

## Article

# Computerized Hammer Sounding Interpretation for Concrete Assessment with Online Machine Learning

Jiaxing Ye <sup>1,\*</sup>, Takumi Kobayashi <sup>2</sup>, Masaya Iwata <sup>2</sup>, Hiroshi Tsuda <sup>1</sup> and Masahiro Murakawa <sup>2</sup><sup>1</sup> National Metrology Institute of Japan (NMIJ), The National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 2, Tsukuba 305-8568, Japan; hiroshi-tsuda@aist.go.jp<sup>2</sup> Artificial Intelligence Research Center (AIRC), The National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8561, Japan; takumi.kobayashi@aist.go.jp (T.K.); m.iwata@aist.go.jp (M.I.); m.murakawa@aist.go.jp (M.M.)

\* Correspondence: jiaxing.you@aist.go.jp; Tel.: +81-29-861-2850

Received: 12 January 2018; Accepted: 5 March 2018; Published: 9 March 2018

**Abstract:** Developing efficient Artificial Intelligence (AI)-enabled systems to substitute the human role in non-destructive testing is an emerging topic of considerable interest. In this study, we propose a novel hammering response analysis system using online machine learning, which aims at achieving near-human performance in assessment of concrete structures. Current computerized hammer sounding systems commonly employ lab-scale data to validate the models. In practice, however, the response signal patterns can be far more complicated due to varying geometric shapes and materials of structures. To deal with a large variety of unseen data, we propose a sequential treatment for response characterization. More specifically, the proposed system can adaptively update itself to approach human performance in hammering sounding data interpretation. To this end, a two-stage framework has been introduced, including feature extraction and the model updating scheme. Various state-of-the-art online learning algorithms have been reviewed and evaluated for the task. To conduct experimental validation, we collected 10,940 response instances from multiple inspection sites; each sample was annotated by human experts with healthy/defective condition labels. The results demonstrated that the proposed scheme achieved favorable assessment accuracy with high efficiency and low computation load.

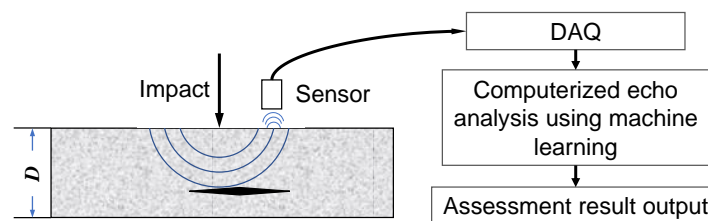
**Keywords:** non-destructive evaluation; hammer sounding; audio signal processing; machine learning; online learning

## 1. Introduction

Aging infrastructure poses significant challenges to human society. It is indispensable to perform the efficient non-destructive evaluation (NDE) to ensure the safety of those critical structures. In this study, we focus on the hammer sounding test, which is one of the most conventional NDE methods for assessment of concrete structures due to the low-cost and high-efficiency [1,2]. In the setting of hammer sounding, a field engineer generates surface impact using a handy hammer and then determines structural condition by listening to the response. It is evident that such judgement is highly subjective and relies on individual experience; thus, it leaves the assessment judgement results open to human errors. Extensive research efforts were made to develop efficient computerized hammer sounding systems to alleviate human efforts, as well as to eliminate the human errors [3–5].

A typical computerized hammer sounding interpretation system consists of two parts, the hardware and data analysis computing system [6,7]. Impact on the surface produces P- and S-waves that travel into the target structure. Then the responses are captured by a transducer nearby, i.e., an air-coupled microphone. Through data acquisition (DAQ) process, the data is analyzed by a program, and the condition assessment result is presented. Figure 1 shows the schematic flow.

This paper mainly addresses the issue of devising an efficient machine learning algorithm for hammer sounding data investigation.



**Figure 1.** The general processing flow of computerized hammering response investigation.

Current learning schemes applied for the hammering response analysis are restricted to the standard batch setting, which assumes that both training and input testing hammer responses reside in the same feature space with the static statistical characteristic; hence, model training can be performed over the pre-collected laboratory-scale database [4,6,7]. In practice, however, such an assumption does not hold. The patterns of response signal can alter significantly with the specifications of concrete structures under evaluation, such as material, shape and years of service [1]. From the viewpoint of machine learning, these factors would make the posterior distribution of the test data drift from that of the pre-collected training samples; thus, degrading the hammer sounding analysis performance.

In this study, we adopt an alternative hypothesis which admits the pre-collected training dataset only covers small range of the complete distribution; moreover, we propose a new formulation of response pattern classification with the online learning paradigm, where efficient model updating schemes have been exploited to minimize the cumulative prediction loss suffered along with the continuous input of data with expert labels. Online learning is a well-established learning scheme which has both theoretical and practical appeals [8,9] and it is particularly well-suited to the hammer sounding, since the large-scale response data can be accessed only in a sequential way.

It is noteworthy that our ultimate goal is to develop an efficient hammering response investigation system with near-human accuracy. In the hammering test, humans are capable of discerning the defect-induced responses of various concrete structures by auditory perception. In this study, we propose an AI-enabled computing system by adopting a formulation of binary classification, which produces labels to indicate healthy or defective concrete, respectively. At validation stage, a loss between the predicted results and expert labels has been employed to compare the performance of the proposed approach to that of humans. The main contributions of the proposed approach can be summarized as follows:

- The objective of this study is to build an efficient hammer sounding analysis system for concrete hammering inspection. To this end, a novel online learning framework had been proposed, which can effectively characterize discriminant information from large-scale response spectrum data in an incremental way.
- Various state-of-the-art online learning algorithms have been reviewed and evaluated for the application of response pattern classification. The side-by-side comparison results can inspire other applications with streaming data input, not limited to the hammer sounding analysis discussed in this study.
- Unlike conventional studies which commonly conduct experiments on laboratory-scale data, a massive dataset has been created during this study, which includes more than 10,000 samples collected from different types of concrete structures. Moreover, each instance has been annotated by professional inspectors with healthy/defective label. The database laid solid fundamentals for learning scheme validation.

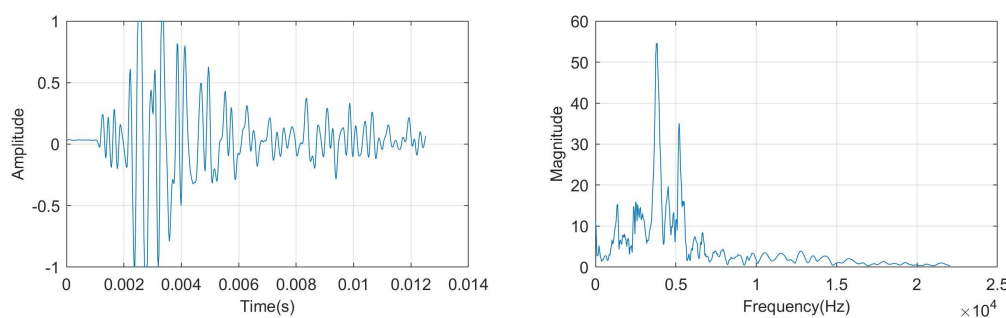
## 2. Related Work

In this section, we present a review of the previous studies conducted on hammering response analysis for concrete condition assessment. The review is comprised of two parts: the first is fundamental research towards the impact-echo method, which is closely related to this study, and the latter is recent advancement in developing a computerized hammering response investigation system with machine learning techniques.

### 2.1. Impact-Echo Method and Air-Coupled Hammer Sounding Inspection

The initial literature describing hammering inspection presented is from the 1970s, and subsequently, extensive studies had been carried out in both theoretical and experimental aspects [2,10,11]. Echo signal analysis is commonly performed through Fourier analysis. Although advanced methods have been employed in recent works, such as using Wavelet transform [12] to alleviate poor time-frequency resolution in Fourier spectra and fuse impulse response with 3D laser scanning for delamination detection [13], Fourier analysis still dominates in practice due to its efficient implementation.

In Figure 2, we present one example of hammering response waveform and its Fourier spectrum, respectively.



**Figure 2.** Hammer sounding waveform (left) and its Fourier spectrum (right).

A well-known formula to determine a void beneath surface of concrete is proposed by [1]:

$$d = \beta \frac{C_p}{2f_{peak}} \quad (1)$$

where  $f_{peak}$  denotes the peak frequency of response signal spectrum,  $C_p$  is the velocity of the longitudinal,  $\beta$  is constant of 0.96 for plate-shape structures wave according to [1] and  $d$  represents depth of inside void. However, some recent studies reveal that the availability of formula (1) is constrained by the size and flatness of the defect area, e.g., if void is not parallel to surface, the resonance behaves differently and thus Equation (1) fails to estimate void depth [14]. Impact-echo is initially a contact inspection method, which is quite time-consuming to fix transducer, especially when dealing with large structures. To enhance the efficiency, a new suggestion emerged to apply air-coupled sensor in impact-echo [15,16]. A designated transducer is employed to capture acoustic response from concrete structure through the air. Furthermore, experimental results show the air-coupled sensor is comparable to contact sensors for delamination detection and grouting quality evaluation tasks. Until now, impact-echo method, due to its high diagnosis accuracy and favorable stability, remains to be an active research topic in the non-destructive test field, and efforts will be continuously delivered to the topic.

### 2.2. Data-Driven Hammer Sounding Investigation System for Non-Destructive Test of Concrete Structure

The last five years have seen remarkable progress in machine learning research, and a spreading trend emerged to develop human-level machine learning systems to relieve people from laborious and

exhausting tasks in the structural health monitoring field [3,5]. In order to substitute the human role in hammering response interpretation, great efforts were carried out to establish a data-driven machine learning system to discern anomalous responses from healthy ones [4,6,7]. We present a review on current research status as follows.

The early systems commonly tackle the hammer sounding investigation problem with statistical pattern classification, in which the response spectra were used as feature vector and various conventional classifiers were employed, such as Gaussian mixture models (GMM) [4], Artificial Neural Network (ANN) [17] and Support Vector Machines (SVM) [18], to characterize discriminant information of healthy / defective responses. In recent years, significant progress has been made in noise robust feature representation learning. Advanced signal descriptors developed by the bag-of-words model (BoW model) [6] and sparse coding approaches [7] have proved to be effective for anomalous response identification under a hostile acoustic environment. It is noteworthy that this literature commonly assumes that all training and test hammer responses are sampled from the same population; the experimental dataset was confined to be the laboratory-scale as well. It is anticipated to be problematic when we directly apply the analysis model trained by lab-scale data to the practical hammer sounding test, because the pre-collected training data is quite limited to rendering sufficient discriminant information to deal with complex real data.

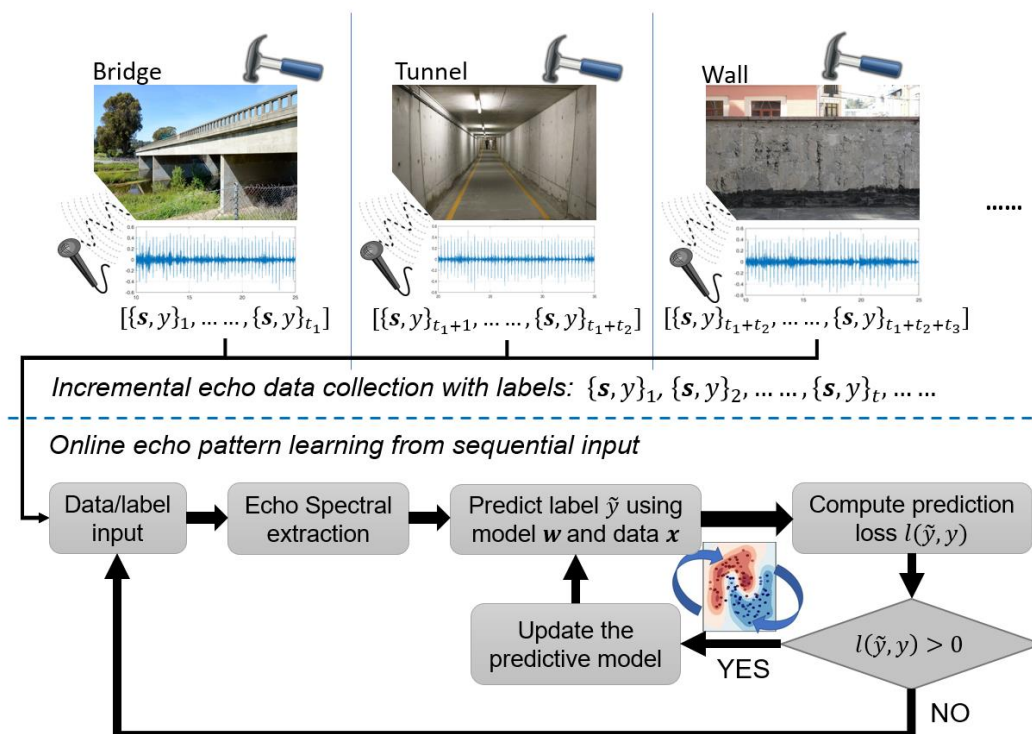


Figure 3. Flow chart of online learning formulation for hammering response investigation.

### 3. The Proposed Online Learning Framework for Hammering Response Pattern Analysis

We introduce the details of the proposed online machine learning-enabled hammering response analysis system in this section. The processing flow has been shown in Figure 3. We assume that the responses have been received in a streaming way  $\{s, y\}_t, t \in [1, \dots, T]$  and all the data had been annotated by a professional inspector with healthy or anomalous labels. Notably, during the data collection, the material, shape, and other specifications of target structures can vary from place to place. For instance, in Figure 3 the inspection place shifted from bridge to tunnel, meanwhile the captured data/label indexes range varied from  $1 \sim t_1$  to  $t_1 + 1 \sim t_1 + t_2$ , respectively. Our goal is to continuously update the analysis model so as to achieve near-human performance for flaw-induced

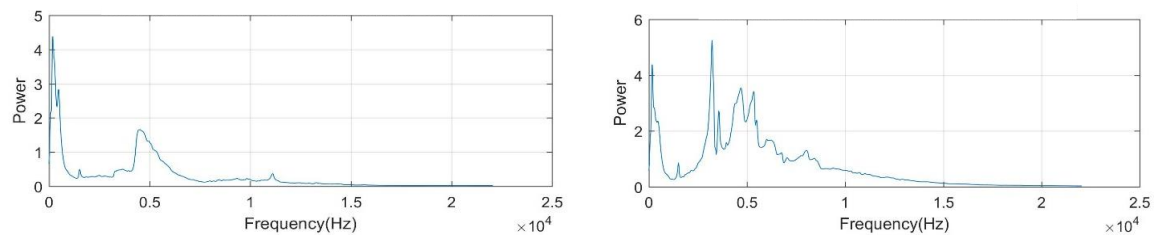
hammering response identification. To this end, it is crucial to devise the efficient online learning scheme and various state-of-the-art algorithms have been reviewed and compared. The details are presented as follows.

### 3.1. Feature Extraction

Over decades, spectral analysis has been the dominant method for hammering response characterization [10,11], in which Fourier transform (FT) is employed to generate the spectrum. Then, further pattern investigation can be performed. The Fourier analysis can be expressed as:

$$x(f) = \int_{-\infty}^{\infty} s(t)e^{-2\pi ift} dt \quad (2)$$

where  $s(t)$  is waveform of hammering response to be analyzed and  $x(f)$  is the extracted spectrum. In the following contents, we will use  $\mathbf{x}_t \in \mathbb{R}^d, t \in [1, \dots, T]$  to denote the spectrum collected at time stamp  $t$ , with  $d$  frequency-bins. Our pattern classification process is based on Fourier spectrum representation of response signal. In Figure 4, we present two spectrum examples, which were collected from normal and defective concrete structures, respectively. According to the plots, differences in spectral distributions can be clearly observed. In addition, at a low frequency region below 500 Hz, high noise power can be seen which is irrelevant to the specimen. We employed high pass filter to eliminate the ambient noises presented in the band lower than 500 Hz. Meanwhile, to reduce feature dimension, we also discarded the frequency bands higher than 15 kHz, since there existed no discriminant information.



**Figure 4.** Examples of hammering response spectrum: normal (left) and defective cases (right).

### 3.2. Online Learning Algorithms in Evaluation

As the core feature of this research, online learning algorithms are employed to deal with the case that hammering response data (with expert annotations) arrives incrementally with time stamps. Concretely, at time  $t$ , the online learning algorithm analyses the input data and expert label, i.e.,  $\{\mathbf{x}_t, y_t\}$ , through three steps: the first is to predict its label  $\hat{y}_t \in \{-1, +1\}$ , in which the two digits represent defective and healthy status, respectively. Then, we compare the predicted label  $\hat{y}_t$  with true label  $y_t \in \{-1, +1\}$  by using a well-defined loss function  $l(y_t, \hat{y}_t)$ . Finally, if the computed prediction loss exceeds a threshold, the classification model will be updated in an analytical way. Overall, the cumulative mistake through whole data stream can be minimized. In this section, we first present a general algorithmic framework of online machine learning for hammer sounding discriminant analysis in Algorithm 1. Then, we explicitly introduce the algorithms employed for this application.

**Algorithm 1.** Online Learning ( $\mathbf{x}_t, y_t, \mathbf{w}_t$ )**Initialization**  $\mathbf{w}_1 \leftarrow 0$ **for**  $t = 1, 2, \dots, T$ 

**do**  $\left\{ \begin{array}{l} \text{input hammering response is received with label: } \mathbf{x}_t \in \mathcal{X}, y_t \in \mathcal{Y} \\ \text{predict label of input hammering response: } \hat{y}_t = \text{sign}(f(\mathbf{x}_t; \mathbf{w}_t)) \\ \text{compute the prediction loss: } l(\mathbf{w}_t; (\mathbf{x}_t; \mathbf{w}_t)) \\ \text{if } l(\mathbf{w}_t; (\mathbf{x}_t; \mathbf{w}_t)) > 0 \\ \quad \text{then update the classification model:} \\ \quad \mathbf{w}_{t+1} \leftarrow \Delta((\mathbf{w}_t; (\mathbf{x}_t; \mathbf{w}_t))) \end{array} \right.$

**return** ( $\mathbf{w}_{t+1}$ )**3.2.1. Perceptron**

The perceptron algorithm is the initial method for online learning [19]. Given the linearly separable data, the method can converge to a hyperplane to shatter the different classes in a finite number of updates. The prediction function of perceptron is very simple:  $\hat{y}_t = \text{sign}(\mathbf{w}^\top \mathbf{x}_t)$  and the updating rule will be conducted as follows:

$$\mathbf{w}_{t+1} = \mathbf{w} + y_t \mathbf{x}_t \quad \text{if } \hat{y}_t \neq y_t \quad (3)$$

There is neither parameter nor optimization constraints in the perceptron algorithm. The perceptron algorithm has several limitations. First, it can only classify linearly separable sets of vectors. If the class-conditional data distribution is inherently nonlinear, perceptron will never reach a point where all vectors are classified properly. Second, since there is no constraint applied during model training, perceptron is vulnerable to noise. To alleviate the problems, substantial modification had been carried out. Representative works can be referred in [9].

**3.2.2. Online Gradient Descent (OGD)**

Gradient descent updating is another efficient approach for online learning [20]. In this evaluation, we selected logistic loss to measure the prediction error:

$$l(\mathbf{w}; (\mathbf{w}_t, y_t)) = \log(1 + \exp(-y_t(\mathbf{w}_t \cdot \mathbf{x}_t))) \quad (4)$$

Subsequently, the updating rule can be represented as:

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta_t y_t \mathbf{x}_t \cdot \frac{1}{1 + \exp(y_t(\mathbf{w}_t \cdot \mathbf{x}_t))}, \quad \eta_t = C / \sqrt{t} \quad (5)$$

**3.2.3. Passive-Aggressive Learning Algorithm [PA]**

Passive-Aggressive Learning Algorithm [PA] is one state-of-the-art first order online learning approach. The optimization formulation can be expressed as follows:

$$\mathbf{w}_{t+1} = \underset{\mathbf{w} \in \mathbb{R}^d}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{w} - \mathbf{w}_t\|^2, \quad \text{s.t. } l(\mathbf{w}; (\mathbf{w}_t, y_t)) \quad (6)$$

where the loss function is based on the hinge loss:

$$l(\mathbf{w}; (\mathbf{w}_t, y_t)) = \begin{cases} 0 & \text{if } y_t(\mathbf{w} \cdot \mathbf{x}_t) \geq 1 \\ 1 - y_t(\mathbf{w}_t \cdot \mathbf{x}_t) & \text{otherwise} \end{cases} \quad (7)$$



The updating rule can be derived analytically:

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta_t^{PA} y_t \mathbf{x}_t, \quad \text{where } \eta_t^{PA} = \frac{l(\mathbf{w}_t; (\mathbf{w}_t, y_t))}{\|\mathbf{w}_t\|^2}. \quad (8)$$

In addition, several variants of PA method had been investigated [21]. The core idea is to add slack variable  $\xi$ -induced penalty to handle non-separable cases.

### 3.2.4. The Second Order Perception (SOP)

Aiming at better characterizing the hammering response structure, advanced second-order online learning approaches were developed. Unlike the above-mentioned first-order algorithms, the second-order online learning is designated to exploit the underlying relationship between features. Concretely, it assumes the weight vector exhibits Gaussian distribution  $\mathbf{w} \sim \mathcal{N}(\mu, \Sigma)$ . At initialization stage, two additional hyperparameters are commonly set to  $\mathbf{w}_1 = 0, \Sigma_1 = a\mathbf{I}$ . Furthermore, the prediction function is noted as:

$$\hat{y}_t = \text{sign}(\mathbf{w}^\top \mathbf{x}_t), \quad \mathbf{w}_t = (\Sigma_t + \mathbf{x}_t \mathbf{x}_t^\top)^{-1} \mu_t \quad (9)$$

The following updating process is conducted as the predicted label is inconsistent with the true label:

$$\mu_{t+1} = \mu_t + y_t \mathbf{x}_t, \quad \Sigma_{t+1} = \Sigma_t + \mathbf{x}_t \mathbf{x}_t^\top \quad (10)$$

A representative work for second order perception can be referred to [22].

### 3.2.5. The Confidence-Weighted Learning Algorithm (CW)

CW method is an advanced second-order online learning [23]. In contrast to SOP approach, CW methods perform the Kullback-Leibler divergence minimization between the new weight distribution and the old one with constraint so that the probability of correct classification can be improved. The updating rule of CW is shown as below:

$$(\mu_{t+1}, \Sigma_{t+1}) = \underset{\mu, \Sigma}{\operatorname{argmin}} D_{KL}(\mathcal{N}(\mu, \Sigma), \mathcal{N}(\mu_t, \Sigma_t)), \quad \text{s.t. } Pr_{\mathbf{w} \sim \mathcal{N}(\mu, \Sigma)}[y_t(\mathbf{w} \cdot \mathbf{x}_t)] \geq \eta \quad (11)$$

A closed-form solution can be derived as:  $\mu_{t+1} = \mu_t + \alpha_t y_t \Sigma_t \mathbf{x}_t$ ;  $\Sigma_{t+1} = \Sigma_t - \beta_t \Sigma_t \mathbf{x}_t \mathbf{x}_t^\top \Sigma_t$ , where the updating coefficients can be calculated as follows:  $\alpha_t = \max\{0, \frac{1}{v_t \zeta}(-m_t \psi + \sqrt{m_t^2 \frac{\phi^4}{4} + v_t \phi^2 \zeta})\}$ ;  $\beta_t = \alpha_t \phi / (\sqrt{u_t} + v_t \alpha_t \phi)$ . More detail parameters setting discussion can be found in [23].

### 3.2.6. Adaptive Regularization of Weight Vectors (AROW)

Regularization is regarded as a useful trick to enhance both accuracy and robustness of the online learning algorithm. AROW method added an adaptive regularizer to restrict the sudden changes of weight during online learning [24]. The formulation of AROW is presented as follows:

$$(\mu_{t+1}, \Sigma_{t+1}) = \underset{\mu, \Sigma}{\operatorname{argmin}} D_{KL}(\mathcal{N}(\mu, \Sigma), \mathcal{N}(\mu_t, \Sigma_t)) + \frac{1}{2\gamma} l^2(\mu; (\mathbf{x}_t, y_t)) + \frac{1}{2\gamma} \mathbf{x}_t^\top \Sigma_t \mathbf{x}_t \quad (12)$$

where

$$l^2(\mu; (\mathbf{x}_t, y_t)) = (\max\{0, 1 - y_t(\mu \cdot \mathbf{x}_t)\})^2 \quad (13)$$

The updating coefficients can be obtained by solving optimization problem:

$$\mu_{t+1} = \mu_t + \alpha_t \Sigma_t y_t \mathbf{x}_t, \quad \Sigma_{t+1} = \Sigma_t - \beta_t \Sigma_t \mathbf{x}_t \mathbf{x}_t^\top \Sigma_t \quad (14)$$

$$\alpha_t = l(\mu_t; (\mathbf{x}_t, y_t))\beta_t, \beta_t = \frac{1}{\mathbf{x}_t^T \Sigma_t \mathbf{x}_t + r} \quad (15)$$

### 3.2.7. Soft Confidence-Weighted Learning (SCW-II)

SCW is a more advanced second-order learning algorithm that improves over the original CW by adding the capability to handles the non-separable cases, and also improves over AROW by adding the adaptive margin property [25]. The classification suffer loss for input hammer sounding data is defined as  $l_t = \max\{0, 1 - y_t \mathbf{w}_t^T \mathbf{x}_t\}$ . If  $l_t > 0$ , the classification model will be updated:

$$\mu_{t+1} = \mu_t + \alpha_t y_t \Sigma_t \mathbf{x}_t; \Sigma_{t+1} = \Sigma_t - \beta_t \Sigma_t \mathbf{x}_t \mathbf{x}_t^T \Sigma_t, \quad (16)$$

where

$$\begin{aligned} \alpha_t &= \min\{C, \max\{0, \frac{1}{v_t \zeta} (-m_t \psi + \sqrt{m_t^2 \frac{\phi^4}{4} + v_t \phi^2 \zeta})\}\}; \beta_t = \alpha_t \phi / (\sqrt{u_t} + v_t \alpha_t \phi) \\ u_t &= \frac{1}{4} (-v_t \alpha_t \phi + \sqrt{\alpha_t^2 v_t^2 \phi^2 + 4v_t})^2, v_t = \mathbf{x}_t^T \Sigma_t \mathbf{x}_t, m_t = y_t (\mu_t \cdot \mathbf{x}_t) \\ \gamma_t &= \phi \sqrt{\phi^2 m_t^2 v_t^2 + 4n_t v_t (n_t + v_t \phi^2)} \text{ and } n_t = v_t + \frac{1}{2C} \end{aligned} \quad (17)$$

### 3.3. Hammering Response Data Visualization

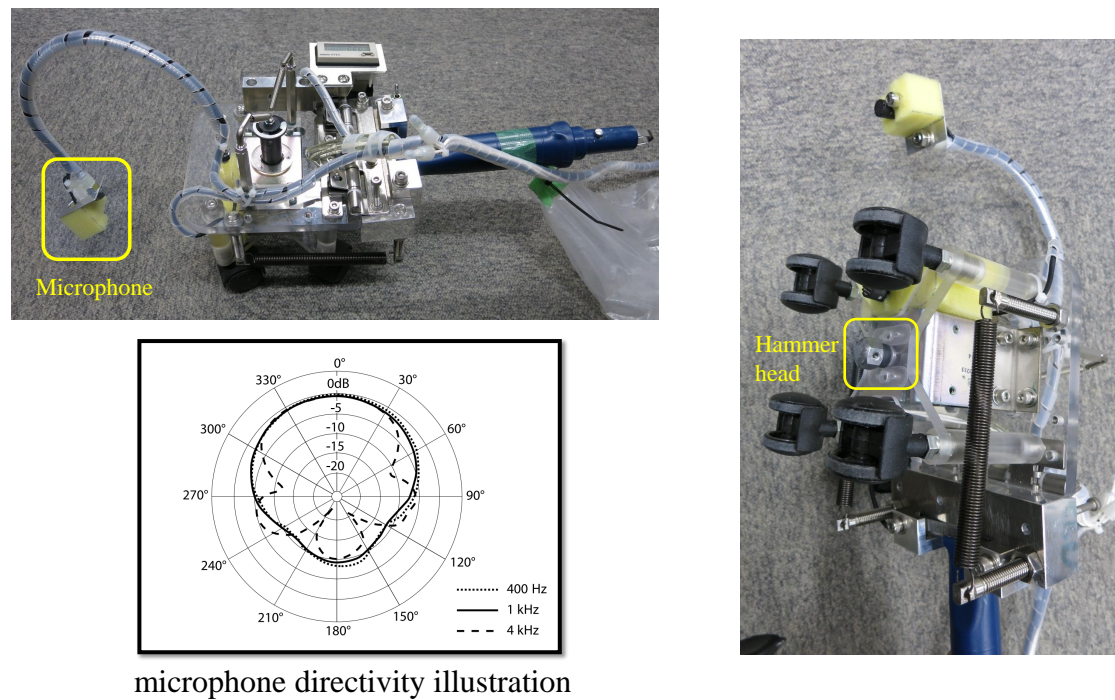
Data visualization is widely recognized as one integral part of today's data analysis systems, which makes complex data more accessible, understandable and usable. In our hammer sounding pattern investigation system, we incorporate data visualization function so as to let end-users browse and understand the massive data distributions. We adopted the fundamental method principal component analysis (PCA), which is a standard way of visualizing data. The basic principle of PCA is to find the low dimension linear subspace such that the variations of data can be maximized. The detail procedures can be found in [26]. We present the whole hammering response data visualization results in the experimental analysis section.

## 4. Experimental Validations

### 4.1. Data Collection

In this section, we introduce the hammer sounding dataset we created to evaluate the proposed system. First, we present the hardware we used for data collection in Figure 5, including solenoid hammer device and microphone. Then, Table 1 shows more detail specifications. For response data recording, the sampling rate was set to 44.1 kHz and resolution was fixed to 16-bit depth. We visited 12 inspection sites to capture hammer sounding data. Meanwhile, binary expert annotations, i.e., the response indicating normal or anomalous concrete, has also been collected. In Figure 6, we show the photos of two inspection sites. The defective area had been tagged with pink color by inspector. In addition, we marked multiple parallel lines in yellow, which explains the trace of hammering. Scanning speed was around 80 centimeters per minute (cm/min). The hammer area varies with location. As a result, we obtained 10,940 annotated hammer responses, among which 9349 are normal and 1591 are anomalous instances, respectively. The dataset laid fundamentals for further numerical analysis.

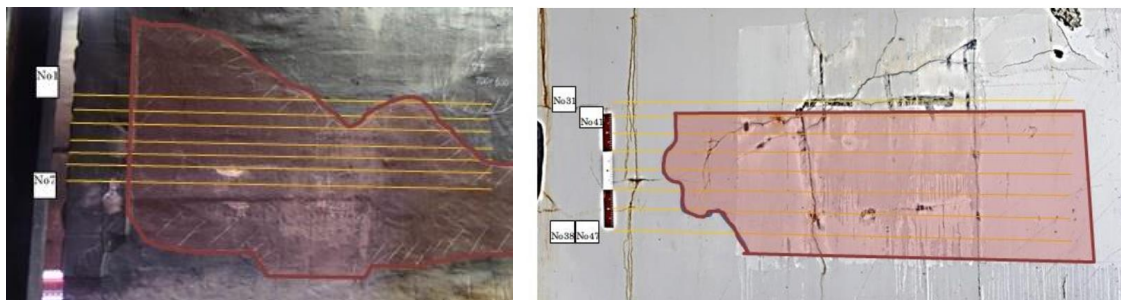




**Figure 5.** The hammer device and microphone used for data capture (upper, right) and the microphone directivity illustration.

**Table 1.** Summary of the hammer sounding hardware.

	Device	Specification
1	Hammer device	Solenoid: Takaha Kikou Co., Ltd., CB15670033
2	Transducer	Low-cost condenser microphone: ECM PC60
3	Recorder	Olympus Voice-Trek V-803



**Figure 6.** Photos of two working sites for hammering response data capture.

#### 4.2. Experimental Settings

At hammer response feature extraction stage, we determine the Fourier analysis window length to be 1024. Band pass filter is applied to focus on the frequencies ranging from 500 to 15,000 Hz. At online learning stage, parameter tuning plays a key role in achieving accurate pattern classification. In this study, we evaluate seven state-of-the-art online learning approaches with the massive data. The detail parameter settings are presented in Table 2.

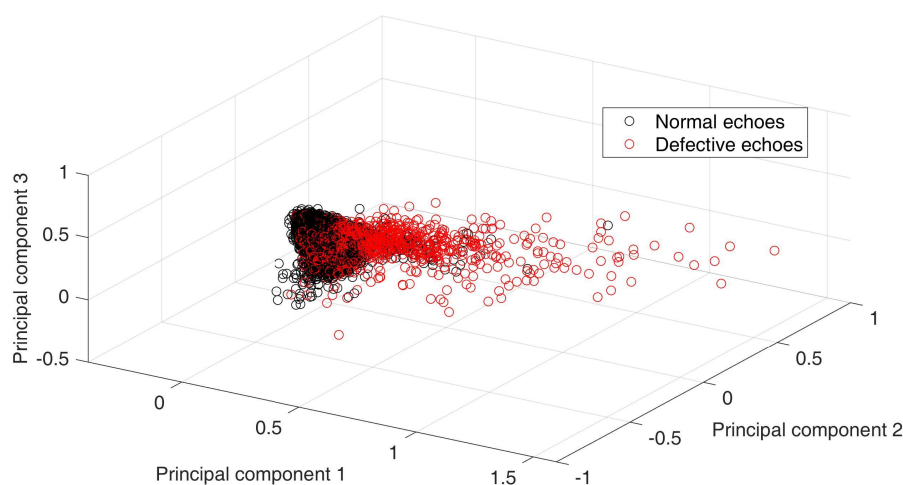
**Table 2.** Summary of the parameter setting by algorithms.

Index	Algorithm	C	$\alpha$	$\eta$	Others
1	Perceptron	/	/	/	parameter free
2	OGD	$C = 1$	/	/	$\eta_t = C/\sqrt{t}$
3	PA	/	/	/	parameter free
4	SOP	/	$\alpha = 1$	/	parameter free
5	CW	/	$\alpha = 1$	$\eta = 0.7$	$\Sigma = a * I$
6	AROW	$C = 1$	$\alpha = 1$	/	$r = C, \Sigma = \alpha * I$
7	SCW-II	$C = 1$	$\alpha = 1$	$\eta = 0.75$	$\Sigma = a * I$

The first parameter  $C$  governs the trade-off between the fitting loss term and regularization term in machine learning model training. In the second order algorithms, the parameter  $\alpha = 1$  is used to initialize the covariance matrix, i.e.,  $\Sigma = \alpha * I$ , where  $I$  is identity matrix. Parameter  $\eta$  is used to define loss function in the confidence-weighted learning algorithms, i.e., in CW and SCW-II. The experiments had been performed in the same vein as a real scenario, in which the labeled data was fed to the online learning system in a sequential manner. The experiments were conducted over 20 random permutations for the whole dataset. At each iteration, we divided the dataset into 15 sub data sets. During online learning, we recorded the evaluation results, i.e., response-wise classification accuracies and computation time costs, when one subset had been processed. Those information will help us understand the learning behavior of the algorithm. Finally, the results are presented by averaging of total 20 trials.

#### 4.3. Echo Data Visualization

As introduced in Section 3.3, data visualization is a useful approach to understanding the data. In Figure 7, we present the distribution of the dataset using principal component analysis (PCA). In the visualization, binary class labels were noted with different colors, i.e., the normal hammer responses were marked with black and flaw-induced ones were colored in red. According to the distribution plot, we have several major findings: 1. damage-induced responses produced more scattered distribution compared to that of healthy ones. It is reasonable because the damaged concrete usually generated more complex spectrum; 2. The boundary between normal and anomalous responses is not clear; in other words, there exists strong non-linearity between the two-class distribution. From the machine learning aspect, the methods which were designated to deal with the inherently nonlinear data may perform superiorly. Grounded on the above understanding of the data collection, we start the algorithmic analysis as follows.

**Figure 7.** Visualization of hammering response dataset using principal component analysis (PCA).

#### 4.4. Empirical Evaluation Results

In this part, we present results of experimental validation. The comparison has been drawn of three aspects: normal/defective hammer response classification accuracy, processing efficiency and computation complexity. As for the first comparison—accuracy, we adopted two metrics: mistake rate transition curve and the cumulative classification error rate. It can be anticipated that with more data being examined, the mistake rate would decrease monotonically. To exploit the performance of online analysis models, we presented the cumulative error rate after the whole online learning process was done. Figure 8 exhibited the overall errors statistics during the online learning process. First of all, by examining the overall mistakes, we found that second-order algorithms, i.e., SOP, CW, AROW and SCW-II usually outperform first-order algorithms, including perception, OGD and PA; also margin based algorithms, such as CW and SCW-II, usually outperform non-margin based methods.

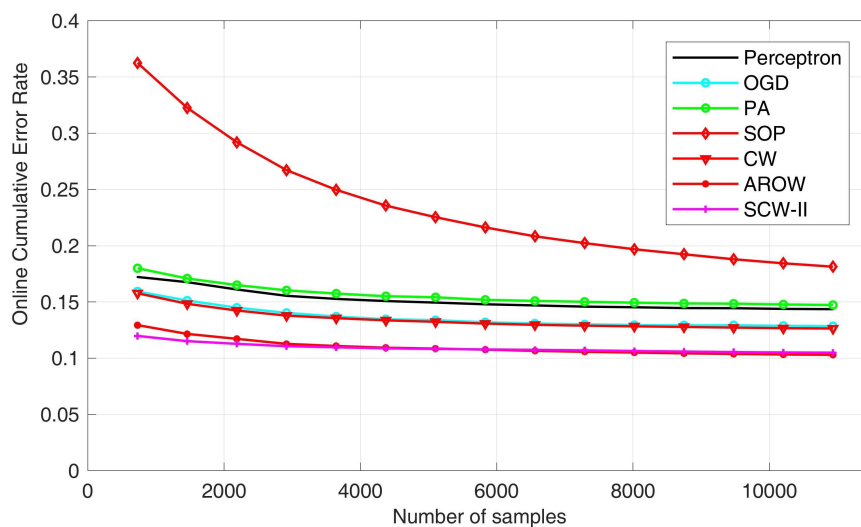


Figure 8. Summary of online cumulative classification error rate.

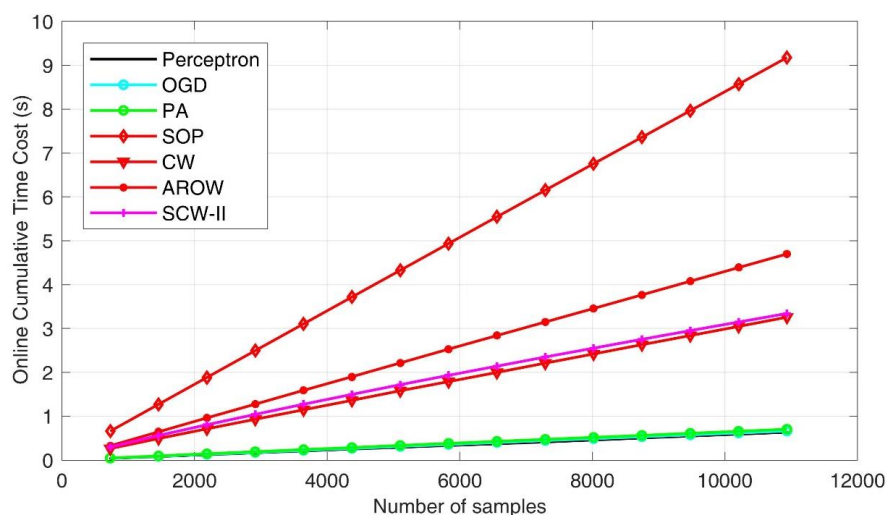
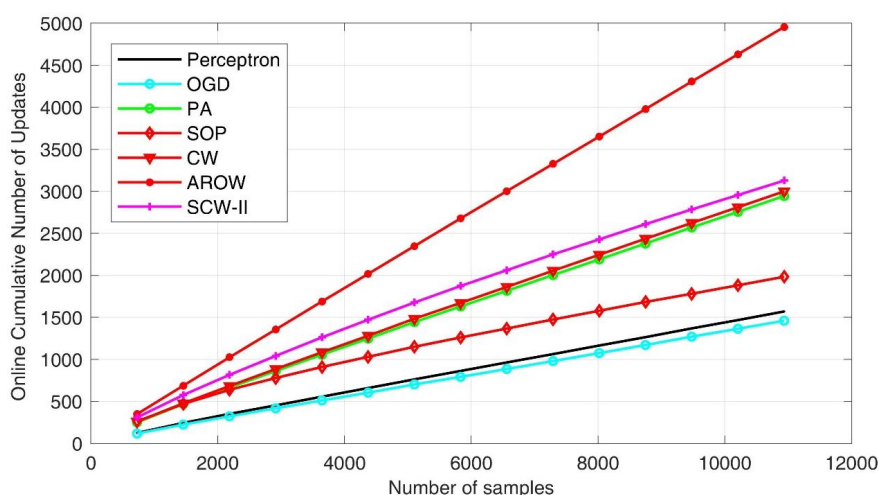


Figure 9. Processing time cost comparison.

As for the second evaluation criteria—computation efficiency, we presented the cumulative time costs for all the seven online learning algorithms under evaluation in Figure 9. We found that the first-order schemes exhibited superior efficiency due to their simpler formulation. The SOP method took the longest time in process, which is one of initial second order approaches. Confident-weighted

learning methods, including both CW and SCW, achieved favorable performance in balancing accuracy and efficiency. To demonstrate the complexity of online learning algorithms, we further showed cumulative number of updates. In general, fewer update steps indicates the algorithm is efficient in establishing a more robust pattern classification hyperplane such that input feature distribution shift can be accommodated. By examining Figure 10, we can see that first order methods usually produced smaller numbers of updates. However, the classification accuracies were inferior. AROW scheme made significantly larger number of updates, which can induce high time cost in processing.



**Figure 10.** Comparison of number of updating steps.

To clarify the comparison, we further prepared Table 3 to summarize the key experimental results, including cumulative mistake rate, size of support vectors (SVs) and cpu run time. Such quantitative information is complementary to the above charts. From the table, we found that SCW-II outperformed all other methods in hammer sounding pattern classification accuracy. Meanwhile, the method achieved superior efficiency among all the second order online learning algorithms. Besides, we also investigated the number of support vectors (SVs) used by different learning schemes. SVs are defined as the samples used to determined max-margin hyperplane for classification. Since the hammer responses are highly non-linear, we can see that larger numbers of SVs had been used at classification stage.

**Table 3.** Summary of all experimental results.

Algorithm:	Mistake Rate ( $M \pm \text{Std}$ )	Size of SVs ( $M \pm \text{Std}$ )	Cpu Time ( $M \pm \text{Std}$ )
Perceptron	$0.144 \pm 0.002$	$1570.8 \pm 25.2$	$0.647 \pm 0.056$
OGD	$0.128 \pm 0.005$	$1460.7 \pm 59.3$	$0.718 \pm 0.045$
PA	$0.147 \pm 0.002$	$2945.8 \pm 41.2$	$0.699 \pm 0.037$
SOP	$0.181 \pm 0.002$	$1983.8 \pm 27.0$	$9.708 \pm 0.608$
CW	$0.126 \pm 0.002$	$3000.1 \pm 36.6$	$3.460 \pm 0.248$
AROW	$0.115 \pm 0.004$	$6378.9 \pm 278.2$	$6.357 \pm 0.429$
SCW-II	$0.105 \pm 0.002$	$3128.4 \pm 61.6$	$3.577 \pm 0.244$

As a result, among all the compared algorithms, SCW-II produced the best performance in terms of accuracy; for other metrics including number of updates, and running time cost, it also outperformed other second order methods. Therefore, the method can be the optimal selection for the application of hammering response investigation. It is noteworthy that errors may exist in the labels, because inspectors usually take location information into account. It can be regarded as performing

region-based smoothing over each individual expert label nearby. In contrast, our quantitative evaluation was conducted in a point-wise way. Such a factor can be one major reason that led to the error rate over 10%. The empirical evaluation validated the effectiveness of the online learning approach.

## 5. Discussion and Conclusions

To tackle the growing problem of aging infrastructure, the need for non-destructive evaluation (NDE) methods with high efficiency and low-cost has become a priority. In this context, engineers are still using these old-fashioned hammer sounding tests that are very subjective and open to human error. In this paper, we attempted to develop the efficient learning framework which is able to mimic human expert ability in hammering response investigation. Specifically, we formulate the task by using binary statistical classification between flaw-induced responses and normal ones. In order to deal with large-scale data, we employed online learning algorithm which can adaptively update itself so as to minimize the cumulative error rate as data and label are received in a sequential manner. To validate the proposed system, we created a massive dataset with professional annotations. The experimental result demonstrated the effectiveness of the proposed learning approach.

Moreover, the proposed system presented several favourable characteristics to facilitate the practical inspection: 1. An objective criterion can be established by using advanced machine learning algorithms for hammer response interpretation, which can effectively eliminate man-made errors; 2. Compared with human inspection, the proposed system exhibited higher efficiency in processing and it can be performed in real time; 3. The online learning framework can be adapted to deal with other types of impact-echo data, not limited to the hammer sounding application addressed in this study. In the next stage, we hope to upgrade the hardware part, such as employing wireless/wearable MIC; also, it can be anticipated that as we collect more hammering sounding data, the condition assessment precision will be further improved.

**Acknowledgments:** This study was partly supported by Cross-ministerial Strategic Innovation Promotion Program (SIP)/Infrastructure Maintenance, Renovation and Management, and supported by the New Energy and Industrial Technology Development Organization (NEDO), Japan. We are grateful to Shutoko Engineering Co., Ltd., and East Nippon Expressway Co., Ltd. for the collection of data, and Techny Co., Ltd. for developing the hammering device. The authors would like to thank the editor and anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

**Author Contributions:** Jiaying Ye provided the method and conducted the experiments. Takumi Kobayashi was in charge of experiment design and parameter optimization. Masaya Iwata helped in the collection and organization of data. Hiroshi Tsuda and Masahiro Murakawa supervised the theoretical statement of the problem. All authors revised the paper for intellectual content.

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

## References

1. Sansalone, M.J.; William, B.S. Impact-echo. *Nondestructive Evaluation of Concrete and Masonry*; Bullbrier Press: Jersey Shore, PA, USA, 1997.
2. Schubert, F.; Kohler, B. Ten lectures on impact-echo. *J. Nondestruct. Eval.* **2008**, *27*, 5–21.
3. Farrar, C.R.; Worden, K. *Structural Health Monitoring: A Machine Learning Perspective*; John Wiley Sons: New York, NY, USA, 2012.
4. Igual, J.; Salazar, A.; Safont, G.; Vergara, L. Semi-supervised Bayesian classification of materials with impact-echo signals. *Sensors* **2015**, *15*, 11528–11550.
5. Zhang, J.K.; Yan, W.; Cui, D.M. Concrete condition assessment using impact-echo method and extreme learning machines. *Sensors* **2016**, *16*, 447.
6. Ye, J.; Takumi, K.; Iwata, M.; Murakawa, M.; Higuchi, T.; Kubota, T.; Yui, T. Noise Reduction Methods for Hammering Impact Acoustic Inspection: An Experimental Comparison. In Proceedings of the International Conference on Structural Health Monitoring, Stanford, CA, USA, 1–3 September 2015.



7. Ye, J.; Takumi, K.; Iwata, M.; Murakawa, M. Noise robust hammering echo analysis for concrete structure assessment under mismatch conditions: A sparse coding approach. In Proceedings of the IEEE Sensors Applications Symposium (SAS), Glassboro, NJ, USA, 13–15 March 2017; pp. 1–6.
8. Gepperth, A.; Hammer, B. Incremental learning algorithms and applications. In Proceedings of the European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, 27–29 April 2016; pp. 357–368.
9. Hoi, S.C.H.; Wang, J.; Zhao, P. Libol: A library for online learning algorithms. *J. Mach. Learn. Res.* **2014**, *15*, 495–499.
10. Carino, N.J.; Sansalone, M.; Hsu, N.N. A point source-point receiver, pulse-echo technique for flaw detection in concrete. *J. Am. Concr. Inst.* **1986**, *2*, 199–208.
11. Sansalone, M. Impact-echo: The complete story. *Struc. J.* **1997**, *94*, 777–786.
12. Yeh, P.L.; Liu, P.L. Application of the wavelet transform and the enhanced Fourier spectrum in the impact echo test. *NDT Int.* **2008**, *41*, 382–394.
13. Hola, J.; Sadowski, L.; Sadowski, K. Nondestructive identification of delaminations in concrete floor toppings with acoustic methods. *Autom. Constr.* **2011**, *20*, 799–807.
14. Gibson, A.; Popovics, J.S. Lamb wave basis for impact-echo method analysis. *J. Eng. Mech.* **2005**, *131*, 438–443.
15. Zhu, J.; Popovics, J.S. Imaging concrete structures using air-coupled impact-echo. *J. Eng. Mech.* **2007**, *133*, 628–640.
16. Groschup, R.; Grosse, C.U. Development of an Efficient Air-Coupled Impact-Echo Scanner for Concrete Pavements. In Proceedings of the World Conference on Non-Destructive Testing, Munich, Germany, 13–17 June 2016.
17. Sadowski, Ł. Non-destructive identification of pull-off adhesion between concrete layers. *Autom. Constr.* **2015**, *57*, 146–155.
18. Li, B.; Ushiroda, K.; Yang, L.; Song, Q.; Xiao, J. Wall-climbing robot for non-destructive evaluation using impact-echo and metric learning SVM. *Int. J. Intell. Robot. Appl.* **2017**, *1*, 255–270.
19. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *7*, 551–585.
20. Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In Proceedings of the International Conference on Machine Learning (ICML), Pittsburgh, PA, USA, 25–29 June 2003; pp. 928–936.
21. Crammer, K.; Dekel, O.; Keshet, J.; Shalev-Shwartz, S.; Singer, Y. Online passive aggressive algorithms. *J. Mach. Learn. Res.* **2006**, *7*, 551–585.
22. Cesa-Bianchi, N.; Conconi, A.; Gentile, C. A second-order perceptron algorithm. *SIAM J. Comput.* **2005**, *34*, 640–668.
23. Crammer, K.; Dredze, M.; Pereira, F. Exact convex confidence-weighted learning. In Proceedings of the Neural Information Processing System Conference (NIPS), Vancouver, BC, Canada, 8–11 December 2008; pp. 345–352.
24. Crammer, K.; Kulesza, A.; Dredze, M. Adaptive regularization of weight vectors. In Proceedings of the Neural Information Processing System Conference (NIPS), Vancouver, BC, Canada, 7–10 December 2009; pp. 414–422.
25. Dredze, M.; Crammer, K.; Pereira, F. Confidence-weighted linear classification. In Proceedings of the International conference on machine learning (ICML), Edinburgh, UK, 26 June–1 July 2012; pp. 121–128.
26. Christopher, M.B. *Pattern Recognition and Machine Learning*; Springer Science Business Media: New York, NY, USA, 2006.

