comparable impulse peak amplitudes. It is important to note that there is no manual adjustment of peak values (y-axis) between the FE model and analytical solution, ensuring the accuracy of the comparison.



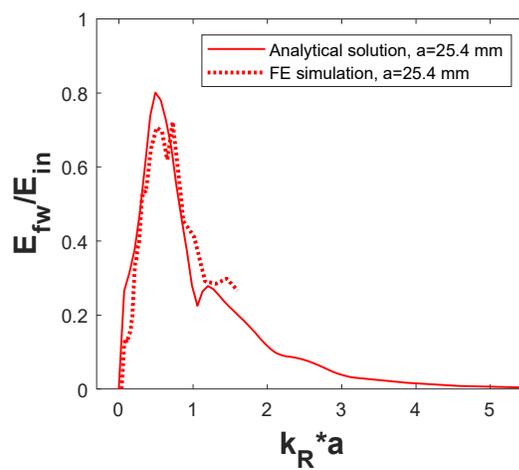
(a)



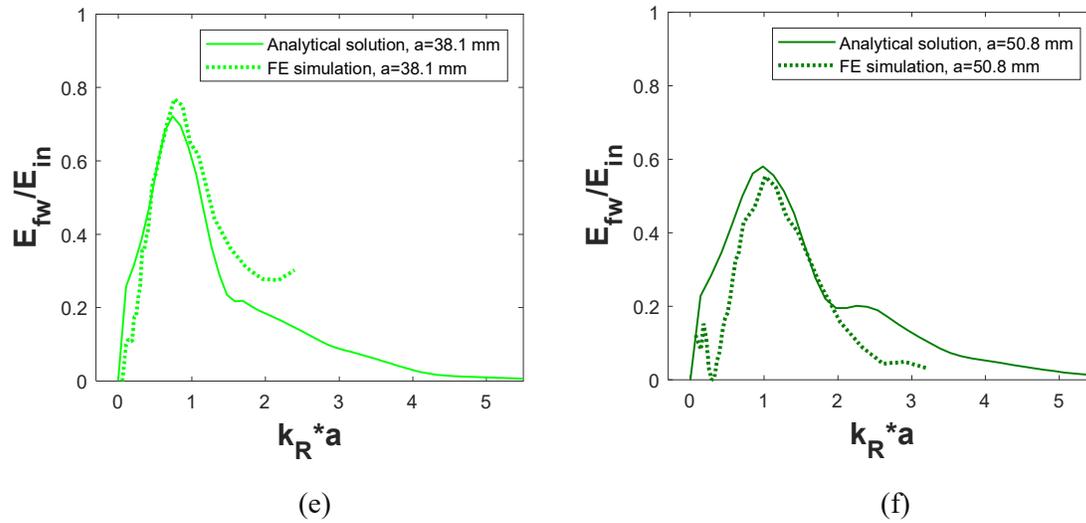
(b)



(c)



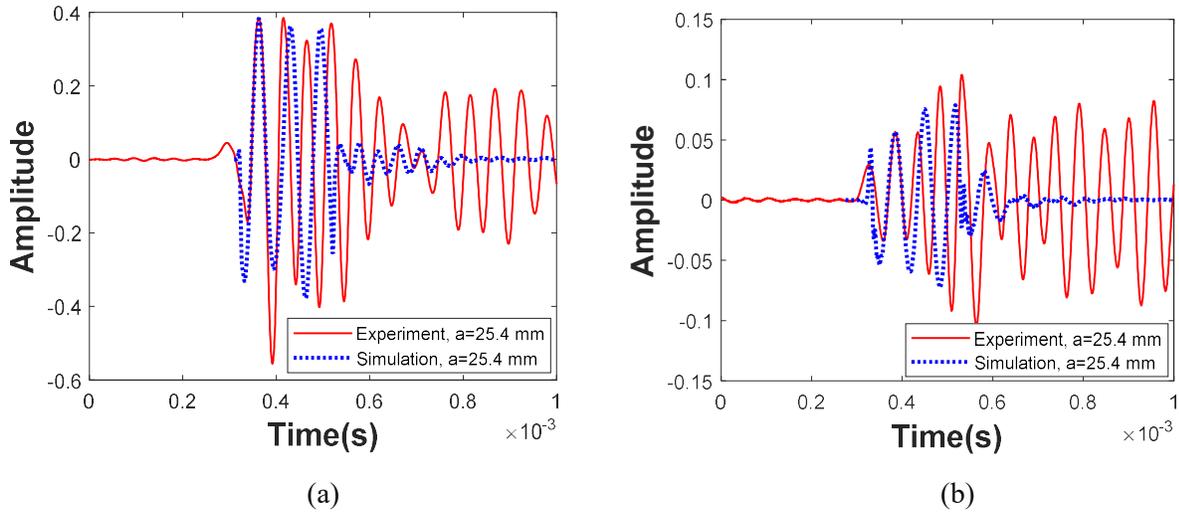
(d)



**Figure 5.15.** Normalized WRV of FE simulation and analytical solutions in HM. On the y-axis, the normalized forward wave is depicted, representing the ratio of the first peak by dividing the forward wave  $E_{fw}$  amplitude by the incident wave amplitude  $E_{in}$ . The x-axis represents the product of "a" and wavelength. (a) the analytical solution results with different a. (b) the FE simulation results with different a. (c)-(f) shows the comparison between FE simulation and analytical solution with different a. The shallower crack case ( $a=12.7\text{mm}$ ) exhibits a peak at a lower impulse frequency, while the deeper crack ( $a=50.8\text{mm}$ ) shows the WRV peak at a higher frequency. This variation occurs because different impulse frequencies have varying abilities to propagate through deeper mediums, influenced by their wave speed and wavelength.

#### 5.4.2 IHM numerical simulation model verified with laboratory test

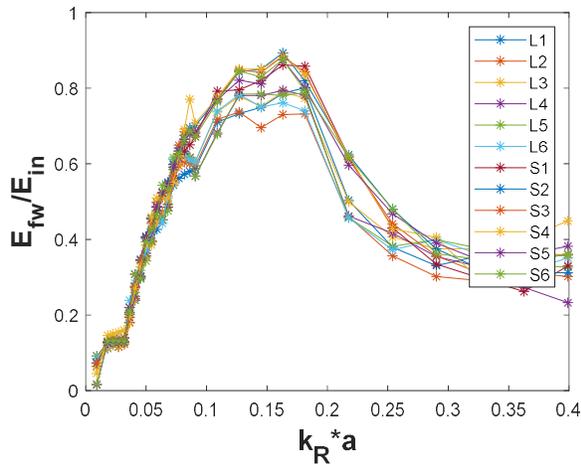
Figure 5.16 presents a comparison of sensing data between the experiment and FE simulation in IHM. The ratio of the incident wave to the forward wave in both the experiment and FE simulation is approximately 0.11. Compare with HM sensing data (Figure 5.14), IHM exhibits more reflected signals, signifying a more intricate interaction between random aggregates and their random distribution within the medium.



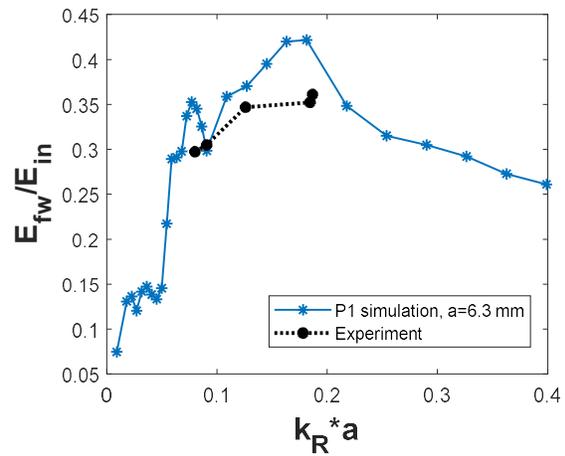
**Figure 5.16.** The example of sensing data comparison between experiment and simulation. (a) the incident wave, and (b) the forward wave.

Figure 5.17, the study investigates different aggregate sizes with random distribution, utilizing the models introduced in Table 5.3. The WRV results, depicted Figure 5.17 (a), reveal a slightly lower amplitude for the large aggregate group (L1-L6) compared to the small aggregate group (S1-S6). This difference may stem from larger aggregates dissipating more energy, causing the wave to be influenced by the aggregate size before passing through completely. The random distribution of aggregates further impacts WRV, evident in the varied amplitude of L1-L6, despite having a similar aggregate portion with the same size. However, organizing the diverse patterns among different aggregate sizes and their distributions proves challenging. Figure 5.17 (b) illustrates the comparison between experimental and simulation results. The limited number of "a" specimens and a fixed impact frequency in laboratory tests resulted in a restricted range of results. Nevertheless, these outcomes still reflect WRV trends, affirming the feasibility of the FE model. Figure 5.17 (c) and (d) present box plots for large and small aggregate groups, respectively. These plots calculate the range of one standard deviation between each WRV for various  $k_R$  values. In Figure 5.17 (c), WRV for low  $k_R$  values exhibit similar amplitudes, becoming more diverse at higher  $k_R$  values. This variation might be attributed to the lower frequency impulse's initial peak having a lower speed to pass through larger aggregates, while the signal's tail part has not yet entered the aggregate. Conversely, higher

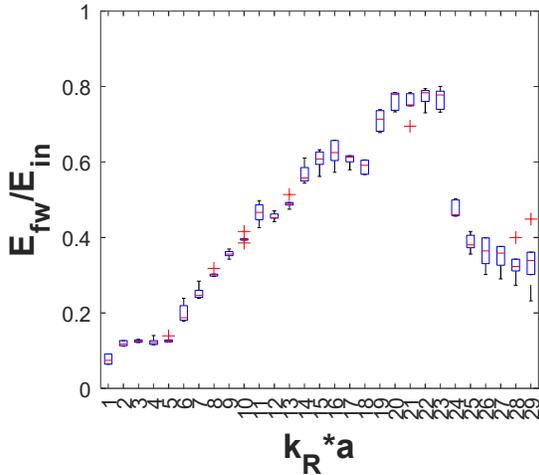
frequency impulses possess greater speed, and the three-cycle pulse wave is simultaneously affected by the aggregates, resulting in a more complex reflection. A similar trend is observed in the small aggregate group in Figure 5.17 (d), where higher impulse frequency leads to a more varied amplitude, commencing from lower frequency ranges due to the smaller aggregate size. Figure 5.18 compares WRV trends between HM and IHM for the same crack case, demonstrating similar patterns. However, identifying the parameters and their impact on the WRV phenomenon remains challenging.



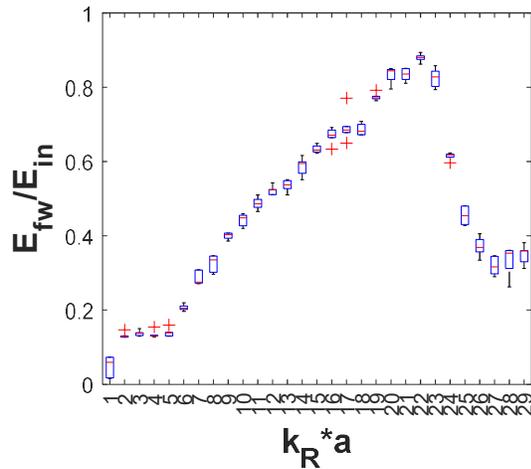
(a)



(b)

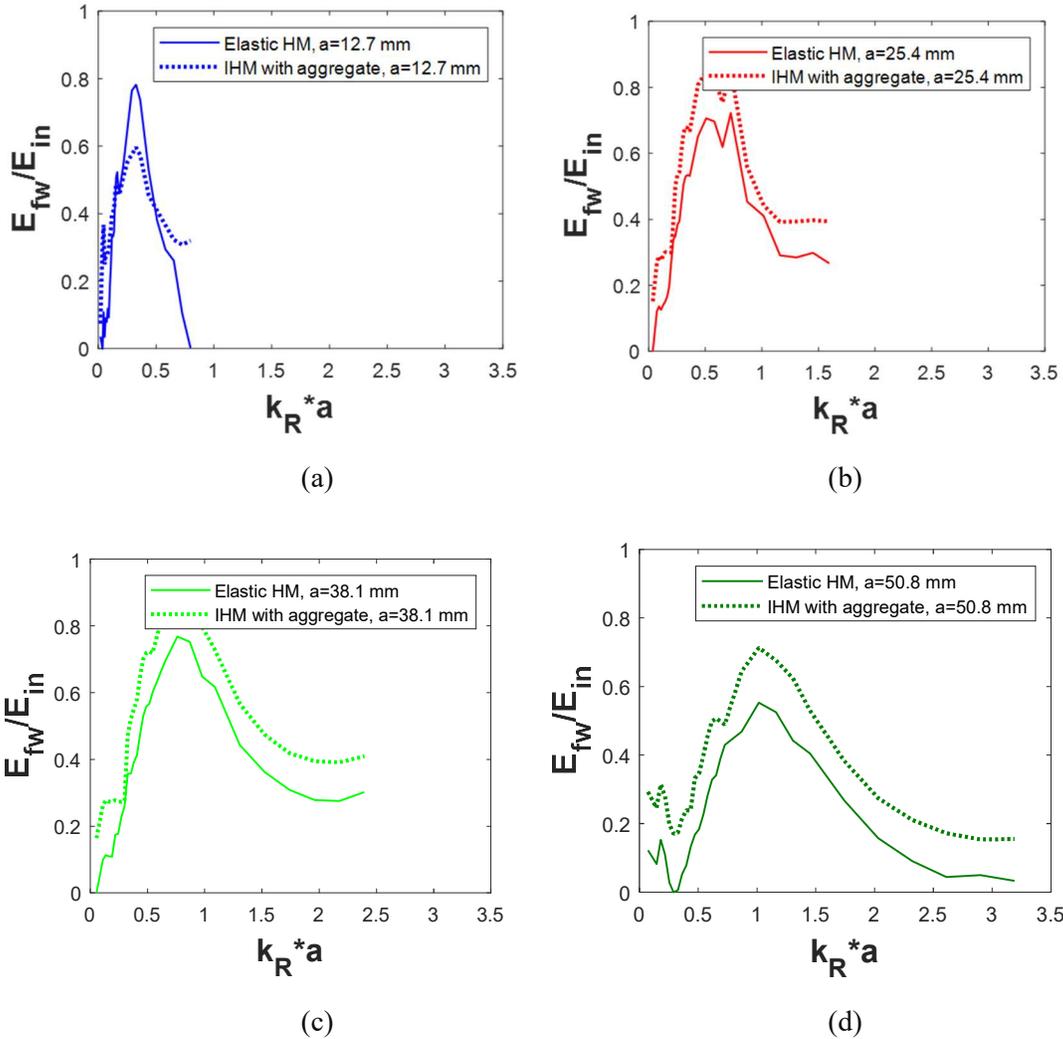


(c)



(d)

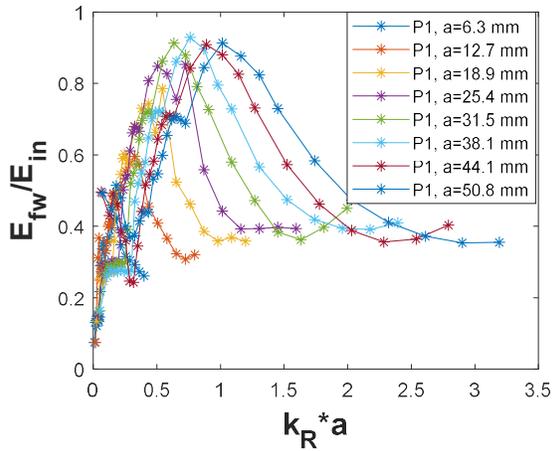
**Figure 5.17.** The WRV pattern with different aggregate size with random distribution (a) the WRV with two aggregate sizes, (b) the comparison between experiment and P1 simulation (c) the boxplot of large aggregate group and (d) the boxplot of small aggregate group. It reveals a slightly lower amplitude for the large aggregate group (L1-L6) compared to the small aggregate group (S1-S6). This difference may stem from larger aggregates dissipating more energy, causing the wave to be influenced by the aggregate size before passing through completely.



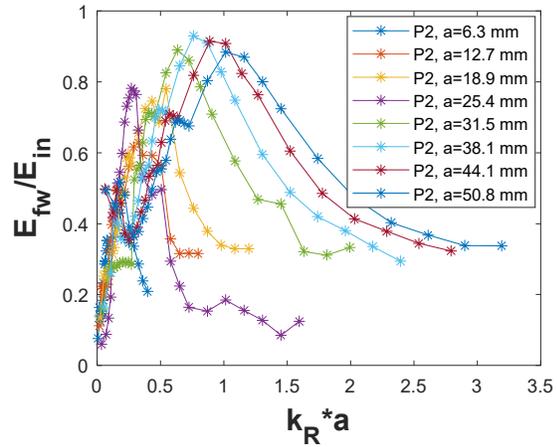
**Figure 5.18.** The WRV comparison between HM and IHM. On the y-axis, the normalized forward wave is depicted, representing the ratio of the first peak by dividing the forward wave  $E_{fw}$  amplitude by the incident wave amplitude  $E_{in}$ . The x-axis represents the product of "a" and wavelength. (a)  $a=12.7\text{mm}$ , (b)  $a=25.4\text{mm}$ , (c)  $a=38.1\text{mm}$ , and (d)  $a=50.8\text{mm}$ .

Figure 5.19 presents ten distinct WRV patterns arising from random aggregate sizes and distributions with the same aggregate proportion. Among these patterns, crack depth,  $a=6.3$  consistently yields lower amplitude, while  $a=50.8\text{mm}$  exhibits higher amplitude, indicating that forward wave energy is readily diffused in shallow cracks. A comparison of patterns like P1 and P2 at  $a=6.3\text{mm}$  demonstrates that aggregate size and distribution impact energy attenuation, even at the same depth-to-crack ratio. The WRV results are intricate due to reflections between particles, making them challenging to interpret. Figure 5.20, different

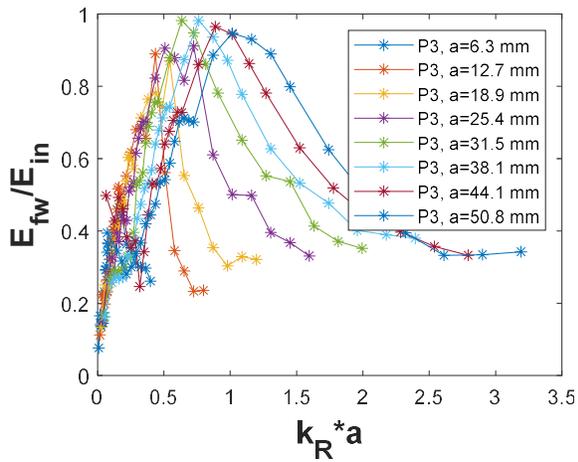
"a" case within the same crack exhibit similar WRV waveforms but varying amplitudes. The WRV curves are influenced by the random distribution and aggregate size. The box plots in Figure 5.21 depict the diverse and intricate WRV values due to different aggregates in IHM. The red plus marks denote values falling outside the standard deviation range, indicating significant differences in forward amplitude from other cases.



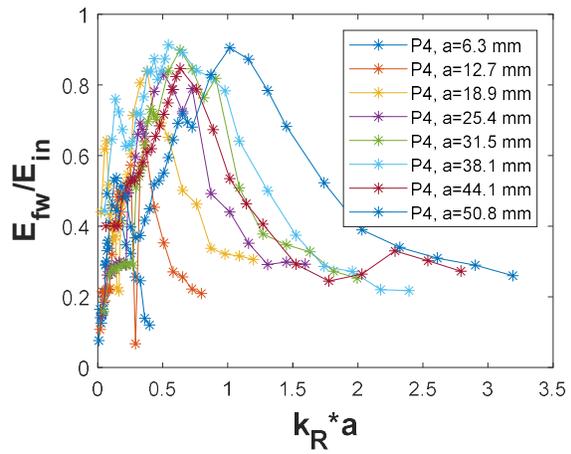
(a)



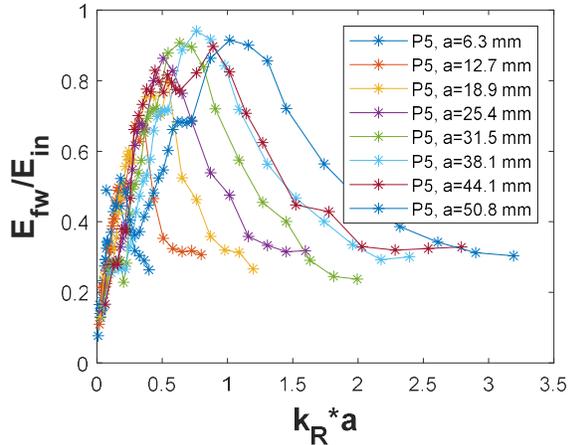
(b)



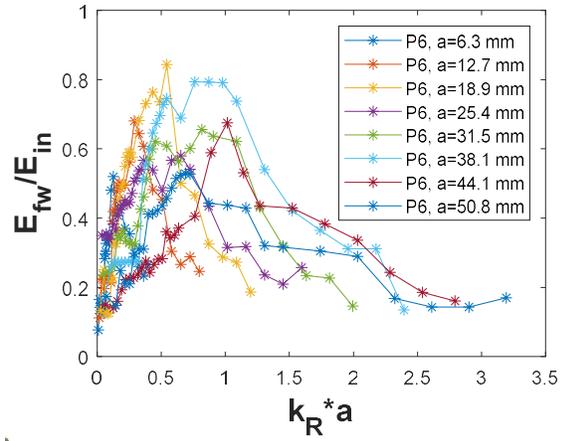
(c)



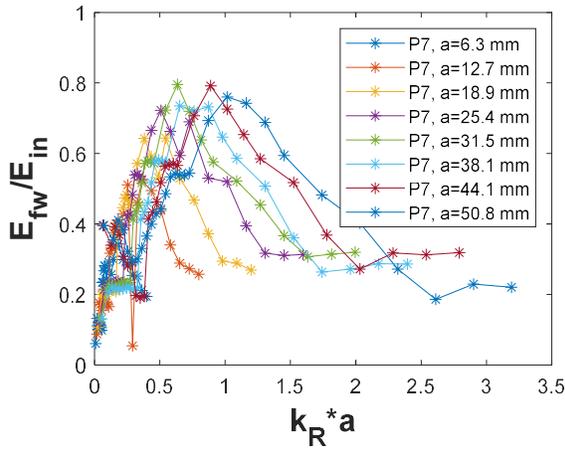
(d)



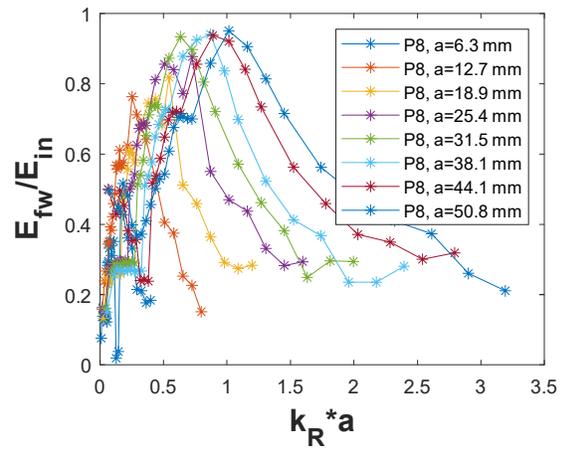
(e)



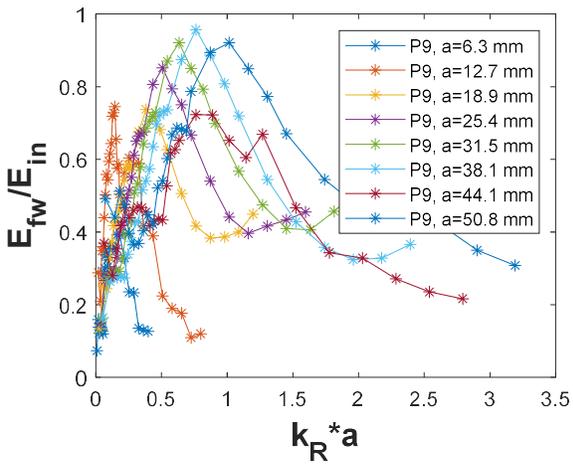
(f)



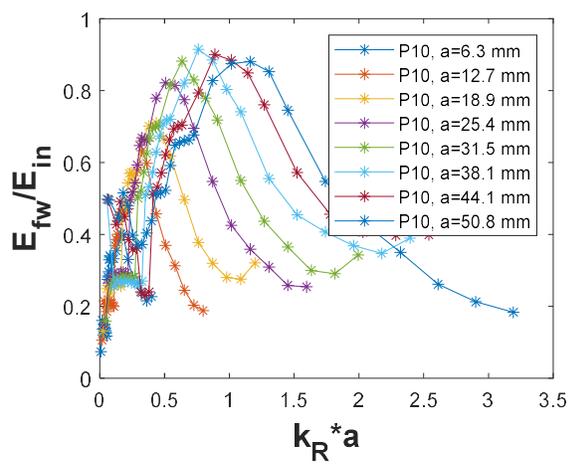
(g)



(h)

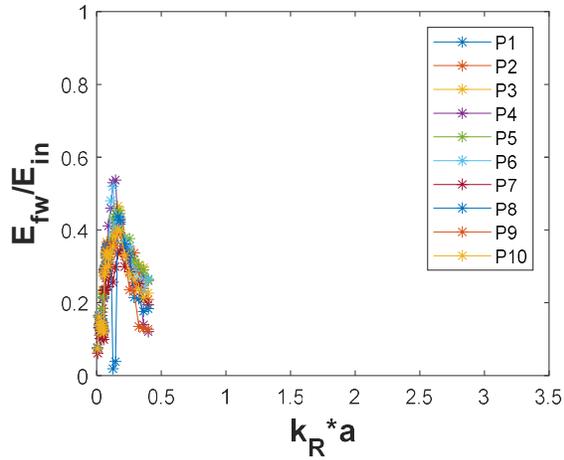


(i)

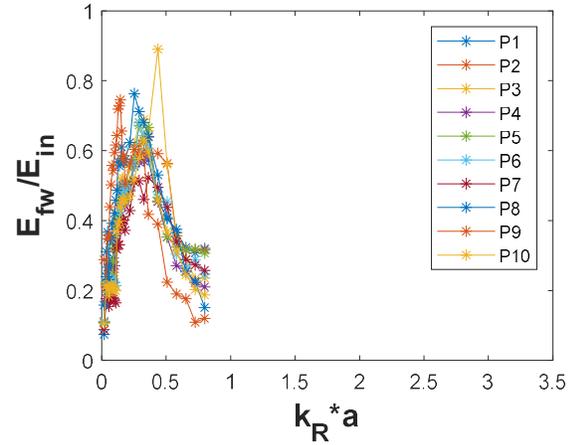


(j)

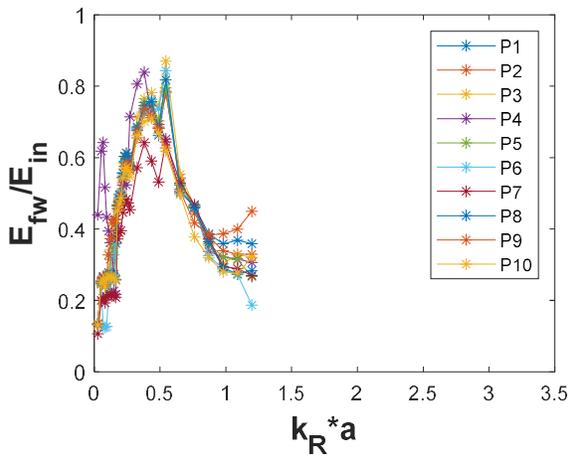
Figure 5.19. Ten pattern WRV results with 8 different “a” value. On the y-axis, the normalized forward wave is depicted, representing the ratio of the first peak by dividing the forward wave  $E_{fw}$  amplitude by the incident wave amplitude  $E_{in}$ . The x-axis represents the product of “a” and wavelength. (a) P1, (b) P2, (c) P3, (d) P4, (e) P5, (f) P6, (g) P7, (h) P8, (i) P9 and (j) P10.



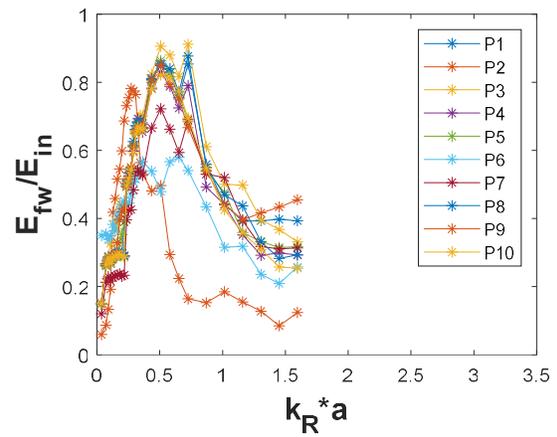
(a)



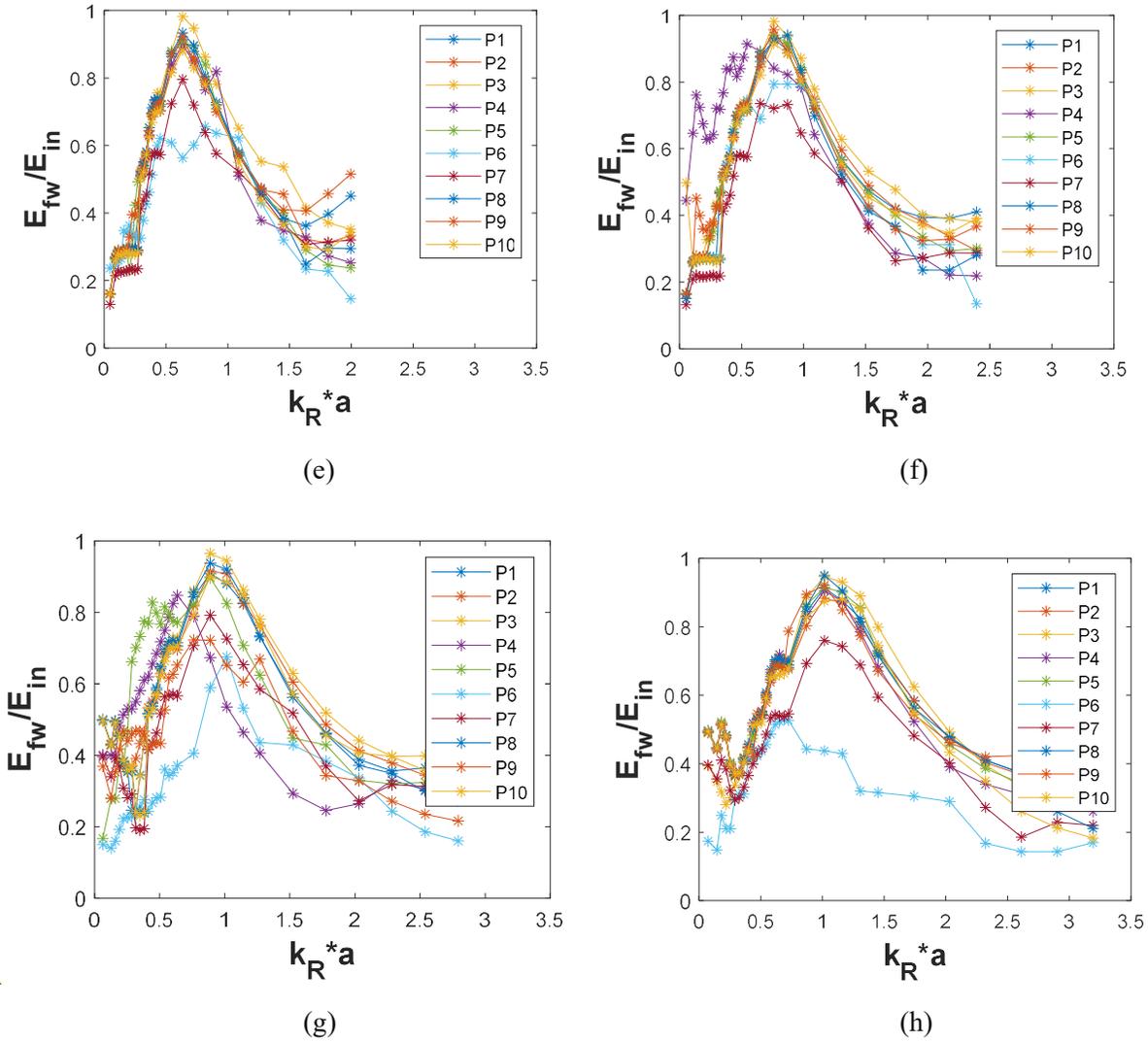
(b)



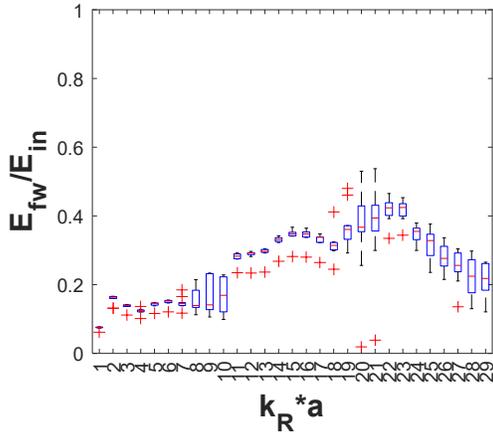
(c)



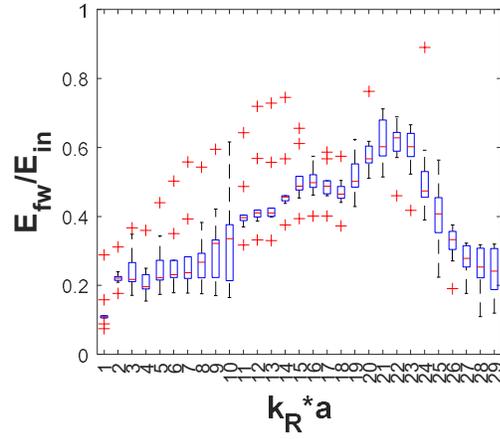
(d)



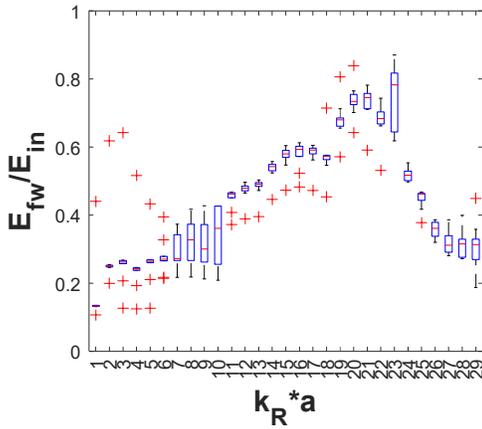
**Figure 5.20.** The ten-pattern comparison by different “ $a$ ” value. On the y-axis, the normalized forward wave is depicted, representing the ratio of the first peak by dividing the forward wave  $E_{fw}$  amplitude by the incident wave amplitude  $E_{in}$ . The x-axis represents the product of “ $a$ ” and wavelength. (a)  $a=6.3$ mm, (b)  $a=12.7$ mm, (c)  $a=18.9$ mm, (d)  $a=25.4$ mm, (e)  $a=31.5$ mm, (f)  $a=38.1$ mm, (g)  $a=44.1$ mm, and (h)  $a=50.8$ mm. Within the same crack, different “ $a$ ” cases exhibit similar WRV waveforms but varying amplitudes. The WRV curves are influenced by the random distribution and aggregate size.



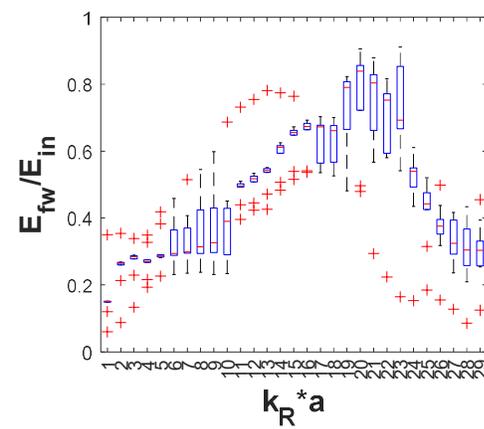
(a)



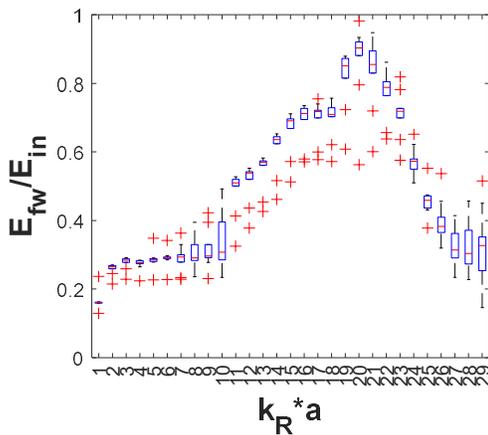
(b)



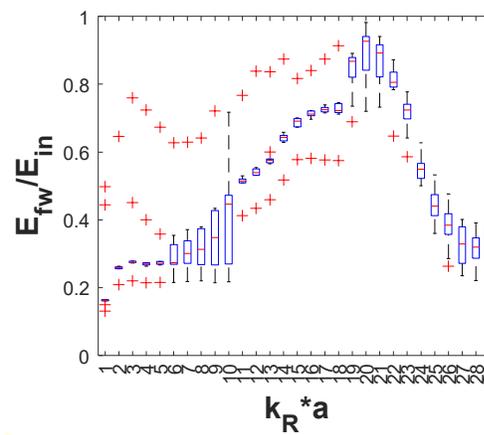
(c)



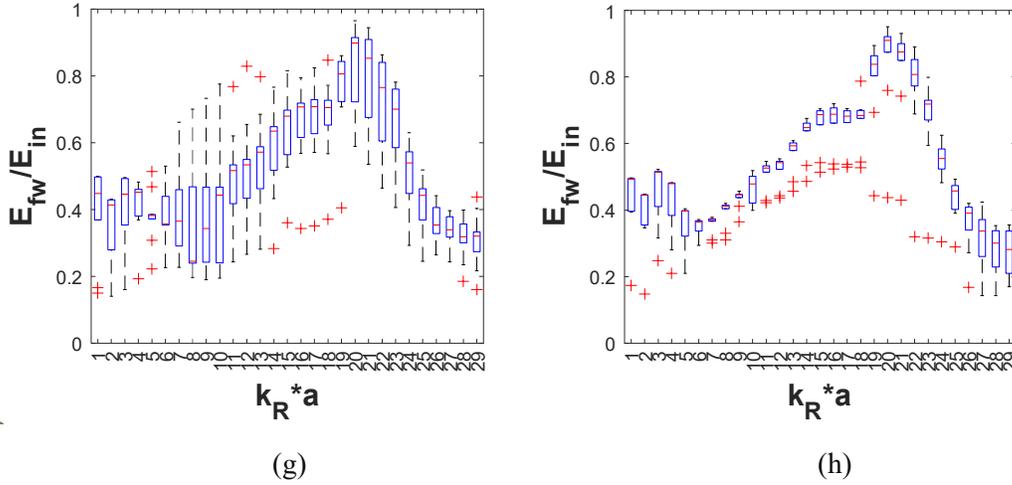
(d)



(e)



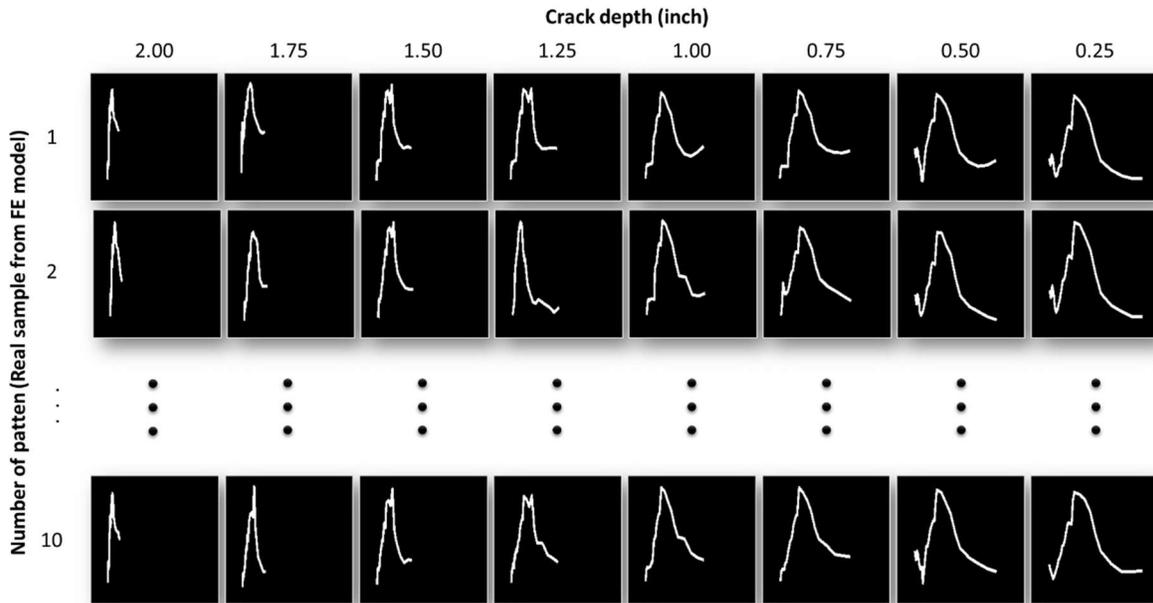
(f)



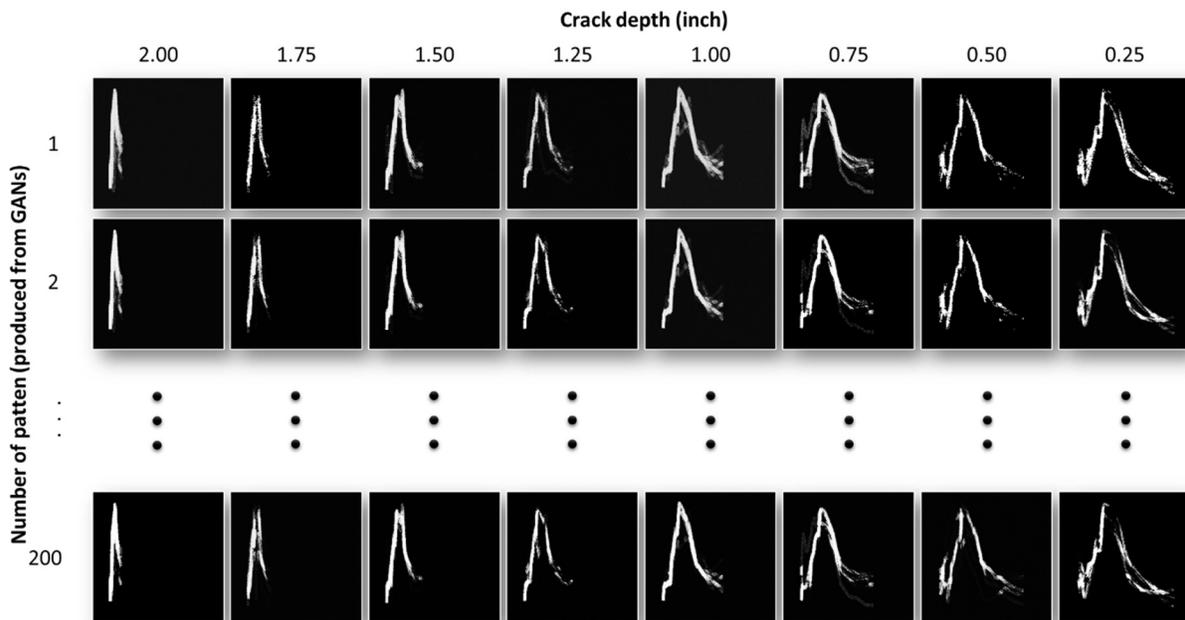
**Figure 5.21.** The boxplot of ten pattern comparison by different “a” value. (a)  $a=6.3\text{mm}$ , (b)  $a=12.7\text{mm}$ , (c)  $a=18.9\text{mm}$ , (d)  $a=25.4\text{mm}$ , (e)  $a=31.5\text{mm}$ , (f)  $a=38.1\text{mm}$ , (g)  $a=44.1\text{mm}$ , and (h)  $a=50.8\text{mm}$ . The red plus marks denote values falling outside the standard deviation range, indicating significant differences in forward amplitude from other cases.

### 5.4.3 Field test damage prediction

The WRV pattern is highly intricate and challenging to interpret, making it difficult to discern differences among various crack depth, “a” value. To overcome this challenge, machine learning techniques such as GANs are employed to classify the WRV patterns more effectively. Since further CNN classification necessitates a larger database of images, GANs are utilized to generate new image samples with features similar to our real training data. The GAN training input consists of WRV waveforms from the FE model, accompanied by eight labels representing the eight different “a” value, as illustrated in Figure 5.22. Figure 5.23 showcases the results obtained from the GANs, producing 200 images for each “a” value group and totaling 1600 images as input for CNN damage prediction. In GANs, a black background provides high contrast against the generated image, enhancing visibility and enabling a detailed assessment of the generated image quality. Consequently, the real sample is inverted to achieve a black background with a white waveform.



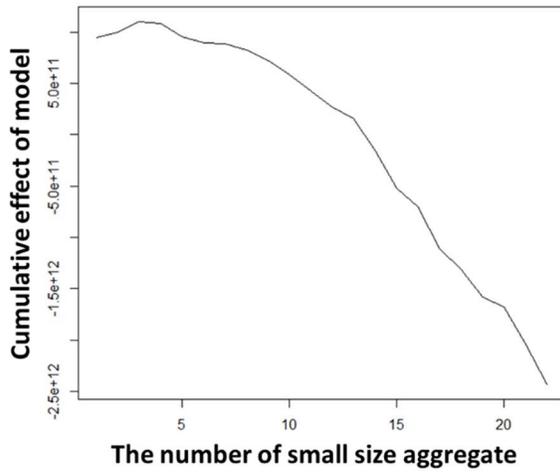
**Figure 5.22.** Example of training input of GANs. There are 8 different groups depends on “a” value with 10 patterns.



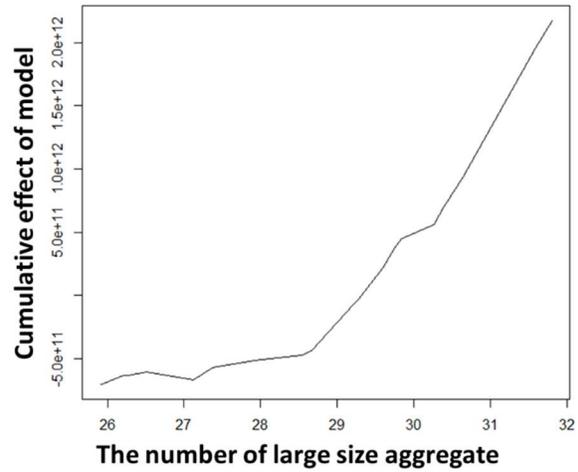
**Figure 5.23.** The output of GANs. In each “a” value, there are 200 images, total has 1600 images.

ALE is a technique employed in machine learning and statistical modeling to analyze the relationship between a feature (or predictor) and a model's output. The ALE plot offers insights into the predicted

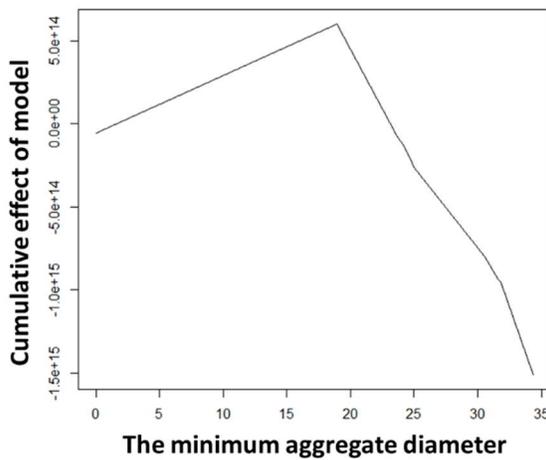
fluctuations, specifically in crack geometry within this paper, when adjusting a single factor while keeping others constant. Figure 5.24 displays the ALE plot, showcasing the results of the ALE analysis conducted with the developed FE model. The investigation encompasses four features: the number of aggregates with small diameter, the number of aggregates with large diameter, the minimum aggregate diameter, and the maximum aggregate diameter. According to the findings, the number of small aggregates exhibits a decreasing effect, while the number of large aggregates shows an increasing effect on the model's output. This observation aligns with the conclusions as depicted in Figure 5.19 (a), where larger aggregates exert a more substantial impact (e.g., reducing the WRV amplitude) on the WRV pattern compared to smaller ones. Figure 5.24 (c) and (d) portray the ALE results for the minimum and maximum aggregate diameters, respectively. The outcomes indicate that a decrease in the minimum aggregate diameter results in a minor effect on WRV, whereas an increase in the maximum aggregate diameter significantly influences the WRV outcomes. While these four ALE results offer valuable insights into the varying WRV patterns associated with changes in aggregate size and number, there remain other unidentified factors that require further exploration in the future, such as aggregate distribution.



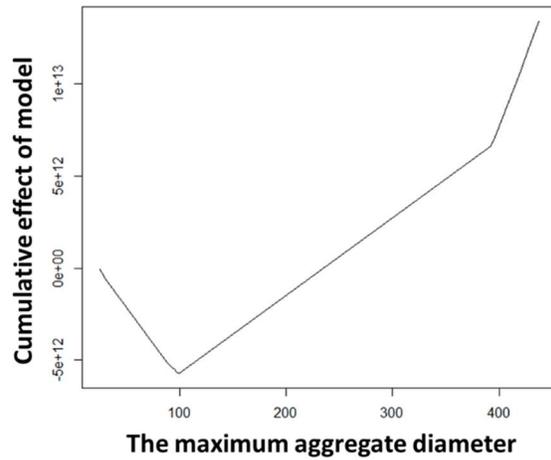
(a)



(b)



(c)



(d)

**Figure 5.24.** *The ALE plot: (a) the number of small size aggregate, (b) the number of large size aggregate, (c) the minimum aggregate diameter, and (d) the maximum aggregate diameter. According to the findings, the number of small aggregates exhibits a decreasing effect, while the number of large aggregates shows an increasing effect on the model's output. This observation aligns with the conclusions as depicted in Figure 5.19(a), where larger aggregates exert a more substantial impact (e.g., reducing the WRV amplitude) on the WRV pattern compared to smaller ones. Figure 5.24(c) and (d) portray the ALE results for the minimum and maximum aggregate diameters, respectively.*

The two CNN testing accuracy confusion matrices are presented in Figure 5.25. Figure 5.25 (a) depicts the CNN model results only without using any fake GAN samples, while Figure 5.25 (b) illustrates the CNN model results with the additional 1080 fake GAN samples. The model without using fake samples achieves a 76.9% accuracy, whereas the model using 1080 fake samples attains an accuracy of 87.7%. The increase in accuracy with the number of fake samples underscores the effectiveness of WRV data in identifying wave propagation patterns. The accuracy rate of 87.5% for the testing model demonstrates its proficiency in identifying diverse wave propagation patterns associated with different aggregate sizes and distributions in IHM. These outcomes strongly suggest that machine learning techniques have the potential to accurately predict damage using WRV data, opening promising avenues for further research and practical applications.

The study of the number of fake samples is presented in Figure 5.26. The CNN accuracy curve increases as the number of fake samples increases, demonstrating that GANs can be used to enhance the dataset and improve the training of the CNN model. The difference in accuracy between 400 fake samples and 1080 samples is only 1%, indicating that the accuracy saturates after using 400 fake samples.

		Actual crack depth (inch)									
		2.00	1.75	1.50	1.25	1.00	0.75	0.50	0.25		
Predicted crack depth (inch)	2.00	4 1%	2 0.5%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	66%	
	1.75	1 0.25%	3 0.75%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	75%	
	1.50	0 0%	0 0%	5 1.25%	1 0.25%	0 0%	0 0%	0 0%	0 0%	83%	
	1.25	0 0%	0 0%	0 0%	4 1%	0 0%	0 0%	0 0%	0 0%	100%	
	1.00	0 0%	0 0%	0 0%	0 0%	5 1.25%	1 0.25%	0 0%	0 0%	83%	
	0.75	0 0%	0 0%	0 0%	0 0%	0 0%	4 1%	0 0%	0 0%	100%	
	0.50	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	1 0.25%	2 0.5%	33%	
	0.25	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	4 1%	3 0.75%	42.80%	
			80%	60%	100%	80%	100%	80%	20%	60%	76.90%

		Actual crack depth (inch)									
		2.00	1.75	1.50	1.25	1.00	0.75	0.50	0.25		
Predicted crack depth (inch)	2.00	10 2.5%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	100%	
	1.75	0 0%	10 2.5%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	100%	
	1.50	0 0%	0 0%	10 2.5%	0 0%	0 0%	0 0%	0 0%	0 0%	100%	
	1.25	0 0%	0 0%	0 0%	10 2.5%	0 0%	0 0%	0 0%	0 0%	100%	
	1.00	0 0%	0 0%	0 0%	0 0%	10 2.5%	4 1%	0 0%	0 0%	71%	
	0.75	0 0%	0 0%	0 0%	0 0%	0 0%	6 1.5%	0 0%	0 0%	71%	
	0.50	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	10 2.5%	10 2.5%	50%	
	0.25	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0%	
			100%	100%	100%	100%	100%	60%	100%	0%	82.5%

(a)

(b)

		Actual crack depth (inch)								
		2.00	1.75	1.50	1.25	1.00	0.75	0.50	0.25	
Predicted crack depth (inch)	2.00	19 4.75%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.75	1 0.25%	20 5%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	95%
	1.50	0 0%	0 0%	20 5%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.25	0 0%	0 0%	0 0%	20 5%	0 0%	0 0%	0 0%	0 0%	100%
	1.00	0 0%	0 0%	0 0%	0 0%	20 5%	7 1.75%	0 0%	0 0%	71%
	0.75	0 0%	0 0%	0 0%	0 0%	0 0%	13 3.25%	2 0.5%	0 0%	85%
	0.50	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	18 4.5%	15 3.75%	54%
	0.25	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	5 1.25%	100%
			95%	100%	100%	100%	100%	65%	90%	25%

(c)

		Actual crack depth (inch)								
		2.00	1.75	1.50	1.25	1.00	0.75	0.50	0.25	
Predicted crack depth (inch)	2.00	30 7.5%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.75	0 0%	30 7.5%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.50	0 0%	0 0%	30 7.5%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.25	0 0%	0 0%	0 0%	30 7.5%	0 0%	0 0%	0 0%	0 0%	100%
	1.00	0 0%	0 0%	0 0%	0 0%	26 6.5%	2 0.5%	0 0%	0 0%	96%
	0.75	0 0%	0 0%	0 0%	0 0%	0 0%	24 6%	8 2%	0 0%	84%
	0.50	0 0%	0 0%	0 0%	0 0%	0 0%	4 1%	22 5.5%	18 4.5%	50%
	0.25	0 0%	0 0%	0 0%	0 0%	4 1%	0 0%	0 0%	12 3%	75%
			100%	100%	100%	100%	86.60%	80%	73.30%	40%

(d)

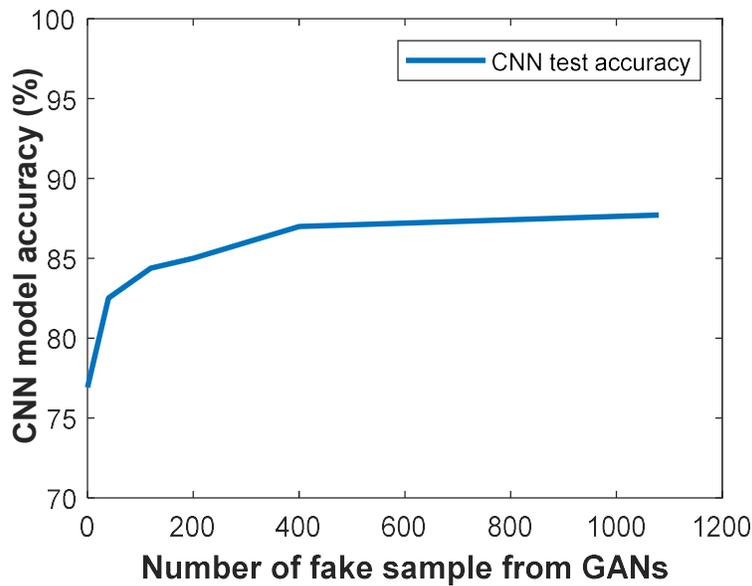
		Actual crack depth (inch)								
		2.00	1.75	1.50	1.25	1.00	0.75	0.50	0.25	
Predicted crack depth (inch)	2.00	55 13.75%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.75	0 0%	52 13%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.50	0 0%	0 0%	51 12.75%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.25	0 0%	0 0%	4 1%	52 13%	6 1.5%	0 0%	0 0%	0 0%	83.80%
	1.00	0 0%	0 0%	0 0%	3 0.75%	49 12.25%	0 0%	0 0%	0 0%	96%
	0.75	0 0%	0 0%	0 0%	0 0%	0 0%	52 13%	7 1.75%	0 0%	84%
	0.50	0 0%	0 0%	0 0%	0 0%	0 0%	3 0.75%	48 12%	30 7.5%	59%
	0.25	0 0%	3 0.75%	0 0%	0 0%	0 0%	0 0%	0 0%	25 6.25%	89%
			100%	94%	94%	94%	86.60%	94%	73.30%	45%

(e)

		Actual crack depth (inch)								
		2.00	1.75	1.50	1.25	1.00	0.75	0.50	0.25	
Predicted crack depth (inch)	2.00	60 15%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.75	0 0%	60 15%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	100%
	1.50	0 0%	0 0%	58 14.5%	3 0.75%	0 0%	0 0%	0 0%	0 0%	95%
	1.25	0 0%	0 0%	2 0.5%	56 14%	0 0%	0 0%	0 0%	0 0%	96%
	1.00	0 0%	0 0%	0 0%	1 0.25%	53 13.25%	4 1%	0 0%	0 0%	91%
	0.75	0 0%	0 0%	0 0%	0 0%	2 0.5%	56 14%	0 0%	0 0%	96%
	0.50	0 0%	0 0%	0 0%	0 0%	5 1.25%	0 0%	48 12%	30 7.5%	57%
	0.25	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	12 3%	30 7.5%	71%
			100%	100%	96%	93%	88%	93%	80%	50%

(f)

Figure 5.25. The confusion matrix of CNN model: (a) the CNN model without using fake samples, (b) with 40 fake sample, (c) with 120 fake samples, (d) with 200 fake samples, (e) with 400 fake samples and (f) with 1080 fake samples. The increase in accuracy with the number of fake samples underscores the effectiveness of WRV data in identifying wave propagation patterns.



**Figure 5.26.** *The CNN accuracy curves vary with different numbers of fake samples. The curves show variations with different numbers of fake samples. The accuracy is 76% without fake samples, while using 400 fake samples can boost accuracy to 86%, reaching a saturated level.*

The impact echo delamination field test results are depicted in Figure 5.27. The image displays two distinct regions of interest, indicating the presence of defects on the tested surface. A comparative analysis between the CNN predictions and crack length evaluations obtained from Rayleigh wave analysis is presented in Figure 5.28. The WRV can accurately identify and predict internal cracks with an error margin of 0.2 inches. Area-A prediction results indicate that the field test estimated a crack depth of 0.8 inches, whereas the CNN prediction using WRV suggests a depth of 0.6 inches. For Area-B, the field test estimated a crack depth of 0.4 inches, while the CNN prediction using WRV indicates a depth of 0.7 inches.

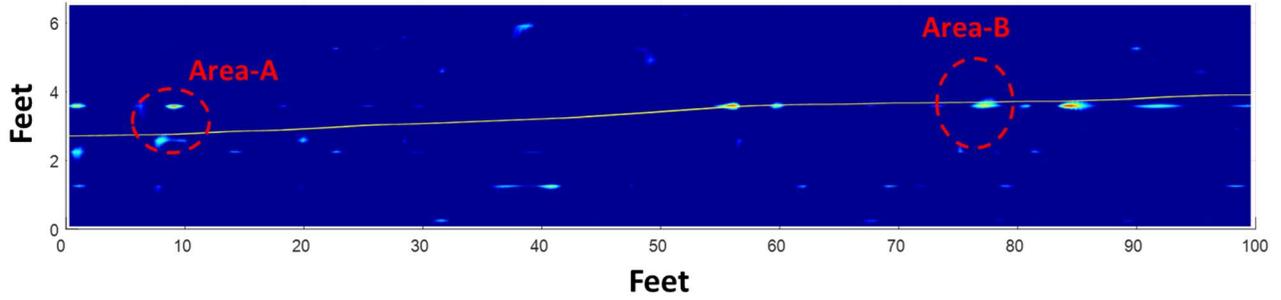


Figure 5.27. The delamination results of impact echo. The yellow solid line denotes the scanning path.

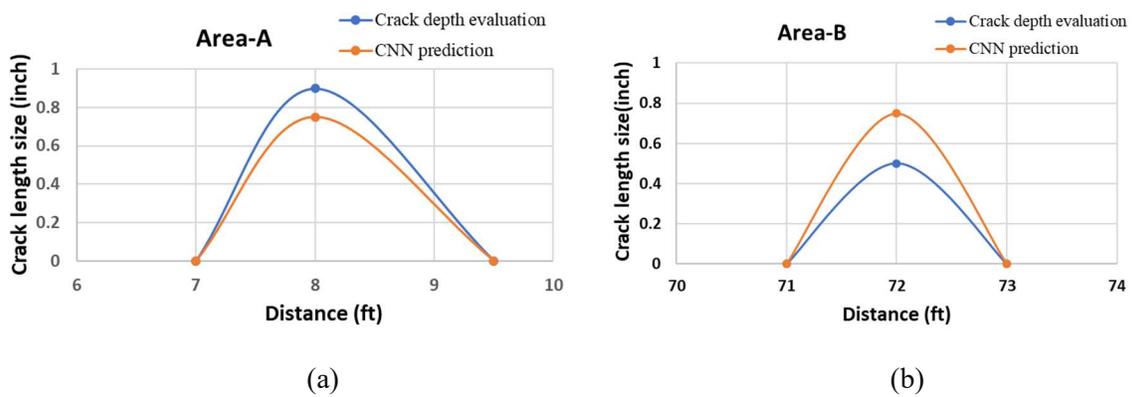
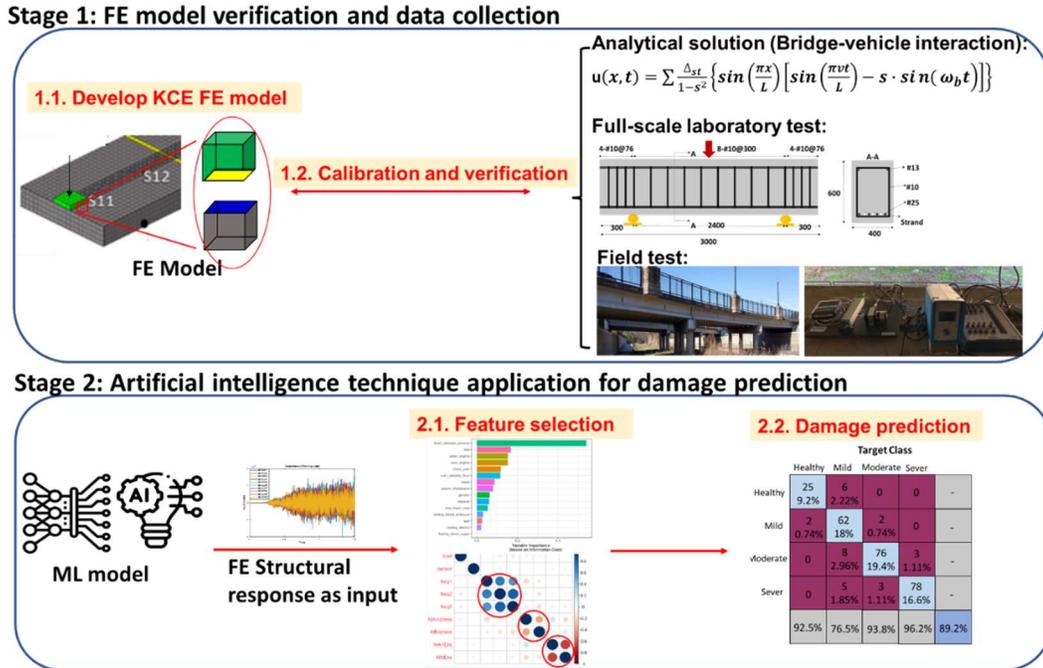


Figure 5.28. The comparison between CNN crack length prediction and the crack estimation from Rayleigh wave(Kang et al., 2021). (a) the area-A prediction result. The field test estimated crack depth is 0.8 inch, while the CNN prediction using WRV is 0.6 inch. (b) the area-B prediction result. The field test estimated crack depth is 0.4 inch, while the CNN prediction using WRV is 0.7 inch.

## CHAPTER 6 STRUCTURAL DAMAGE PREDICTION

### 6.1 Chapter overview

The aim of this chapter is to fill the aforementioned information gaps by developing a reliable FE model that incorporates SHM inspection for use in efficient AI, with a focus on deep learning techniques for predicting structural damage. Three objectives were identified: 1) to present the results of a fundamental study on FE models with a time-dependent moving load, verified through laboratory and field tests, 2) to design a reliable FE model with a moving load and its responses to structural damage, and 3) to apply several ML models to different types of damage, resulting in the development of a highly accurate and reliable damage prediction model. The research is divided into two stages, as shown in Figure 1. Stage 1 provides realistic structural responses using both static and moving loads, with the FE model calibrated and verified through analytical solutions, two laboratory tests, and one field test. In Stage 2, feature selection is applied to extract the most important data collected from Stage 1 as input for ML models, improving the accuracy of damage prediction. Results of the comparison between four different ML models are presented, demonstrating prediction accuracy.



**Figure 6.1.** A flowchart depicting the two stages of developing a more accurate structural prediction model. Stage 1 involves data collection from FE modeling, and Stage 2 shows the implementation of AI application for feature selection and damage prediction.

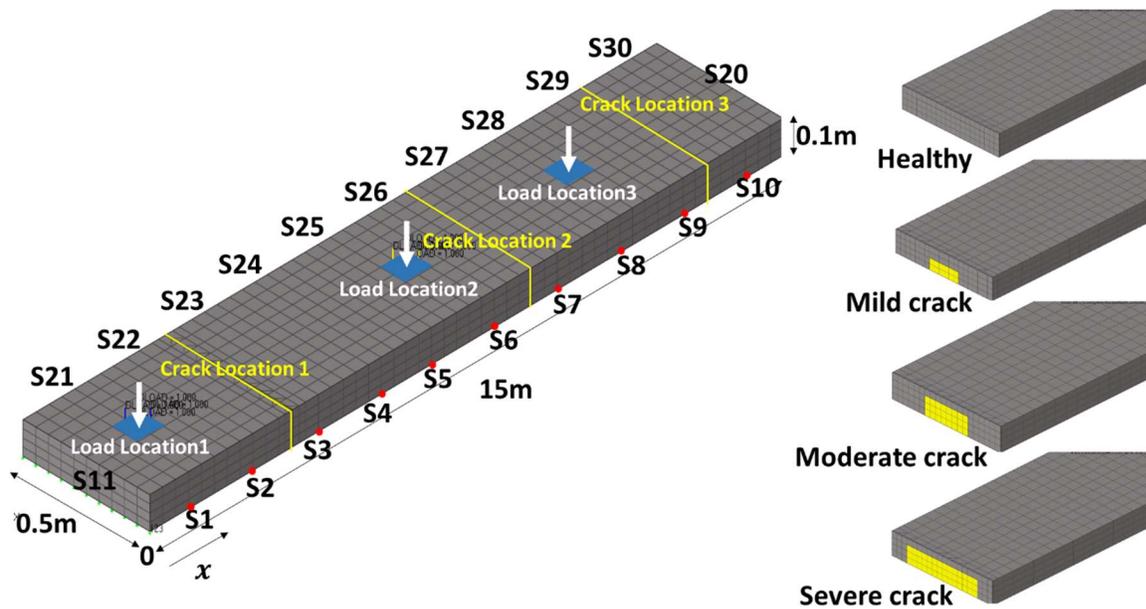
## 6.2 Develop moving load FE model

### 6.2.1 Kinetic contact enforcement method

A 3-D FE model of concrete structures is simulated using ABAQUS/EXPLICIT. The model includes an isotropic, elastic plate with the same Young's modulus ( $E$ ), Poisson's ratio, and thickness as the structure. The element type used is a four-node plane strain element called C3D8R. The mesh size for the solid 3-D simulation is set at 25 mm. The material properties assumed for concrete are typical values ( $2400 \text{ kg/m}^3$ ,  $E = 30 \text{ GPa}$ , and  $\nu = 0.2$ ), while those for tires are  $700 \text{ kg/m}^3$ ,  $E = 1 \text{ MPa}$ , and  $\nu = 0.3$ . The computations are carried out using a computer workstation with eight CPUs clocked at 1.90 GHz, 16 GB RAM, and a 250 GB hard drive.

For the static load FE model, the dimensions used are 0.5 m x 15 m x 0.1 m. In Figure 6.2, 30 sensing points are shown under the bridge deck and numbered from S1 to S30. The red dots represent S1 to S10, which

are applied on the right bottom edge, while S11 and S20 are applied on the bottom of the centerline. S21 to S30 are applied on the left bottom edge but are not marked in the figure to maintain clarity. The figure also displays three different crack locations and three different static load locations. Figure 6.2 shows a sectional view of the three different internal cracks, which are identified by the yellow area. The crack size is determined by taking different fractions of the side section area of each simulated concrete specimen, such as 10%, 25%, and 45%. Each missing element is then defined as a mild, moderate, and severe crack, respectively. The study of the static load-based model involves 12 models, which provide 900 data sets collected from 30 sensing points originating from various models with three crack locations (CL) and three load locations (LL). Table 6.1 displays the CL and LL of the static load model.



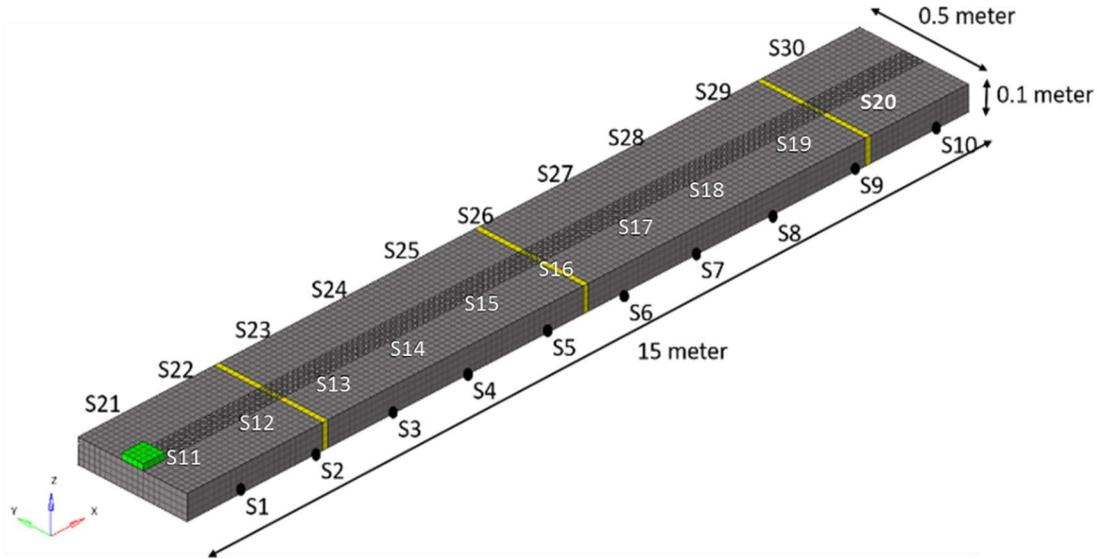
**Figure 6.2.** The model of static load design. The design of the static load FE model with three different load locations (LL 1-3) and three different crack locations (CL1-3). The dimensions of the model are  $1\text{ m} \times 5\text{ m} \times 0.02\text{ m}$ , and it features thirty sensing points (S1-S30) positioned underneath. The model simulates mild, moderate, and severe cracks with 10%, 25%, and 45% of the missing elements, respectively, in the side section area.

**Table 6.1. The locations (x-axis) in Figure 2 of the static load model's crack load (CL) and (LL) load. Unit: meter**

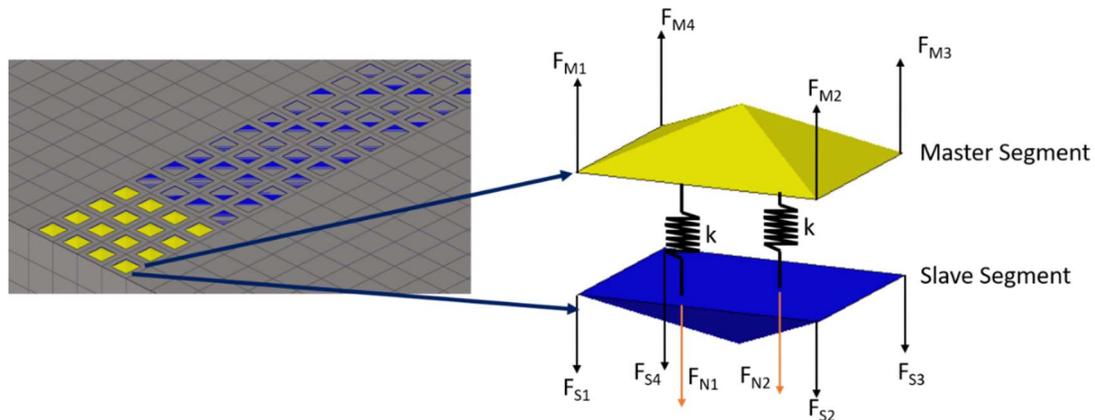
LL1	LL2	LL3	CL1	CL2	CL3
$x = 1$	$x = 6$	$x = 11$	$x = 3$	$x = 8.5$	$x = 13$

The one-way moving-load model is implemented using the KCE method shown in Figure 6.3, where the green area represents the region of contact. The KCE method is utilized for simulating the actual movement of a vehicle. This method employs a master layer and a slave layer, where the force is transmitted through contact with the surface of these two layers. The master layer is the area where the external force is applied and is transferred to the slave layer. The contact conditions between the master and slave layers are typically referred to as contact discretization, and they are an example of surface-to-surface conditions, as illustrated in Figure 6.4. For linear problems, the contact surface must satisfy equilibrium, the expression:

$$F_N = \sum_{i=1}^4 F_{Mi} \tag{6.1}$$



**Figure 6.3. The one-way FE model moving load applied with KCE**



**Figure 6.4. The 3-D model of surface-to-surface contact in KCE, where  $F_S$  and  $F_M$  denote the nodal force on both the slave surface and master surface;  $k$  is the contact stiffness between the master and slave elements, and  $F_N$  represents an equilibrium of forces.**

In order to simulate an actual moving load using KCE, a specific surface interaction boundary condition is required. The augmented Lagrangian method (Hirmand et al., 2015; Simo & Laursen, 1992) and Coulomb's law of friction are used to establish a "linear" FE simulation, as the time-dependent moving load can be transferred to the bridge as a linear impulse. By defining the contact condition, such as contact discretization, and surface interaction property, the transmitting force can be calculated using the augmented Lagrangian method. The force is then used to simulate the moving load by assigning a specific direction

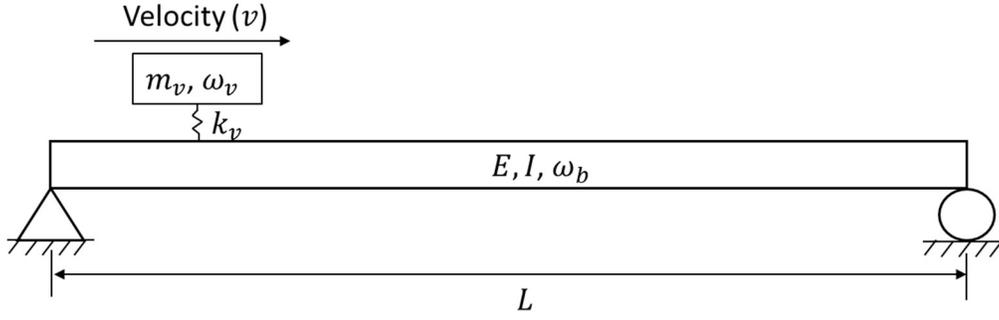
and velocity (Hirmand et al., 2015; Simo & Laursen, 1992). Table 6.2 provides the material properties and details for the KCE-based one-way moving load model, which shares the same properties as the static-load model.

**Table 6.2. The setting for one-way moving load detail and material properties**

Velocity	Concrete Density	Concrete Poisson ratio	Concrete Elastic modulus	Tire Density	Tire Poisson ratio	Tire Elastic modulus	Load Applied
6 mph	2400 kg/m <sup>3</sup>	0.2	30 GPa	700 kg/m <sup>3</sup>	0.3	1 GPa	1 Hz

### 6.2.2 Analytical solutions for moving vehicle

In order to demonstrate the feasibility of the KCE-based moving load FE simulation results, the analytical solution of VBI is utilized for comparison. The VBI concept is based on the interdependence between moving vehicles and bridge responses. The model considers a passing vehicle and a simple bridge, as depicted in Figure 6.5, where  $L$  represents the length of the bridge, while  $E, I$ , and  $\omega_b$  are the bridge moduli of elasticity, the bridge second moment of area, and the bridge's first natural frequency, respectively. Additionally, the mass of the vehicle, vehicle's first natural frequency, and vehicle stiffness are denoted as  $m_v$ ,  $k_v$ , and  $\omega_v = \sqrt{k_v/m_v}$ , respectively.



**Figure 6.5.** The simple model considered passing vehicle and bridge, where  $L$  is the length of the bridge,  $E, I$ , and  $\omega_b$  are the bridge modulus of elasticity, bridge second moment of area and bridge first natural frequency, while  $m_v$ ,  $k_v$ , and  $\omega_v = \sqrt{k_v/m_v}$  are the mass of the vehicle, vehicle first natural frequency, and vehicle stiffness, separately.

The equation of motion for the moving mass over a beam can be written as (Yang et al., 2014):

$$\ddot{q}_v(t) = \frac{\Delta_{st}\omega_v^2}{2(1-s^2)} \left[ A_1 \cos \omega_v t + A_2 \cos \frac{2\pi vt}{L} + A_3 \cos \left( \omega_b - \frac{\pi v}{L} \right) t + A_4 \cos \left( \omega_b + \frac{\pi v}{L} \right) t \right] \quad (6.2)$$

and the first mode structural response of the bridge can be expressed as:

$$u(x, t) = \sum \frac{\Delta_{st}}{1-s^2} \left\{ \sin \left( \frac{\pi x}{L} \right) \left[ \sin \left( \frac{\pi vt}{L} \right) - s \cdot \sin (\omega_b t) \right] \right\} \quad (6.3)$$

and

$$\ddot{u}(x, t) = \sum \frac{\Delta_{st}}{1-s^2} \left\{ \sin \left( \frac{\pi x}{L} \right) \left[ (\omega_b^2 \cdot s) \sin (\omega_b t) - \left( \frac{\pi v}{L} \right)^2 \cdot \sin \left( \frac{\pi vt}{L} \right) \right] \right\} \quad (6.4)$$

where  $\ddot{q}_v(t)$  is the acceleration of the vehicle at the moment  $t$ ;  $u(x, t)$  and  $\ddot{u}(x, t)$  are the displacement and acceleration of the bridge with a specific location( $x$ ) and time( $t$ );  $\Delta_{st}$  is the approximate static deflection at the beam's mid under the gravity action of  $m_v$ ;  $s$  is the bridge coefficient related with bridge properties, and  $A_1, A_2, A_3$ , and  $A_4$  determine the relative contributions of each component to the total acceleration response (Yang et al., 2014). Figure 6.3 displays the unit of properties used as inputs. Figure 6.4 showcases the results of three different case studies conducted with varying material properties and input values such as elastic modulus, density, depth, and width for both FE modeling and analytical solutions. In Case 1,

different values of elastic moduli are examined. In Case 2, different densities are explored, and in Case 3, varying depths are analyzed.

**Table 6.3. The properties of bridges and moving vehicle**

Properties	Unit	Symbol
Length	$m$	$L$
Mass of bridge	$kg$	$m$
Modulus of elasticity	$N/m^2$	$E$
First moment of inertia	$kg - m^2$	$I$
First natural frequency	$Hz$	$\omega_b$
Mass of vehicle	$kg$	$m_v$
Speed of vehicle	$m/s$	$v$
Stiffness of vehicle	$N/m$	$k_v$

**Table 6.4. Detailed variables of parameters between FE modeling and analytical solutions**

	Elastic moduli (GPa)	Density ( $kg/m^3$ )	Depth (m)	Width (m)
Case 1	30, 33, 35, 38, 41	2400	0.1	0.5
Case 2	30	100, 1000, 2400, 3000, 6000	0.1	0.5
Case 3	30	2400	0.1, 0.2, 0.3, 0.4, 0.5	0.5

### 6.2.3 Feature selection and damage prediction model

Data features are important for training the ML model and are extracted from structural responses in FE simulation results. They include statistical indicators such as acceleration, stress, displacement, and spectrum analysis from fast Fourier transform. The authors also considered these structural responses as features. The statistical indicator is shown in *Table 6.5*.

**Table 6.5. Statistical indicator and FE modeling information used for data features** (Finotti et al., 2019)

Peak	Mean	Mean square	Root Mean square	Variance
$\max x $	$\frac{1}{n} \sum_{i=1}^n x_i$	$\frac{1}{n} \sum_{i=1}^n (x_i)^2$	$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i)^2}$	$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Standard deviation	Skewness	Kurtosis	Crest factor	K-factor
$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$	$\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\sigma^3}$	$\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\sigma^4}$	$\frac{x_{peak}}{x_{rms}}$	$x_{peak} \cdot x_{rms}$

To avoid overfitting during training, the authors use feature selection to rank the positive correlations between features and predicting results. The embedded approach is used, incorporating other ML algorithms like random forest to calculate the weight of features. Random forest, composed of several decision trees, calculates impurity by randomly selecting features and assigning weights based on their importance. This approach filters out irrelevant and redundant features and provides better feature selection results. The four most commonly used types of impurity are shown in Table 6.6. The authors also used the Gini impurity.

**Table 6.6. Four commonly used impurities in random forest**

Gini impurity	Entropy	Mean Square Error(MSE)	Mean Absolute Error(MAE)
$\sum_{i=1}^n x_i(1 - x_i)$	$\sum_{i=1}^n -x_i \log_2 x_i$	$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$	$\frac{1}{n} \sum_{i=1}^n  x_i - \bar{x} $

The methods include two types of ML (DT and SVM) and two types of DL algorithms (BP and XGBoost). DT is an optimized regression method using a tree structure (Loh, 2011), SVM is an algorithm for regression and classification tasks (H. F. Chen, 2009)

**Table 6.7. The final setting of each machine learning model**

Machine learning models	BP	SVM	DT	XGBoost
Settings	Neurons:3 Layer:4 Active function: Sigmoid	Kernel: Sigmoid Gamma: scale Degree:3	Criterion: Gini Splitter: best Depth:5	Eta: 0.4 Depth:5 Learning: Gamma regression

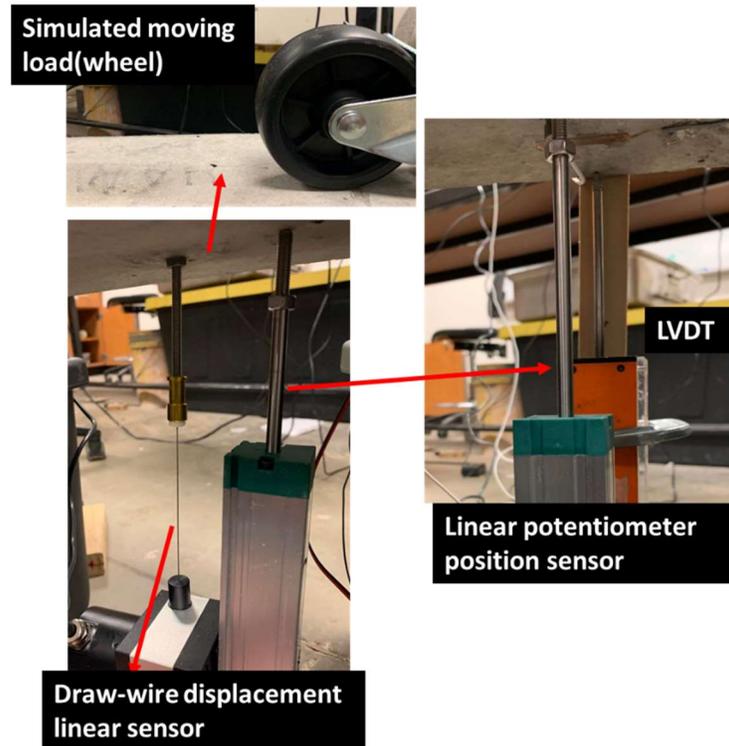
### 6.3 Verification of FEM model

This section presents two laboratory tests and one field test which are used to calibrate and verify the developed KCE model. The purpose of these tests is to gather reference data for comparing the performance of experimental data with the FE model simulation. Once the confirmation between experiment and simulation data is completed, the FE model simulation can provide reliable data to the ML work. The laboratory tests were conducted at the civil engineering laboratory building (CELB) in UTA.

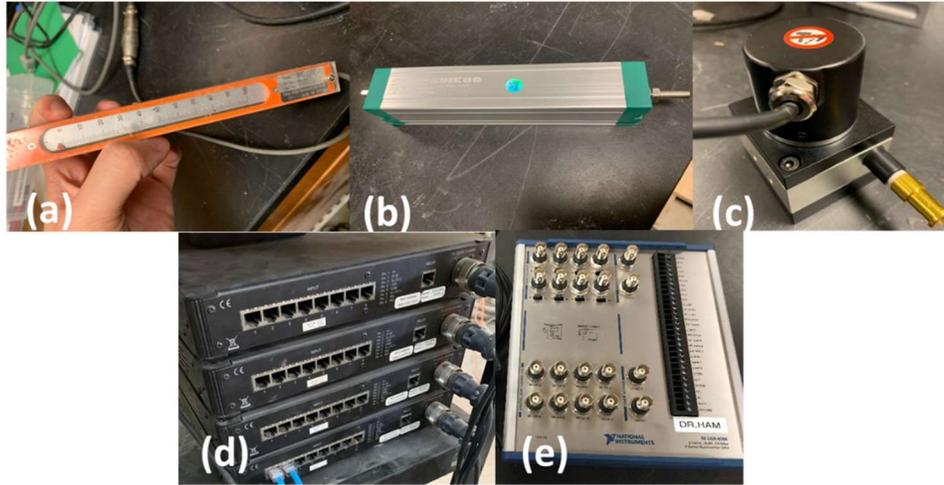
#### 6.3.1 Laboratory sensor calibration test

The primary objective of the first laboratory experiment is to validate the sensitivity, calibration, and overall performance of the sensors. The experimental setup for this test is depicted in Figure 6.6, with a test specimen that measures 20 inches in length, 6 inches in height, and 6 inches in width. In this configuration, the moving element is a wheel, and the sensors are positioned underneath the middle of the concrete beam. Three sensor types are utilized, which include linear variable differential transformer (LVDT), linear potentiometer position sensor, and draw-wire displacement linear sensor. Data from the various sensors are gathered using two distinct data acquisition (DAQ) systems. The Vishay Precision Group (VPG) DAQ is

used for LVDT data collection, while the National Instrument (NI) DAQ is employed for the linear potentiometer position sensor.

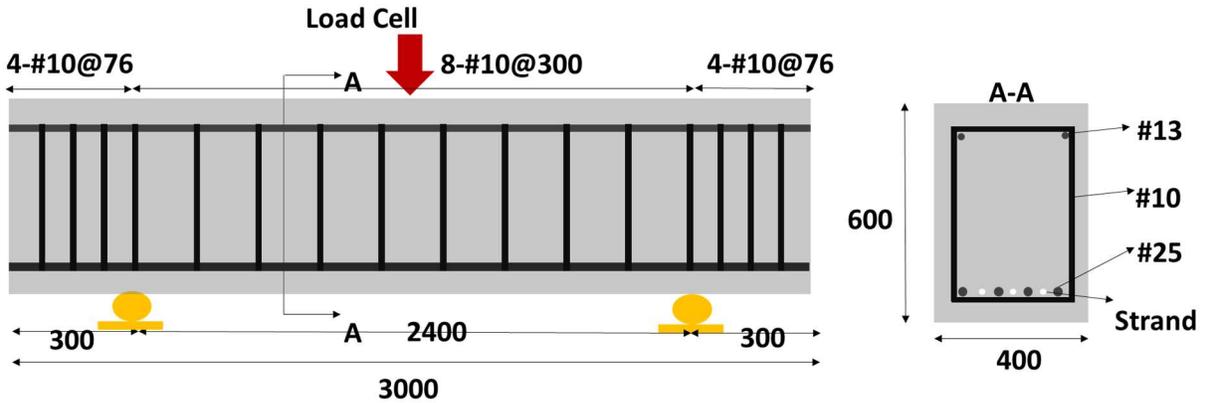


*Figure 6.6. Testing configuration of the sensor performance laboratory test.*



**Figure 6.7. Testing sensors and DAQ. (a) LVDT, (b) linear potentiometer position sensor, (c) draw-wire displacement linear sensor, (d) VPG DAQ, and (e) NI DAQ.**

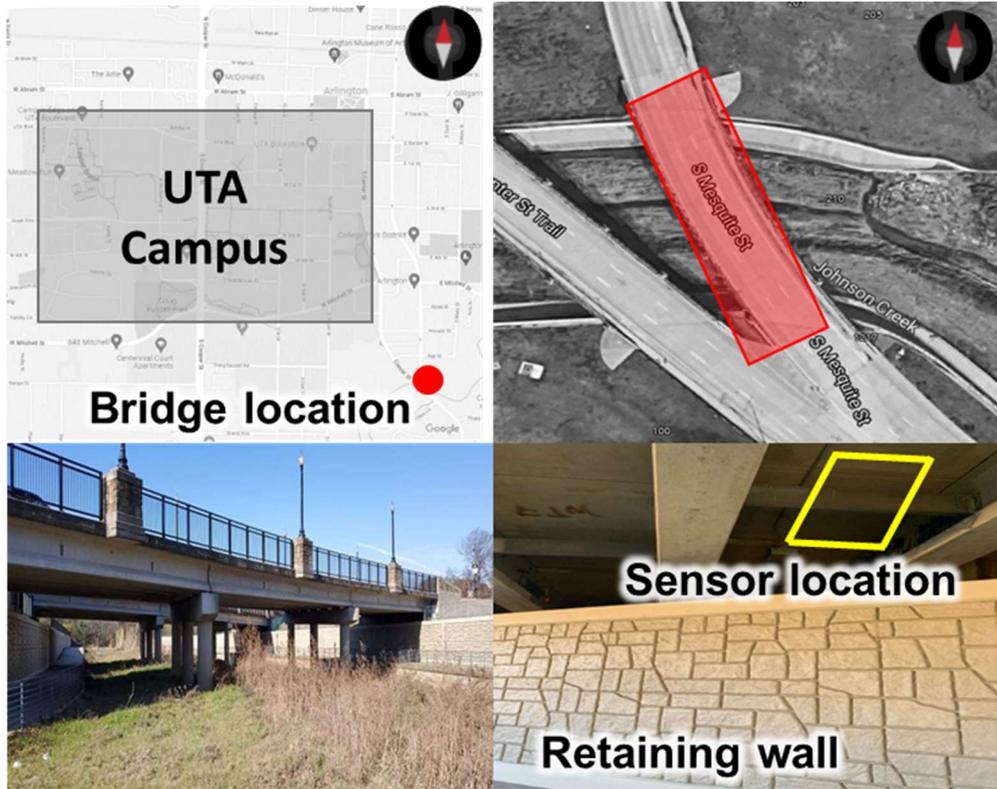
The full-scale laboratory test involves a large beam structure with LVDT sensors placed beneath the middle of the specimen, as depicted in Figure 6.8. The specimen has dimensions of 3m (118 in.) in length, 0.4m (16 in.) in width, and 0.6m (24 in.) in height. Five prestressing strands with a diameter of 0.5 inches (ASTM A416, Grade 270, stress relieved) and a total of 16 #10 stirrups are utilized. Each strand was initially prestressed at 1201MPa (174 ksi), resulting in an average initial prestress of 2.48 MPa (359 psi) in the beams. In addition to the prestressing strands, non-prestressed mild steel reinforcement (Grade 60) is also employed. The shear span to effective depth ratio is set at 3.0, which corresponds to slender beams and typically results in the lowest shear strength for beams without stirrups. A static load is applied to the top middle surface of the beam, and data is primarily collected on the displacement and strain prior to the failure of the concrete beam.



*Figure 6.8. A testing configuration with concrete beam in full-scale laboratory test.*

### 6.3.2 Field test

This field test aims to validate the accuracy of the KCE-based moving load FE simulation results and to evaluate the effectiveness of the sensors and DAQ systems in collecting data. The field test location for this study is S Mesquite St, Arlington. A retaining wall is available for device support, which eliminates the requirement for sensing wire length as shown in Figure 6.9. The bridge has a total of 7 girders and 3 one-way lanes. For this study, two sensors, namely LVDT and linear potentiometer position sensor, are deployed on the middle lane under the bridge deck at mid-span. The VPG DAQ and NI DAQ are used to collect data from the different sensors. Four different types of vehicles are used for the field test, including trucks, vans, SUVs, and sedans, with GVW of 9000 lbs, 5950 lbs, 4455 lbs, and 4431 lbs, respectively. The test is conducted at three different speeds, namely 10 mph, 25 mph, and 40 mph. To ensure accurate sensing data, tests are scheduled during non-peak traffic hours with one vehicle passing at a time.



**Figure 6.9.** *The location of bridge*

## 6.4 Result and discussion

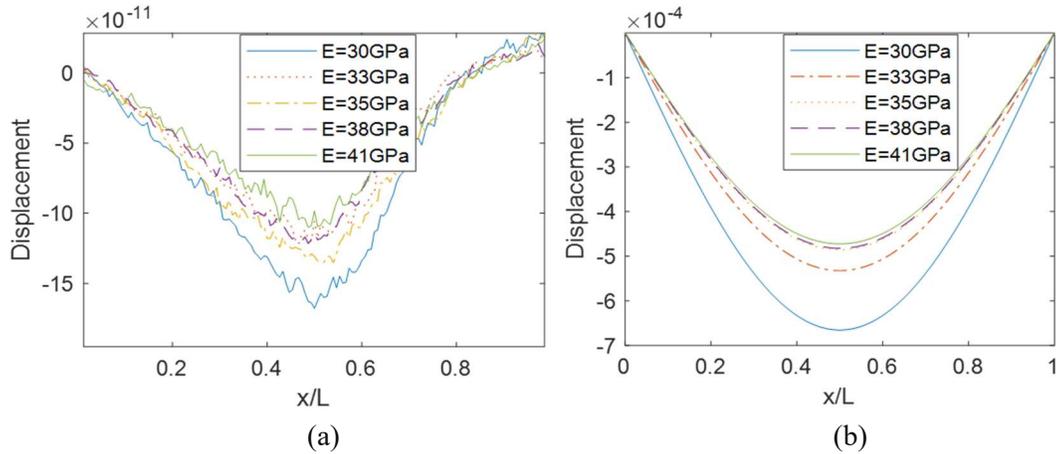
### 6.4.1 *FEM simulation and analytical solution result*

Bridge displacements obtained from KCE-based FE simulation under moving loads with varying parameters: elastic modulus (E), bridge density (D), depth (H), and width (b). The comparison of elastic modulus results between simulation and analytical solution the from Eq. 6.4 is shown in Figure 6.10. The displacement results show linear change with different elastic moduli because the bridge frequency is =

$\frac{\pi^2}{L^2} \sqrt{\frac{EI}{m}}$ . Additionally, the FE simulation and analytical results have different amplitudes due to stiffness

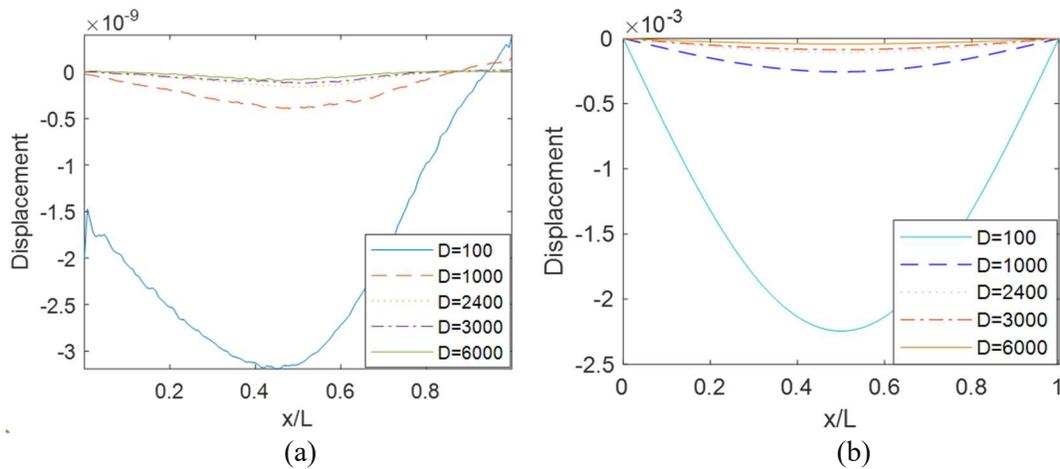
updates between contact layers (master and slave layer) during penalty calculation and contact shell movement. Therefore, the FE model may not produce an exact value as the analytical solution, but it does

confirm the same trend of bridge displacement. The KCE approach uses the Lagrangian method to automatically determine stiffness (Hirmand et al., 2015; Simo & Laursen, 1992).



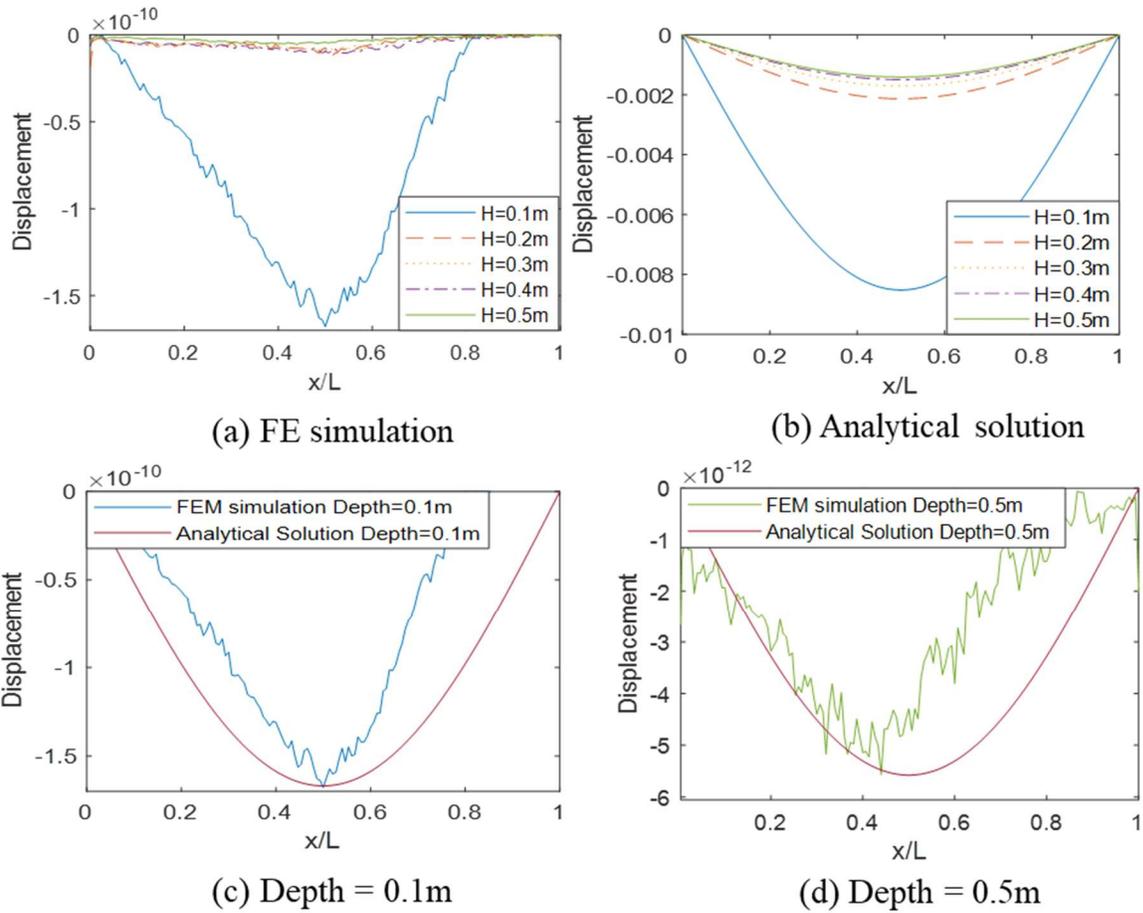
**Figure 6.10. The bridge displacement of (a) the FE simulation and (b) the analytical solution.**

The comparison density ( $D$ ) results between simulation and analytical solution are shown in Figure 6.11. The results show a similar trend when  $D = 100 \text{ kg/m}^3$ , presenting the negative maximum displacement of different density cases. The lower density model has a higher displacement amplitude in both results. In the simulation results, the displacement does not start at 0 when the bridge density is  $100 \text{ kg/m}^3$  due to the higher moving element density  $700 \text{ kg/m}^3$ , causing deformation occurrence in the beginning.

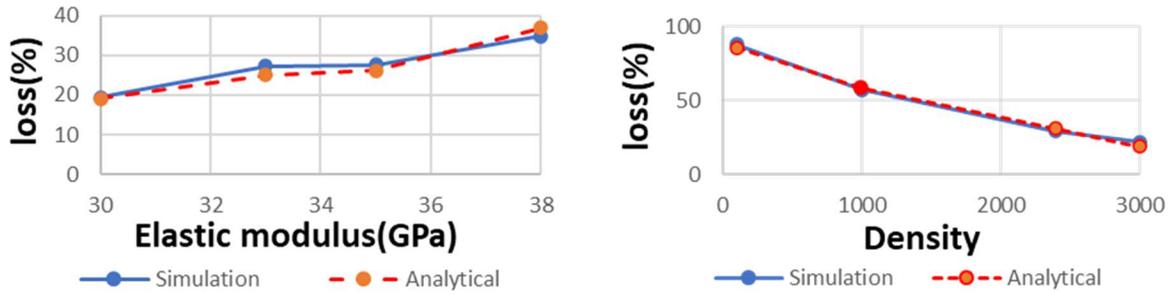


**Figure 6.11. The comparison of density results between FE simulation and analytical solution. The plot shows a similar trend among different density cases, verifying the accuracy of the FE model. The displacement in the simulation results starts at a non-zero value (when  $D = 100 \text{ kg/m}^3$ ) due to the higher density of the moving element ( $700 \text{ kg/m}^3$ ) compared to the bridge density ( $100 \text{ kg/m}^3$ ).**

Figure 6.12 shows displacement responses for various bridge depths (H). Thinner depth bridges show a higher displacement amplitude, but the waveform is affected by high frequency reflected waves from the bottom sides due to vehicle vibration. According to Figure 6.12 (c) and (d), the  $H=0.1 \text{ m}$  model ( $10^{-10}$ ) shows more reflection than the  $0.5 \text{ m}$  model ( $10^{-12}$ ). The larger specimen has less reflection in FEM. The  $H=0.5 \text{ m}$  depth model shows more high-frequency noise with its maximum displacement shifting to the left due to different depths and number of layers; the  $H=0.1 \text{ m}$  model has only one layer while the  $H=0.5 \text{ m}$  model has five layers. More layers with thicker dimensions may cause more reflected waves. Figure 6.13 shows the loss (%) of maximum bridge displacement, which exhibits similar values among all cases that demonstrate FE simulation from KCE. The elastic modulus study shown in Figure 6.13 (a) and density study shown in Figure 6.13 (b). Each study shows similar patterns between analytical solution and KCE model results.



**Figure 6.12.** The comparison of depth results between simulation and analytical solutions. The results show the trend where the lower depth is shown as (e) a higher displacement. However, the waveform is substantially affected by high frequency reflected waves from (c) the bottom sides caused by (d) a thicker dimension.



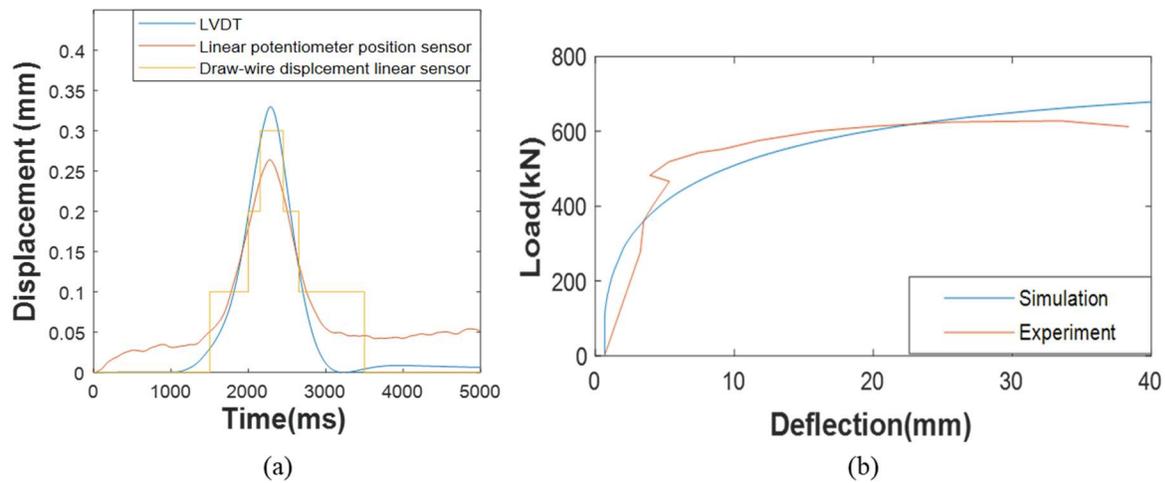
(a)

(b)

**Figure 6.13. The loss (%) of maximum displacement. The value between FE simulation and analytical solution demonstrates similar energy loss, which means the developed FE simulation from KCE and analytical solution has a similar pattern between moving vehicles. (a) elastic modulus study, and (b) density study.**

#### 6.4.2 Laboratory test results

The purpose of the first laboratory test is to calibrate sensor sensitivity. Lab test in Figure 6.14 (a) shows LVDT provides high-resolution continuous data of max amplitude, while linear position sensor slightly differs from LVDT. This is because the movable steel bar in the linear position sensor can only be pushed or pulled by the bridge deck displacement, whereas the LVDT has an additional spring that can push the pointer back. Draw-wire displacement linear sensor has 0.1 mm resolution and only detects variations  $\geq 0.1$ mm. It was not used in the field test due to potential lower displacement. In the sensor calibration laboratory test, LVDT provided the best resolution among all displacement sensors. Thus, LVDT data was chosen as a reference for verifying FE simulation results. Figure 6.14 (b) compares deflection results from the full-scale laboratory test with FE simulation. The x-axis represents mid-span sensor deflection, and the y-axis represents load. The concrete beam deflection is recorded before it fractures. Both the simulation and experiment show the same pattern as the load increases. At 400kN load, the deflection increases at a faster rate in both cases. It should be noted that the FE simulation uses static load simulation, while the field test applies a moving load.



**Figure 6.14. The Laboratory test results: (a) the displacement result of the sensor calibration laboratory test. The draw-wire displacement linear sensor provides low-resolution results and, therefore, will not be used in the field test. (b) The comparison between full-scale laboratory tests and KCE simulation.**

### 6.4.3 Field test results

Figure 6.15 shows results of the field test with two displacement sensors. Figure 6.15 (a) is for a truck with a speed of 10mph, while Figure 6.15 (b) is for a truck with a speed of 40mph. The LVDT (blue dash line) provided higher displacement compared to the linear potentiometer (red line) in both cases. The LVDT had a clearer waveform and was less affected by vibration after the vehicle passed due to the presence of a spring. The displacement duration for the 10mph case was longer than for the 40mph case due to the slower vehicle speed. The displacement amplitudes were similar in both cases, with the maximum displacement of the LVDT being 0.1mm and the linear potentiometer position sensor being 0.07mm. The comparison between KCE FE model and field test as shown in Figure 6.16. The van and sedan cases at 25 mph shown in Figure 6.16 (a) and (b), separately. A van moving at 25 mph caused a maximum displacement of 0.1 mm, while a sedan caused a maximum displacement of 0.04 mm. These results demonstrate the reliability of the developed FE simulation model with moving load in simulating reality VBI. Overall, the weight of the vehicle affects displacement and acceleration, and the moving speed affects acceleration response, which are important factors to consider in the data analysis of the field test.

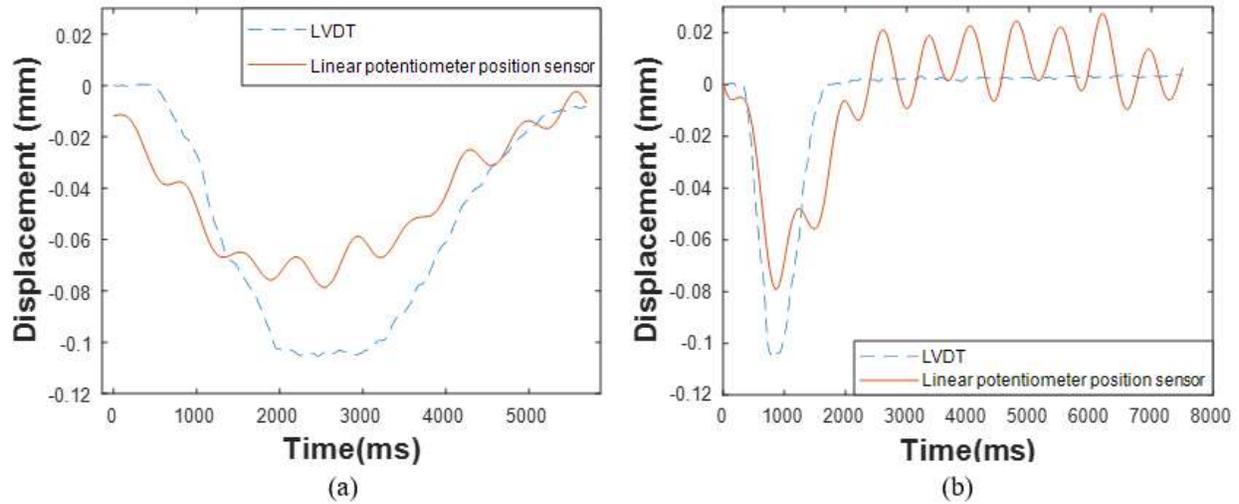


Figure 6.15. The example of field test results: (a) with truck 10mph, and (b) with truck 40mph.

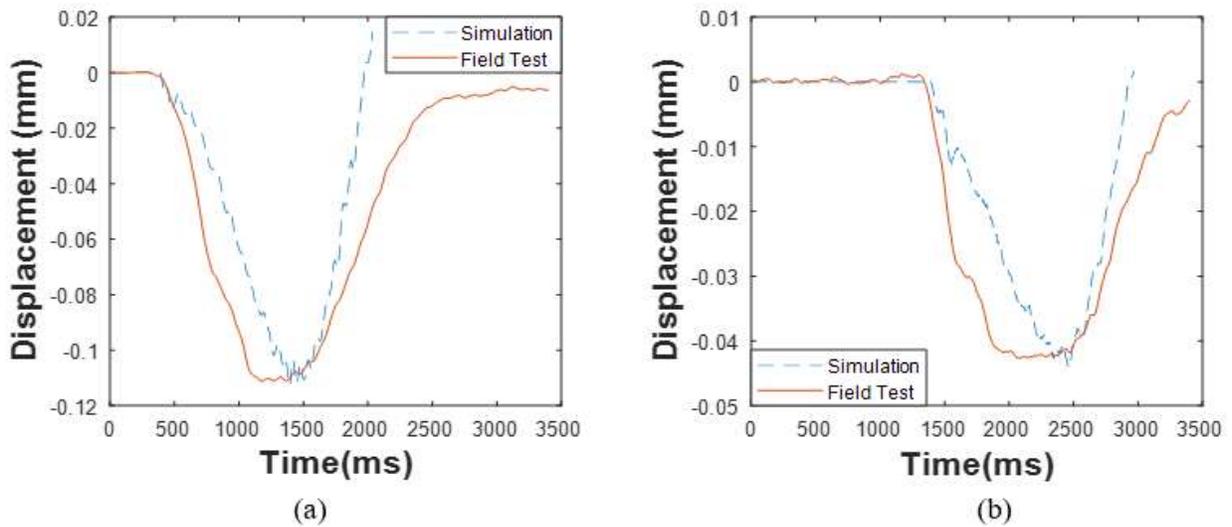
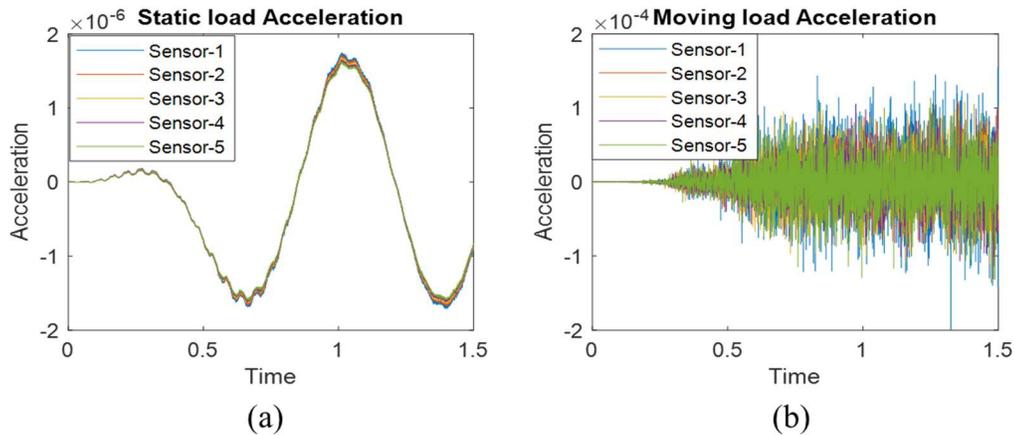


Figure 6.16. The comparison between KCE FE model and field test: (a) with van 25 mph and (b) with sedan 25 mph

#### 6.4.4 Feature selection results

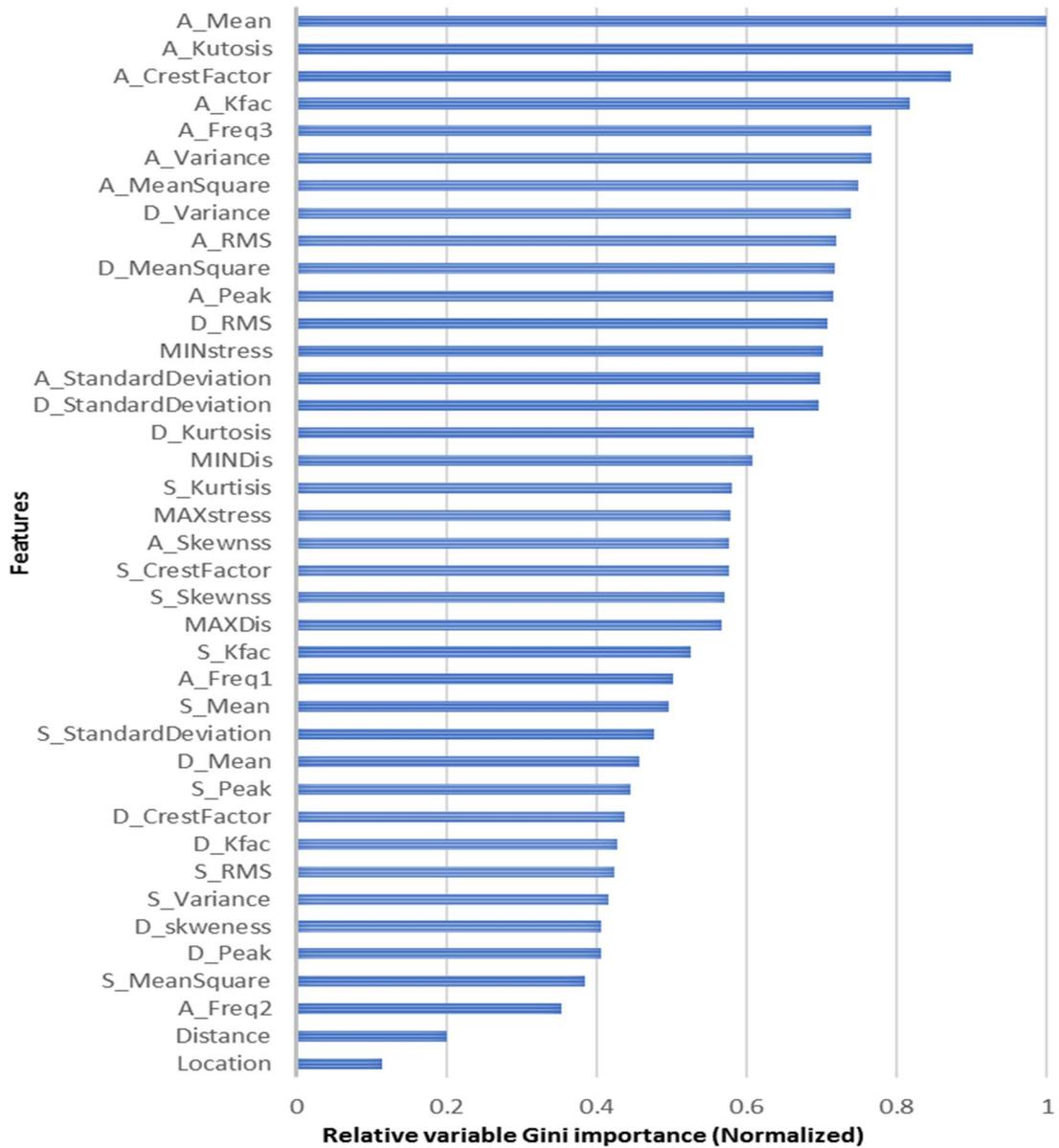
Once the FE model was calibrated and verified with analytical solutions, laboratory tests, and field tests, a set of FE models was created for the damage prediction database. The structural response to static and

moving loads with CL1 is shown in Figure 6.17. Moving load acceleration shows more complex patterns than static load acceleration due to the vehicle's continuous impulse generation through vibrations. Despite having the same crack location and size, the structural response varies significantly based on the load applied. ML is used to identify patterns in different crack cases across crack and load locations.



**Figure 6.17. Examples of structural acceleration responses: (a) under static loads and (b) under moving loads, presenting more complex wave response patterns.**

Due to the black-box nature of most ML techniques, feature significance and model interpretation can be complex. Only a small number of features may have a significant impact on the response, while the majority of features may be irrelevant. Gini importance can be used to determine the contribution of each feature in predicting the response, providing a better understanding of feature relevance and their correlation to the ML prediction model. Features with higher importance have a greater effect on the prediction model. Figure 6.18 shows the importance rank calculated from acceleration (A), displacement (D), and stress (S), with mean acceleration having the highest importance for damage prediction. Using the Gini importance, features with a score over 0.3 were selected for training an ML model, with the permutation method providing correction for bias and reducing prediction uncertainty.



**Figure 6.18.** The importance ranking is calculated from the random forest with Gini impurity. The first capital alphabet presents what data used for indicators. “A” means the indicator calculated from acceleration; “D” means the indicator calculated from displacement and “S” means the indicator calculated from stress. The acceleration provides the highest importance feature for ML prediction results. The features with Gini importance over 0.3 are used in ML training data for damage prediction.

### 6.4.5 Damage prediction results

In this section, the damage prediction of four classes of static loads and moving loads in FE models are explored, using four different ML models. While static loads are more straightforward to comprehend, moving loads are closer to real-world scenarios. To obtain damage prediction results, the static load simulation outcomes as depicted in Figure 6.19. The four classes of damage severity are healthy (intact), mild cracking, moderate cracking, and severe cracking, based on the FE model design. The predicted results are presented as matching matrices, where the actual class is the known information from FE simulation results, and the predicted class is the output from the ML model. The accuracy of damage prediction is shown by the diagonal line. Accuracy for prediction means that the predicted class matches the actual class. For example, according to Figure 6.19 (a), the accuracy of healthy, mild cracking, moderate cracking, and severe cracking between the real class and predicted class is 27.7% for a predicted healthy class, 14.4% for a predicted mild cracking class, 22.6% for a predicted moderate cracking class, and 25.6% for a predicted severe cracking class. The 27.7% value is calculated based on the fraction of all testing samples, including the healthy real class sample ( $78/270 * (75/78) * 100\%$ ), where 78 is the total of healthy samples in a real class; 270 is the sum total of all tested samples and  $(75/78) * 100\%$  is the accuracy percentage of ML prediction in the healthy class compared to the other class totals. The total accuracy is the sum of these four values, demonstrating an accuracy of 90.3%. For the static load model, XGBoost showed the highest accuracy (92.6%) in predicting cracks. BP achieved an accuracy of 90.3%. SVM and DT had lower prediction outcomes.

		Healthy	Mild crack	Moderate crack	Severe crack	
Predicted Class	Healthy	75 27.7%	4 1.48%	0 0%	0 0%	94.9%
	Mild crack	1 0.37%	39 14.4%	4 1.48%	0 0%	88.6%
	Moderate crack	0 0%	5 1.85%	61 22.6%	4 1.48%	87.1%
	Severe crack	2 0.74%	6 2.22%	0 0%	69 25.6%	89.6%
		96.2%	72.2%	93.8%	94.5%	90.3%

(a) BP

		Healthy	Mild crack	Moderate crack	Severe crack	
Predicted Class	Healthy	64 23.7%	6 2.22%	1 0.37%	0 0%	88.95%
	Mild crack	4 1.48%	55 20.4%	4 1.48%	0 0%	87.3%
	Moderate crack	0 0%	5 1.85%	45 16.7%	10 3.7%	75%
	Severe crack	3 1.11%	0 0%	9 3.33%	64 23.6%	84.2%
		90.1%	83.3%	76.2%	86.4%	84.4%

(b) SVM

		Healthy	Mild crack	Moderate crack	Severe crack	
Predicted Class	Healthy	74 27.4%	6 2.22%	1 0.37%	9 3.33%	88.95%
	Mild crack	6 2.22%	47 17.4%	4 1.48%	0 0%	87.3%
	Moderate crack	0 0%	8 2.96%	39 14.4%	2 0.74%	75%
	Severe crack	7 2.59%	4 1.48%	3 1.11%	60 22.2%	84.2%
		90.1%	83.3%	76.2%	86.4%	81.4%

(c) DT

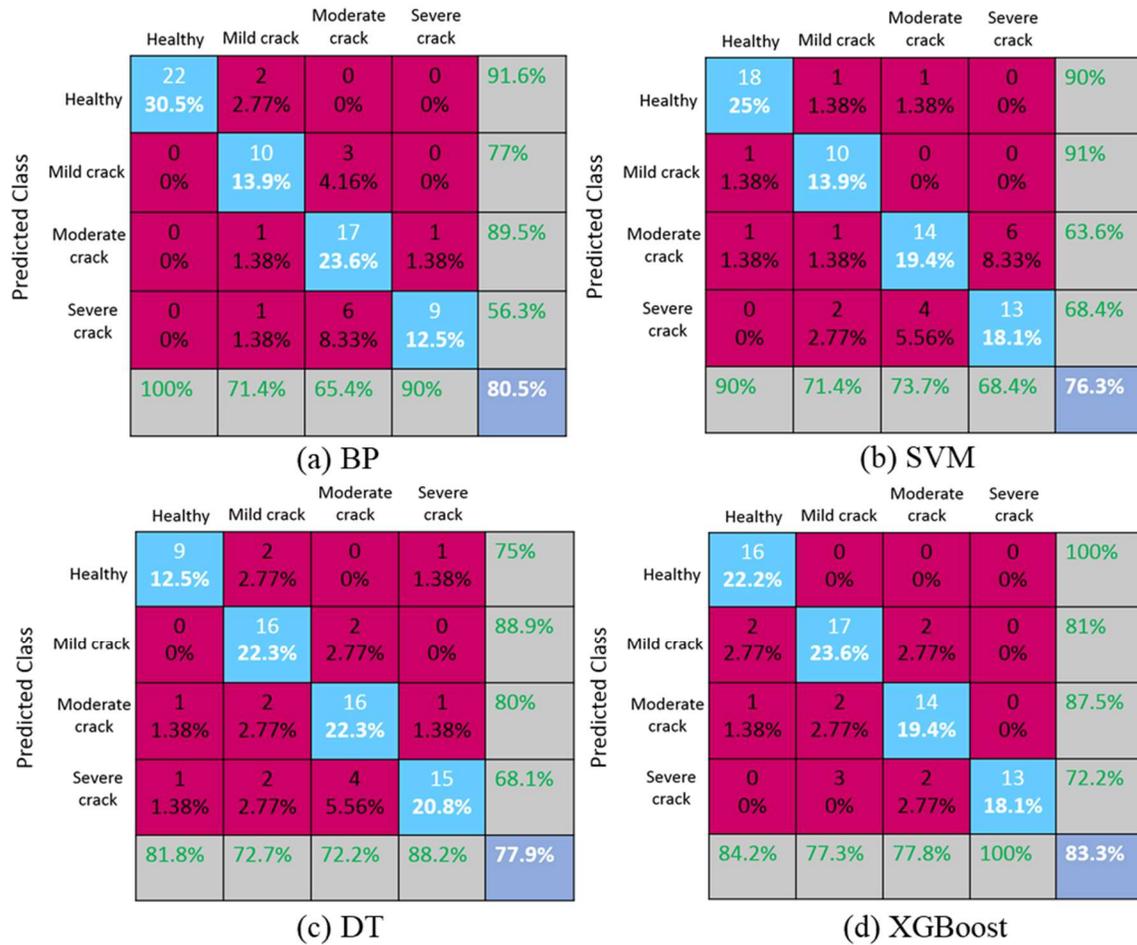
		Healthy	Mild crack	Moderate crack	Severe crack	
Predicted Class	Healthy	62 23%	0 0%	0 0%	0 0%	100%
	Mild crack	4 1.48%	51 18.9%	6 2.22%	4 1.48%	78.5%
	Moderate crack	0 0%	5 1.85%	72 26.7%	4 1.48%	88.9%
	Severe crack	0 0%	3 1.11%	0 0%	65 24%	95.6%
		93.9%	86.4%	92.3%	89%	92.6%

(d) XGBoost

**Figure 6.19. The damage prediction results were derived by using static load FE simulation results from (a) BP, (b) SVM, (c) decision tree and (d) XGBoost algorithms.**

The prediction results under each moving load are shown in Figure 6.20. XGBoost has the highest prediction accuracy (83.3%), followed by BP, DT, and SVM with an accuracy of 80.5%, 77.9%, and 76.3%, respectively. In comparison to the static load prediction model, the moving load model exhibits lower accuracy in damage prediction due to the more intricate input signals from structural response. In the moving load model, the impulse signal is consistently impacted by the reflection generated from passing vehicles, while the bridge response is influenced by complex impulse signals, resulting in a lower accuracy

in damage prediction, as anticipated, summarizes the prediction accuracy of each model, with XGBoost showing the best damage prediction ability for the four crack cases.



**Figure 6.20.** The damage prediction results after processing the moving load FE simulation results from (a) BP, (b) an SVM, (c) a decision tree, and (d) the XGBoost algorithm

## CHAPTER 7 CONCLUSION

The dissertation focuses on infrastructure damage detection using advanced AI techniques, encompassing IE signals, WRV, and structural performance. In chapter 3, the damage identification using various impact echo signals, including time-domain, frequency-domain, and STFT signals is explored. Unlike conventional methods reliant on a specific frequency range (1kHz to 5kHz), which can be influenced by testing conditions, this research explores the potential of using diverse input signals for damage identification through CNN. Among the four models tested, the one employing composite images generated from frequency and STFT data achieves the highest testing accuracy. The study emphasizes the significance of validation accuracy as a critical metric for assessing model performance, ensuring it doesn't overfit the data. Furthermore, the research investigates the efficiency of CNN models with different image resolutions and varying numbers of images. The slope analysis between computation time and resolution provides crucial insights, highlighting thresholds for optimal input settings. By exploring multiple input signal types and evaluating their performance comprehensively, this study contributes valuable insights into enhancing the accuracy and efficiency of damage identification methods using advanced ML techniques.

Based on our findings, the following conclusions can be drawn:

- The multi-input approach combining frequency and STFT referred to as FSTFT achieves an impressive 87% testing accuracy in identifying damages during impact echo tests.
- Among all CNN models, FSTFT demonstrates superior capabilities in identifying specific damages (D1 to D6), while F and TFSTFT exhibit comparable accuracy in identifying D3 to D5.
- Validation accuracy serves as a crucial metric, ensuring the models effectively handle unknown inputs without overfitting issues, a vital aspect studied in all four models.
- Efficiency analysis reveals a stark trade-off: a significant exponential increase in computation time yields only a marginal 2% increase in accuracy.
- Considering the thresholds of accuracy and computation time is paramount in the efficiency analysis, enabling the identification of an optimal balance between speed and precision.

- The efficiency analysis demonstrates that an image resolution of 200x200 with 100 images strikes the most balanced equilibrium between accuracy and computation time, providing valuable insight for practical implementation.

In chapter 4, the study introduces a novel approach to signal extraction leveraging ANN and CNN to enhance the accuracy of impact echo results while mitigating the influence of environmental factors in field test data. Through deep learning, it was determined that a signal duration of 1 millisecond and a starting time of 0.1 millisecond after the zero-crossing point yielded optimal results for the DL model. These identified DL-suggested signal parameters were incorporated into the impact echo delamination calculation, leading to significant improvements as validated through a thorough examination of both time-domain and frequency-domain signals. This methodology effectively reduced signal dismissals and exaggerations caused by variations in impactor energy and ground surface conditions. The application of this approach, as detailed in this chapter, offers a more accurate and reliable analysis of impact echo results when dealing with field test data. In future research, expanding the signal extraction process to incorporate various signal types, such as short-time Fourier (STFT), and employing multi-input CNN algorithms could enhance the dataset and improve the model's identification capabilities. Based on our findings, the following conclusions can be drawn:

- In a comprehensive DL analysis of NDT time-domain signals, an optimal image parameter for accurate damage identification was identified.
- Deep learning revealed that a signal duration of 1 millisecond and a starting time of 0.1 millisecond after the zero-crossing point provided optimal outcomes for the DL model.
- The signal extraction approach significantly reduces signal dismissals and exaggerations caused by environmental interferences in field test data.
- Utilizing DL-suggested signal parameters ensures a more precise and dependable analysis of impact echo results, particularly in the context of field test data.

In chapter 5, the comparison and examinations of wave response patterns between HM and IHM is studied. The wave response pattern in HM is validated through laboratory tests and analytical solutions, revealing differences in noise and wave reflection between experimental and simulated signals. Despite dissimilarities, the signal frequency and waveform align for the first three significant cycles, affirming the suitability of the FE model for generating subsequent wave response pattern data. The study explores variations in the wave response pattern for cracks at different depths, noting distinct peaks for shallower and deeper cracks. In IHM, larger aggregates exert a greater impact on the wave response pattern, causing more energy attenuation in the forward wave compared to smaller aggregates. The study utilizes ML techniques, including ALE and CNN, to analyze and predict cracks with high accuracy. The application of GANs contributes to generating synthetic samples for training purposes. While shedding light on internal damage in IHM using AI technology, future work considerations include developing an analytical solution for IHM, investigating varied depth-to-crack specimens for detailed wave response patterns, and exploring different impact sizes during laboratory testing to enhance understanding and control of wave response characteristics. The summary of this paper is shown below:

- In HM, the shallower crack ( $a=12.7\text{mm}$ ) showed a peak at a lower impulse frequency, while the deeper crack ( $a=50.8\text{mm}$ ) exhibited its WRV peak at a higher frequency.
- In IHM, larger aggregates exerted a more significant impact on the WRV pattern than smaller aggregates, causing increased energy attenuation in the forward wave.
- The combination of random aggregate size and distribution results in a more intricate WRV pattern.
- GANs can generate synthetic samples based on images, creating a robust database for training purposes.
- ALE elucidates the ML model's black box process, specifically highlighting the impact of aggregate information on the WRV.
- The CNN achieved an impressive training accuracy of 87.5% when using WRV images with varying internal crack information.

The study presents the framework for detecting internal damage in IHM using AI technology. However, future work should focus on 1) developing an analytical solution for IHM to enhance parameter control and deepen understanding of WRV; 2) exploring varied depth-to-crack specimens for more detailed WRV curves; 3) employing different impact sizes in laboratory testing to generate detailed WRV patterns with varying impact frequency. These tasks aim to facilitate a more comprehensive study of the WRV pattern.

In chapter 6, the study's primary focus was to develop a moving load FE model capable of simulating vehicle motion using KCE and conducting a comprehensive parameter study of the model. To ensure the model's accuracy, it was verified against VBI analytical solutions, laboratory tests, and field tests. The study examined solutions using ML methods for damage prediction and compared the results of the static load prediction model with the moving load prediction model. The latter showed lower accuracy due to the complicated nature of the structural response signal that serves as input for the ML model. XGBoost, particularly in deep learning, had the highest accuracy for both models, with 92.6% for the static load and 83.3% for the moving load. The results highlight the potential for improved damage prediction through the use of BWIM responses. The main findings are summarized as follows.

- The KCE-based simulation for a moving load is feasible. The structural responses were verified with a VBI analytical solution, and two laboratory tests and field test supported by parameter studies. The loss (%) of maximum displacement between FE simulation results and VBI analytical solutions shows a similar value.
- Feature selection is required for ML to avoid overfitting and to minimize the situation where the ML prediction model is not well trained to increase the accuracy of damage prediction.
- The prediction accuracy of the moving load is lower than static load accuracy because the structural response to a moving load is much more complicated than the response to a static load model
- By comparing the most recent ML methods, XGBoost combined the assembly decision tree and gradient boosting technology, allowing us to predict damage with the highest accuracy.

- The series of procedures presented in this report estimates and predicts structural damage from BWIM signals, which provides a new strategy for lowering SHM. Such information will help structural engineers take preventive or proactive actions to improve the drivers' safety and protect and preserve the transportation infrastructure.

The series of procedures presented in this chapter estimates and predicts structural damage from BWIM signals, which provides a new strategy for lowering SHM. Such information will help structural engineers take preventive or proactive actions to improve the drivers' safety and protect and preserve the transportation infrastructure. Further research of the KCE concept with a moving load will be further verified through the field test with actual semi-truck. And consider two-way vehicle effect. It is possible to monitor the health condition of a bridge structure in real-time by using BWIM responses.

## REFERENCE

- Achenbach, J. D., & Brind, R. J. (1981). Scattering of Surface Waves by a Sub-Surface Crack. *Journal of Sound and Vibration*, 76(1), 43–56. [https://doi.org/https://doi.org/10.1016/0022-460X\(81\)90289-3](https://doi.org/https://doi.org/10.1016/0022-460X(81)90289-3)
- Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1), 100004. <https://doi.org/https://doi.org/10.1016/j.ijime.2020.100004>
- Baggens, O., & Ryden, N. (2015). Systematic errors in Impact-Echo thickness estimation due to near field effects. *NDT and E International*, 69, 16–27. <https://doi.org/10.1016/j.ndteint.2014.09.003>
- Carino, N. J. (2001). The Impact-Echo Method: An Overview. In *Structures 2001: A Structural Engineering Odyssey*, 109, 1–18. [https://doi.org/10.1061/40558\(2001\)15](https://doi.org/10.1061/40558(2001)15)
- Chen, H. F. (2009). In silico log p prediction for a large data set with support vector machines, radial basis neural networks and multiple linear regression. *Chemical Biology and Drug Design*, 74(2), 142–147. <https://doi.org/10.1111/j.1747-0285.2009.00840.x>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/https://doi.org/10.1145/2939672.2939785>
- Finotti, R. P., Cury, A. A., & Barbosa, F. de S. (2019). An SHM approach using machine learning and statistical indicators extracted from raw dynamic measurements. *Latin American Journal of Solids and Structures*, 16(2), 1–17. <https://doi.org/10.1590/1679-78254942>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, 63(11), 27. <http://arxiv.org/abs/1406.2661>

- Gregory, R. D. (1975). The non-existence of standing modes in certain problems of linear elasticity. *Mathematical Proceedings of the Cambridge Philosophical Society*, 77(2), 385–404. <https://doi.org/10.1017/S0305004100051203>
- Gucunski, N., Slabaugh, G., Wang, Z., Fang, T., & Maher, A. (2008). Impact echo data from bridge deck testing: Visualization and interpretation. *Transportation Research Record*, 2050(2050), 111–121. <https://doi.org/10.3141/2050-11>
- Hirmand, M., Vahab, M., & Khoei, A. R. (2015). An augmented Lagrangian contact formulation for frictional discontinuities with the extended finite element method. *Finite Elements in Analysis and Design*, 107, 28–43. <https://doi.org/10.1016/j.finel.2015.08.003>
- Hishamuddin, M. A. H., Zin, S. M. M., & Yusup, E. M. (2021). Non-Destructive Measurement and Evaluation of Surface Cracks Using Ultrasonic Rayleigh Waves - A Review. *Journal of Physics: Conference Series*, 2129(1). <https://doi.org/10.1088/1742-6596/2129/1/012038>
- Hou, L., & Li, Z. (2020). Fault diagnosis of rolling bearing based on tunable Q-Factor wavelet transform and convolutional neural. *International Journal of Online and Biomedical Engineering*, 16(2), 47–61. <https://doi.org/10.3991/ijoe.v16i02.11953>
- Kang, S., Ham, S., & Kim, K. J. (2021). An analytical, numerical, and experimental study of Rayleigh wave scattering for internal vertical crack evaluation. *Construction and Building Materials*, 306, 124838. <https://doi.org/10.1016/j.conbuildmat.2021.124838>
- Kang, S., Wu, Y. C., David, D. S., & Ham, S. (2022). Rapid damage assessment of concrete bridge deck leveraging an automated double-sided bounce system. *Automation in Construction*, 138, 104244. <https://doi.org/10.1016/j.autcon.2022.104244>
- Kang, S., Wu, Y. C., & Ham, S. (2020). Singular integral solutions of analytical surface wave model with internal crack. *Applied Sciences*, 10(9), 3129. <https://doi.org/10.3390/app10093129>

- Kulkarni, P. S., Londhe, S. N., & Deo, M. C. (2017). Artificial Neural Networks for Construction Management: A Review. *Journal of Soft Computing in Civil Engineering*, 1(2), 70–88. <https://doi.org/10.22115/SCCE.2017.49580>
- Liu, Y., Xu, H., Ma, X., Wang, D., & Huang, X. (2023). Study on Impact–Echo Response of Concrete Column near the Edge. *Applied Sciences (Switzerland)*, 13(9), 5590. <https://doi.org/10.3390/app13095590>
- Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 14–23. <https://doi.org/10.1002/widm.8>
- Oh, T., Popovics, J. S., Ham, S., & Shin, S. W. (2012). Practical finite element-based simulations of nondestructive evaluation methods for concrete. *Computers and Structures*, 98–99, 55–65. <https://doi.org/10.1016/j.compstruc.2012.02.003>
- Peris-Sayol, G., Paya-Zaforteza, I., Balasch-Parisi, S., & Alós-Moya, J. (2017). Detailed Analysis of the Causes of Bridge Fires and Their Associated Damage Levels. *Journal of Performance of Constructed Facilities*, 31(3). [https://doi.org/10.1061/\(asce\)cf.1943-5509.0000977](https://doi.org/10.1061/(asce)cf.1943-5509.0000977)
- Reilly, J. L., Cook, W. D., Josée, B., & Mitchell, D. (2014). Effects of Delamination on the Performance of Two-Way Reinforced Concrete Slabs. *Journal of Performance of Constructed Facilities*, 28(4). [https://doi.org/10.1061/\(ASCE\)CF](https://doi.org/10.1061/(ASCE)CF)
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>
- Simo, J. C., & Laursen, T. A. (1992). An augmented Lagrangian treatment of contact problems involving friction. *Computers and Structures*, 42(1), 97–116. [https://doi.org/10.1016/0045-7949\(92\)90540-G](https://doi.org/10.1016/0045-7949(92)90540-G)

- Wickham, G. R. (1977). The forced two-dimensional oscillations of a rigid strip in smooth contact with a semi—infinite elastic solid. *Mathematical Proceedings of the Cambridge Philosophical Society*, *81*(2), 291–311. <https://doi.org/10.1017/S0305004100053378>
- Yang, Y. B., Li, Y. C., & Chang, K. C. (2014). Constructing the mode shapes of a bridge from a passing vehicle: A theoretical study. *Smart Structures and Systems*, *13*(5), 797–819. <https://doi.org/10.12989/sss.2014.13.5.797>
- Yoon, Y. G., Kim, C. M., & Oh, T. K. (2022). A Study on the Applicability of the Impact-Echo Test Using Semi-Supervised Learning Based on Dynamic Preconditions. *Sensors*, *22*(15), 5484. <https://doi.org/10.3390/s22155484>
- Yun, J., Kang, S., Tehrani, A. D., & Ham, S. (2020). Image analysis and functional data clustering for random shape aggregate models. *Mathematics*, *8*(11), 1–21. <https://doi.org/10.3390/math8111903>