In-Depth Steel Crack Analysis Using Photoacoustic Imaging (PAI) with Machine Learning-Based Image Processing Techniques and Evaluating PAI-Based Internal Steel Crack Feasibility

Arbab Akbar 1, Ja Yeon Lee 2, Jun Hyun Kim 2 and Myung Yung Jeong 3,*

1 Department of Cogno-Mechatronics Engineering, Pusan National University, Busan 46241, Republic of Korea; arbab@pusan.ac.kr
2 ESPn Medic, ESPn Medic Cooperation, Busan 46241, Republic of Korea; fish6508@pusan.ac.kr (J.Y.L.); oaksa03@pusan.ac.kr (J.H.K.)
3 Department of Optics and Mechatronics Engineering, Pusan National University, Busan 46241, Republic of Korea
* Correspondence: myjeong@pusan.ac.kr

Abstract: Steel plays an indispensable role in our daily lives, permeating various products ranging from essential commodities and recreational gears to information technology devices and general household items. The meticulous evaluation of steel defects holds paramount importance to ensure the secure and dependable operation of the end products. Photoacoustic imaging (PAI) emerges as a promising modality for structural inspection in the realm of health monitoring applications. This study incorporates PAI experimentation to generate an image dataset and employs machine learning techniques to estimate the length and width of surface cracks. Furthermore, the research delves into the feasibility assessment of employing PAI to investigate internal cracks within a steel sample through a numerical simulation-based study. The study’s findings underscore the efficacy of the PAI in achieving precise surface crack detection, with an acceptable root mean square error (RMSE) of 0.63 ± 0.03. The simulation results undergo statistical analysis techniques, including the analysis of variance (ANOVA) test, to discern disparities between pristine samples and those featuring internal cracks at different locations. The results discern statistically significant distinctions in the simulated acoustic responses for samples with internal cracks of varying sizes at identical/different locations (p < 0.001). These results validate the capability of the proposed technique to differentiate between internal crack sizes and positions, establishing it as a viable method for internal crack detection in steel.

Keywords: photoacoustic imaging; data augmentation; machine learning; regression models; statistical analysis; ANOVA test; internal steel cracks

1. Introduction

Steel finds its use in a variety of daily life applications like building structures. Structures (including buildings, bridges, vehicles, and many more) constitute to be one of our basic necessities after food and clothing and are a trademark of human civilization [1]. The safety (defined as the inhabitants, users, workers, and the neighborhoods are all safe) and reliability (defined as the structure will perform well without accidents for all of its designed lifetime) of all these structures are of core importance [2]. There are several techniques that have been employed to ensure the safety and reliability of steel starting from as simple as visual inspection techniques. For the task of defect detection, researchers have progressively devised new methodologies collectively called nondestructive testing (NDT) techniques [3]. All NDT techniques have their individual strengths and limitations.
depending upon the specific application requirements, result manipulation, time and cost factors, and other parameters.

The oldest known technique for NDT is the visual inspection technique starting with the naked eye, but with the advent of technology, the visual inspection technique is augmented with the power of strong microscopes and high-resolution camera lens setups [4]. Vibration-based methods have also been used in the industry for defect detection; however, they are limited by their low sensitivity and accuracy, especially for small subjects [5]. Electromagnetic (EM) techniques offer a vital contribution in the field of NDT, utilizing techniques like eddy current testing, electromagnetic induction, and the sweep frequency technique [6–8]. However, EM techniques have a strict requirement for being applicable only on electrically conducive test materials and require physical contact. NDT diagnostics based on guided wave propagation are popular in the inspection of large testing areas [9], but their application is limited by the requirement of sensor installation, involving cost and accessibility factors. With advances in computational resources and advanced signal processing techniques, machine learning and deep learning techniques provide robust, accurate, and fast tools to classify and predict the size and location parameters of the damages in both image and numeric data for applications in the field of NDT [10].

Though these and other NDT techniques may be applied in a standalone setting, a clever approach to counter each other’s limitations and benefit from mutual combined strengths is to combine these approaches together. An example of such a combination is the development of the photoacoustic imaging (PAI) technique. As the name suggests, it mutually benefits from laser, acoustic, and imaging techniques. PAI offers a good contrast, relying on the optical principles (the imaging part) and yielding a good penetration utilizing the ultrasonic principles (the acoustic part) [11]. PAI provides a noncontact testing scheme and overcomes the limitations of inaccessibility issues (due to hazardous or other environmental/geometrical factors), coupling agent (water/gel) performance, and other such factors. The photoacoustic (PA) effect relies on the generation of acoustic waves by transferring heat energy from a laser source to a target region, resulting in an increase in the kinetic energy, causing a volumetric expansion of the target region. Figure 1a shows the block diagram for the major steps involved in a complete loop of a PAI-based system. Figure 1b presents a schematic diagram of a photoacoustic microscopy system utilized in this study. The ambient conditions again cause the target region to contract as soon as the laser source is turned off; this rapid volume expansion and contraction generates pressure waves (ultrasonic waves) in the target material, which travel through the material in all directions and carry important embedded information about the internal structure of the material under test. After propagating through the material, these ultrasonic waves are sensed at designated points, and the data are stored in a personal computer. Analysis of the acquired data yields important information about the internal structural details of the material. A variety of analyses techniques may be applied for the desired task, but with quick advances and substantial outcomes in the field of image processing, relying on machine learning/deep learning algorithms proves to be a good choice [12].

PAI has been widely used in biomedical applications for the detection of disease and abnormalities in the skin and vital organs, and for diagnosing cancers and eye diseases [13,14]. However, in the field of NDT, there have been a few studies to detect the size and location of surface cracks. A study conducted by Grégoire et al. showed that in the presence of cracks, the PA spectra appear to have mixed frequencies that may help in crack detection [15]. Another study utilizing metallic plates showed nonlinear PA imaging of the surface-breaking cracks [16]. Jeon et al. detected surface defects in metal plates employing an OR-PAM PAI setup [17]. Earlier, our group performed a numerical simulation-based study evaluating PAI capabilities for steel surface crack detection and mathematical modeling of its acoustic response [18]. The current study utilizes an actual PAI experimentation setup for PAI dataset construction, and then uses deep/machine learning-based techniques to estimate the size of surface cracks in steel samples. Machine learning (ML) models have proven their efficacy and strength for image processing tasks in the last few years [19].
ML models have been utilized in a multitude of image processing applications, with high performance and efficiency results. The ML models' applications in image processing range from object detection and pattern recognition to segmentation including healthcare [20,21], optical microscopy [22], content-based image retrieval [23,24], and much more.

Figure 1. Photoacoustic principle and the photoacoustic setup utilized for experimentation in the study. (a) A block diagram that shows the major steps involved in the photoacoustic imaging data acquisition, starting from laser actuation to photoacoustic image reconstruction. (b) Schematic diagram of an actual photoacoustic imaging microscopy setup utilized for the actual experimentation in the study.

The detection of internal cracks is another crucial research avenue and requires an even more complex and rigorous methodology. The second section of this study demonstrates a numerical simulation-based feasibility study of the PAI technique for the detection of internal steel cracks. The acoustic responses to various internal defects with different sizes and locations are acquired. These responses are mutually compared statistically, and significance scores are calculated based on the analysis of variance (ANOVA) test. The ANOVA is a well-known predictor of analyzing group differences based on the signal mean and variances [25].

The contributions of this study are described as the construction of a dedicated PA image dataset for steel surface cracks using actual PAI experimentation. The objective is to apply machine learning models to the constructed dataset and precisely estimate the length and width of the surface cracks. The purpose is to propose a scheme that not only saves researchers time, but also generates reliable and accurate results. Also, a feasibility analysis of using PAI for internal crack detection of steel using a numerical simulation study and statistical analysis of the results is performed. The innovation of this study comes from the idea to utilize low-resolution images for defect detection instead of the high-resolution images generally used to generate comparable results. The proposed scheme enables us to utilize the powers of machine learning and deep learning models, while saving a huge amount of precious industry time in image acquisition as explained in the following sections.
2. Materials and Methods

This study consists of two sections: The first involves actual experimentation using a PAI setup on steel samples with surface cracks. A dataset is constructed, and machine learning models are applied to estimate the length and width of the surface cracks. The second section describes the numerical simulation process carried out in COMSOL Multiphysics software version 5.6, imitating the PA effect and investigating internal steel cracks by analyzing the acoustic response of the steel sample. The simulation study is repeated for multiple samples with and without internal cracks of different sizes at different locations.

2.1. Photoacoustic Imaging Experimentation and Machine Learning-Based Crack Detection

2.1.1. PAI Experimentation Setup

An original photoacoustic microscopy (PAM) system is utilized for actual experimentation as shown in Figure 2. The target material is a steel sample with surface cracks; the crack dimensions utilized for experimentation are listed in Table 1. The target sample is placed in the container at the desired coordinates using an XY motorized stage (Thorlabs, Newton, NJ, USA, DDS220/M). The container is then filled with oil for better acoustic transmission and measurements. The acoustic excitation in the target sample is achieved using a pulsed laser module (CoLID-I, Connet, Shanghai, China) bearing an approximate pulse energy of 0.42 µJ, repetition rate of 1 kHz, pulse width of 20 ns, and wavelength of 1560 nm. Laser illumination and acoustic data acquisition processes are synchronized using an external trigger. A custom-made ultrasound transducer (Precision Acoustics: center frequency = 20 MHz and diameter = 20 mm) detects the generated PA waves. The acoustic signal is then amplified (Minicircuit, Brooklyn, NY, USA, Zx60-3018G-S+) and digitized (Signatec, Poway, CA, USA, PX14400) at a sampling rate of 400 MHz.

![Figure 2. Actual picture of the PAM setup utilized for experimentation.](image-url)
Table 1. Surface crack dimensions utilized for actual experimentation using the PAM setup to generate raw acoustic response data.

<table>
<thead>
<tr>
<th>Crack #</th>
<th>Length (mm)</th>
<th>Width (µm)</th>
<th>Sample Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1000</td>
<td>Sample 1</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>350</td>
<td>Sample 2</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>500</td>
<td>Sample 3</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

2.1.2. PAI Dataset

The PAM setup described above is utilized to generate an A-line scan at a designated point on the sample surface. An A-line scan constitutes of 1024 amplitude values of the reflected acoustic wave. A B-line scan is generated by repeating the A-line measurements for 1024 points in the x-axis. Finally, a C-scan image is obtained by repeating the B-scan image acquisition process for 256 points in the y-axis, resulting in the final output with dimensions of $1024 \times 1024 \times 256$ for a single-crack image. Such single C-scan images take more than 16 min for acquisition, which will result in huge acquisition time if there is a large number of samples to be inspected.

This study proposes a machine learning approach to use a low-resolution image ($1024 \times 1024 \times 16$, which takes only 1 min of acquisition time) to produce reliable results comparable to those of high-resolution images. After acquiring the raw acoustic response data, PA images are reconstructed. Further, the reconstructed images are downsampling to generate low-resolution images, and the images are resized from $1024 \times 1024$ to a smaller size ($224 \times 224$) to boost the performance of the machine learning crack detection scheme. The number of images uniquely generated by the PAM setup and finally constructed after data augmentation techniques are listed in Table 2.

Table 2. Total unique images generated by the PAM setup and final images constructed after data augmentation techniques.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Crack Images</th>
</tr>
</thead>
<tbody>
<tr>
<td># of cracks</td>
<td>15</td>
</tr>
<tr>
<td>PAM setup’s unique laser excitation angles per crack</td>
<td>4</td>
</tr>
<tr>
<td>Total original images</td>
<td>$15 \times 4 = 60$</td>
</tr>
<tr>
<td>Total Images after applying Data augmentation techniques</td>
<td>4050</td>
</tr>
</tbody>
</table>

Data augmentation (image rotation, reflection, and scaling) techniques presented in Table 3 are applied on the low-resolution images to generate an ample amount of data for the machine learning models. Figure 3 presents the actual steel samples with surface cracks used for PAI experimentation and a few examples PA images before and after data augmentation. The overall scheme starting from performing actual experiments on the steel samples, image preprocessing, and dataset construction is presented in Figure 4.
Figure 3. Steel samples used for actual PAI experimentation are presented in the first column; the second column provides a few example PA images reconstructed from the acoustic responses received after experimentation. Finally, the third column shows a few images after application of data augmentation techniques.

Figure 4. Overall scheme showing the steps involved in the construction of the final PAI dataset starting from PAI experimentation, image preprocessing, image resizing, and data augmentation.
2.1.3. Machine Learning-Based Crack Length and Width Prediction Framework

For the goal of predicting the length and width of the cracks, machine learning capabilities of CNN models are combined with statistical techniques and regression models. The concept is to extract deep features from the input PA images through three deep CNN models (GoogleNet, ResNet50, and VGG16) using transfer learning. Each of these CNN models exhibits unique architectural characteristics to extract discriminative features from input images [26]. VGG16 is characterized by a simpler architecture featuring smaller filter sizes, albeit with a substantial number of trainable parameters [27]. ResNet50 [28] introduces the concept of residual learning to address the challenge of vanishing gradients. GoogleNet, on the other hand, prioritizes the simultaneous capture of multiscale features across successive layers [29]. The CNN models are pretrained on the famous dataset ImageNet [30] containing millions of images and several diverse classes. Since the feature space from these models is quite rich and diverse from each other, a dimensionality reduction statistical technique called principal component analysis (PCA) is utilized [31]. The extracted features from the penultimate layers of each of these models are fed into the PCA block for dimensionality reduction. In this way, the complexity of the proposed scheme is significantly reduced while keeping the maximum information intact [32,33]. The working principle of PCA can be divided into two major steps: (1) the covariance matrix computation for the dataset and (2) identification of the principal components by singular value decomposition of the covariance matrix [34]. The eigenvalue represents the maximum variance in the dataset, and the eigenvector gives the direction of the maximum variance. The order of the variances gives the most vital principal components. Using PCA on the deep features extracted using the 3 deep CNN models, the most significant 128 PCA features are retained. This 128-dimensional vector is then input to the three different regression models (linear, SVM, and random forest regression models) used. The complete schematic is presented in Figure 5.

Table 3. Data augmentation techniques and their corresponding parameter values for generating sufficient volumetric data for a deep learning model.

<table>
<thead>
<tr>
<th>Augmentation Technique</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation</td>
<td>[0, 360] degrees</td>
</tr>
<tr>
<td>Reflection</td>
<td>X and Y [-1, 1]</td>
</tr>
<tr>
<td>Scaling</td>
<td>[0.5, 1]</td>
</tr>
</tbody>
</table>

![Figure 5. Schematic diagram for combinatorial application of deep CNN models for feature extraction, PCA for dimensionality reduction, and machine learning regression models to predict crack length and width.]
2.1.4. Regression Models

Three different regression models are used for crack size estimation. The following subsections give a brief introduction of each of these models:

Linear Regression

In machine learning, linear regression is a statistical method to model the relationship between a dependent variable and one or more independent variables [35]. The model’s objective is to evaluate the best-fitting linear relationship predicting the value of the dependent variable based on the independent variables [36]. Mathematically, linear regression may be represented by Equation (1),

\[ Y = XM + \beta \] (1)

where \( Y \) is the output vector, \( X \) is a matrix of the observed value of independent variables, \( M \) is the vector of coefficients to be optimized during the training to minimize the error between actual and the predicted values of \( Y \), and \( \beta \) represents the error term showing the difference between actual and predicted values.

Support Vector Regression (SVR)

Support vector regression is a powerful technique for regression tasks adapted from the powerful machine learning classifier, the support vector machine (SVM). SVR is a good choice in situations of nonlinear relationships, and where there is an acceptable error margin between the actual and predicted values [37,38]. Mathematically, a simple SVR model may be represented by Equation (2),

\[ y_i = f(X_i) + \epsilon \] (2)

where \( f(X_i) \) represents the predicted value for the \( i \)th input data point \( X_i \), and \( \epsilon \) is the error between the predicted value and the actual value \( y_i \). The objective function of SVR is to minimize feature coefficients and the error margin defined by \( \epsilon \), as defined by Equation (3).

\[
\min_{w,b}\sum_{i=1}^{N} (\xi_i + \xi_i^*) + \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)
\] (3)

The objective function defined above is subjected to the constraints given by Equations (4)–(6),

\[ y_i - (w.X_i + b) \leq \epsilon + \xi_i \] (4)

\[ (w.X_i + b) - y_i \leq \epsilon + \xi_i^* \] (5)

\[ \xi_i, \xi_i^* \geq 0 \] (6)

where \( w \) is the weight vector, \( b \) is the bias term, \( \xi_i, \xi_i^* \) are slack variables that allow for some points to fall outside the margin, \( C \) is a regularization parameter that controls the trade-off between minimizing the error and minimizing the magnitude of the weights, and \( \epsilon \) is a parameter that defines the width of the margin.

Random Forest Regression

Random forest regression (RFR) is a decision tree-based ensemble learning technique that is based on random forest algorithm [39,40]. The mechanism of random forest regression is to build and train multiple trees on different subsets of data predicting the output values (length and width of crack) individually. These individual estimates from all the trees are then averaged to get the final value of the crack size. This scheme makes the model robust and avoids overfitting.
2.1.5. Implementation Details

The tasks of image preprocessing and augmentation were implemented using Python libraries OpenCV (cv2) and Scikit Image (skimage). The CNN models (VGG16, ResNet50, and GoogleNet) with pretrained ImageNet parameters were implemented in Pytorch, and these models were fed with the batch of input images to extract the features. These features’ embeddings were concatenated over the whole training set, and the resultant feature matrix was used for PCA-based feature selection. The top 128 features (k = number of components) were selected on the basis of an optimal explained variance against a minimum number of components. These selective training features were used to train three regression models. The whole dataset was split using a ratio of 8:2 for generating the training and testing datasets, respectively. A total of 10% of the data from the training dataset were used for cross-validation. Python library SciKit Learn (sklearn) was used to perform regression and hyperparameter tuning. In the case of linear regression, generally, hyperparameter tuning is not needed because its performance primarily depends on how well input data follow a linear distribution [41]. To improve the performance of the linear regression model, data were normalized, and regularization was added into the objective function. A grid search method was used to find the optimal hyperparameters for RFR and SVR. RFR only requires the number of trees, maximum depth, and number of evidential features (m) to generate a prediction model. The model’s accuracy may be improved by reducing correlation among trees by reducing the parameter m [42]. To find an optimal combination of these parameters, the range of trees was set between 100 and 300 at intervals of 50 and the search depth between 10 and 30 at an interval of 10. The value of m = 11 was selected using the square root of the input feature size. The best results were obtained for 300 trees, 20 search depth, and m = 11 using GridSearch. After a certain number of trees and depth, improvement in the model’s accuracy is negligible but computational overhead is significant [43]. Root mean square error (RMSE) is used for reporting the performance results of the proposed machine learning models. RMSE is a well-known and reliable performance metric when assessing the accuracy of a predictive model. RMSE measures the average magnitude of the errors between predicted values and actual values by taking the square root of the average of the squared differences between predicted and actual values. Mathematically, it can be expressed in Equation (7) as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]  

(7)

where \( n \) represents the number of observations, \( y_i \) represents the actual values, \( \hat{y}_i \) gives the predicted values, and the resulting \( RMSE \) gives the root mean squared error for the model.

2.2. PAI-Based Internal Crack Detection Feasibility Using Numerical Simulation

Finite element method simulations are performed to study the effects of internal damage on the propagating acoustic wave and investigate the information it carries about the internal material structure. COMSOL MultiPhysics\textsuperscript{®} (version 5.6) software package is utilized to model the geometry of a solid steel block. The physical properties of the steel material are set as given in Table 4. Internal damage in the form of a void of different sizes is introduced inside the steel block at different locations. The cracks are introduced at one of the external edges only to be able to replicate the geometry for an actual experimentation geometry in a future study. Figure 6 shows the model geometry used for simulations with internal damages. Figure 7 presents the resulting mesh grid showing triangular mesh components while also giving an idea of their approximate size. It is important to note here that the mesh is fine near the defect edges for better analysis but is kept relatively coarse at other areas to save computational resources. As described in the literature [44], the element size should be calculated keeping in consideration the minimum wavelength of the propagating acoustic wave, as expressed in the mathematical relation given in
Equation (8). Also, the simulation time step and the maximum frequency are associated with the mathematical relation given in Equation (9).

\[ \Delta I = \frac{\lambda_{\text{min}}}{20} \]  

(8)

Table 4. Properties of steel material used in the numerical simulation setting.

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>7850 kg m(^{-3})</td>
</tr>
<tr>
<td>Thermal conductivity</td>
<td>44.5 W/(m·K)</td>
</tr>
<tr>
<td>Coefficient of thermal expansion</td>
<td>12.3 (\times) 10(^{-6}) K(^{-1})</td>
</tr>
<tr>
<td>Bulk velocity</td>
<td>6200 m s(^{-1})</td>
</tr>
<tr>
<td>Shear velocity</td>
<td>3200 m s(^{-1})</td>
</tr>
<tr>
<td>Young’s modulus</td>
<td>205 GPa</td>
</tr>
<tr>
<td>Poisson’s ratio</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Figure 6. A representation of all the geometries used for simulation. Origin of the coordinate system is the left bottom corner of the steel block. (a) Steel block (dimensions 150 × 50 mm\(^2\)) without any internal crack. (b) Steel block with a crack (crack width 1 mm height 10 mm) at an offset of (120,0). (c) Steel block with a crack (crack width 10 mm height 1 mm) at an offset of (150,40). (d) Steel block with a crack (crack width 1 mm height 10 mm) at an offset of (30,0). (e) Steel block with a crack (crack width 10 mm height 5 mm) at an offset of (150,40). (f) Steel block with a crack (crack width 10 mm height 5 mm) at an offset of (150,20).
The acoustic wave is measured at the designated detection points to mimic the acoustic responses from the defected sample and the acoustic response from the sample coordinate value of (0, 5), visible as a red dot in the figure. The acoustic wave is measured at the designated detection points to mimic the pulse and echo scenario. Figure 8A shows the face of the geometry to which the mechanical actuation is applied using a blue highlighting color. In Figure 8B, the acoustic signal detection point is highlighted as a red-colored dot to pinpoint the sensor location for the measurement of the reflected acoustic wave. All the detection points receive the amplitude intensity of the reflected acoustic wave and store it in data variables. The farthest face of the steel block is configured as a low-reflecting boundary to present it as an infinite geometry to the simulation software and deal with reflections only from inserted internal damages.

Here, $\Delta t$ is the simulation time step and $f_{\text{max}}$ is the highest frequency of interest. A mechanical force of 10 kN is applied as a load at the face of the block to generate the acoustic waves. The acoustic wave is measured at the designated detection points to mimic the pulse and echo scenario. Figure 8A shows the face of the geometry to which the mechanical actuation is applied using a blue highlighting color. In Figure 8B, the acoustic signal detection point is highlighted as a red-colored dot to pinpoint the sensor location for the measurement of the reflected acoustic wave. All the detection points receive the amplitude intensity of the reflected acoustic wave and store it in data variables. The farthest face of the steel block is configured as a low-reflecting boundary to present it as an infinite geometry to the simulation software and deal with reflections only from inserted internal damages.

Figure 7. Geometry details showing the dimension of the steel cylinder, the internal damage, and the approximate mesh size for the numerical analysis.

Figure 8. Actuation point and sensor location set in the simulation. (A) Left face of the geometry is configured as the uniform actuation application area visible as a blue line, (B) Detection point set at a coordinate value of (0, 5), visible as a red dot in the figure.

Statistical Analysis

In order to evaluate our hypothesis, we needed to analyze the differences between the acoustic responses from the defected sample and the acoustic response from the sample in pristine condition. Statistical technique was employed for the purpose, and analysis
of variance (ANOVA) tests were performed using software MATLAB 2017 version 9.3 (MathWorks). A confidence interval of 95% was defined for the ANOVA test input conditions. An ANOVA test determines whether data from several conditions of a parameter have a common mean. In other words, the ANOVA test enables us to find out whether different groups of an independent variable have different effects on the response variable. Mathematically, ANOVA can be expressed as a linear model with $j$ number of different input groups, and there are $i$ number of responses per group, given by Equation (10).

$$y_{ij} = \alpha_j + \varepsilon_{ij}$$ \hspace{1cm} (10)

Here, $y_{ij}$ is the response for the $i$th observation of the $j$th group, $\alpha_j$ is the mean for group number $j$, and $\varepsilon_{ij}$ is assumed to be a random error with zero mean and a constant variance. ANOVA partitions the total variation in the data into following two components to test for the difference in the group means:

- Difference of group means from the overall mean, i.e., $\bar{y}_j - \bar{Y}$ (variation between groups), where $\bar{y}_j$ is the sample mean of group $j$, and $\bar{Y}$ is the overall sample mean;
- Difference of observations in each group from their group mean estimates, $y_{ij} - \bar{y}_j$ (variation within a group).

In other words, ANOVA partitions the total sum of squares (SST) into sum of squares due to the between-groups effect (SSR) and sum of squared errors (SSE), as formulated in Equation (11).

$$\sum_{i} \sum_{j} (y_{ij} - \bar{Y})^2 = \sum_{j} n_j (\bar{y}_j - \bar{Y})^2 + \sum_{i} \sum_{j} (y_{ij} - \bar{y}_j)^2$$ \hspace{1cm} (11)

Here, $n_j$ is the sample size for the $j$th group, where $j = 1, 2, \ldots, k$. An ANOVA test is used to compare the variation between groups to the variation within groups. If the ratio of between-group variation to within-group variation is significantly high resulting in a low $p$-value, it statistically proves that the group means are significantly different from each other. This can be measured using a test statistic that has an $F$-distribution with $(k - 1, N - k)$ degrees of freedom as presented in Equation (12).

$$F = \frac{\sum_{i} \sum_{j} (y_{ij} - \bar{y}_j)^2 / k - 1}{\sum_{j} n_j (\bar{y}_j - \bar{Y})^2 / N - k}$$ \hspace{1cm} (12)

While comparing the acoustic response of the actuation between the damaged samples and the samples in pristine condition, two independent sample $t$-tests were performed, whereas paired $t$-tests were used for the comparison of the acoustic response of the damaged samples with different detector positions. Data are expressed as the mean $\pm$ SD. A difference was considered to be statistically significant if the resulting $p$-value was less than 0.05. To verify the results based on detector location, independent sample and paired $t$-tests were performed. Table 5 presents the description of all the tests performed, along with the test purpose and the geometries utilized in each test.

To statistically determine whether the presence of a crack and its location introduces differences in responses, we compared the acoustic responses from a sample without any internal crack (refer to Figure 6a), sample having a crack at an offset of $(120,0)$ from the origin (refer to Figure 6b), and sample having an internal crack at an offset of $(150,40)$ from the origin (refer to Figure 6c) using a one-way ANOVA test. Furthermore, to investigate the differences based on crack location, we compared the acoustic responses of samples with internal cracks at offsets of $(30,0)$ and $(120,0)$ from the origin (refer to Figure 6d and Figure 6b, respectively), and samples with internal cracks at offsets of $(150,20)$ and $(150,40)$ from the origin (refer to Figure 6f and Figure 6e, respectively) were mutually compared. For addressing the differences based on size of the internal crack, we compared the simulation
datasets of samples with different sizes of cracks at the same location, i.e., an offset of (150,40) from the origin (refer to Figure 6c and Figure 6e, respectively).

Table 5. The purpose of every comparison carried out is listed along with the geometries involved in the comparison.

<table>
<thead>
<tr>
<th>Comp #</th>
<th>Geometries Involved (Shown in Figure 6)</th>
<th>Purpose (to Show if Statistically Significant Differences Are Present in Acoustic Responses of):</th>
</tr>
</thead>
</table>
| 1      | 6a, 6b, 6c                             | • a sample with no crack  
                  |                                                      | • a sample with crack in horizontal axis  
                  |                                                      | • a sample with a crack in vertical axis |
| 2      | 6b, 6d                                 | 2 samples having internal cracks at two different locations along horizontal axis             |
| 3      | 6e, 6f                                 | 2 samples having internal cracks at two different locations along vertical axis                |
| 4      | 6c, 6e                                 | 2 samples having internal cracks of different sizes at same location                         |

3. Results and Discussion

The root mean square error (RMSE) from the regression models was utilized to compare the performance of the CNN and regression model combinations. Table 6 presents the RMSE (mean ± standard deviation) of predicted crack lengths (unit: mm) by the combinations of three CNNs and three regression models utilized, whereas in Table 7, the RMSE (mean ± standard deviation) of the predicted crack width (unit: µm) by these combinations is reported.

Table 6. RMSE means and standard deviations in the predicted crack lengths (unit: mm) by the CNN model and regression model combinations utilized.

<table>
<thead>
<tr>
<th></th>
<th>GoogleNet</th>
<th>ResNet50</th>
<th>VGG16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>0.91 ± 0.07</td>
<td>0.97 ± 0.06</td>
<td>0.93 ± 0.05</td>
</tr>
<tr>
<td>Support vector regression</td>
<td>0.81 ± 0.06</td>
<td>0.75 ± 0.05</td>
<td>0.69 ± 0.05</td>
</tr>
<tr>
<td>Random forest regression</td>
<td>0.79 ± 0.05</td>
<td>0.71 ± 0.04</td>
<td>0.63 ± 0.03</td>
</tr>
</tbody>
</table>

Table 7. RMSE means and standard deviations in the predicted crack widths (unit: µm) by the CNN model and regression model combinations utilized.

<table>
<thead>
<tr>
<th></th>
<th>GoogleNet</th>
<th>ResNet50</th>
<th>VGG16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>9.2 ± 1.2</td>
<td>10.1 ± 0.95</td>
<td>9.9 ± 0.81</td>
</tr>
<tr>
<td>Support vector regression</td>
<td>7.9 ± 0.75</td>
<td>7.3 ± 0.67</td>
<td>7.1 ± 0.53</td>
</tr>
<tr>
<td>Random forest regression</td>
<td>7.4 ± 0.63</td>
<td>6.9 ± 0.51</td>
<td>6.2 ± 0.48</td>
</tr>
</tbody>
</table>

The results show that among the CNN models utilized, the features extracted by VGG16 perform the best in the identification and quantification of the crack, because it provides the least RMSE values for the predicted crack length (RMSE: 0.63 ± 0.03) and width (6.2 ± 0.48). However, on occasion, GoogleNet performs well especially when combined with the linear regression model (RMSE: 0.91 ± 0.07 for crack length estimation). Additionally, random forest regression is the best model to estimate the crack length and width with a lesser RMSE (mean ± standard deviation: 0.63 ± 0.03) as compared with the linear regression and support vector regression models.

Figure 9 shows the simulation results for the acoustic responses of multiple geometries at specific time instances. Figure 9 shows the details related to the acoustic wave propagation and its interaction with different crack fronts that come in its path. The division of the principal wave in the transmitted and reflected parts can also be seen in Figure 9.
When the incident acoustic wave comes in contact with a defect, a part of it is reflected back depending upon the interaction area. The reflections here are found to be directly proportional to the surface area/interaction area of the crack and the propagating wave.

![Simulation snapshot for acoustic pressure](image)

**Figure 9.** Simulation snapshot for the propagation of an acoustic wave within samples with internal cracks and without any defect, showing the acoustic wave interaction with internal defects at different locations for all the simulated geometries.

Figure 10 presents the snapshot of the pressure profile of the propagating acoustic wave on the specific time instances for various geometries. An acoustic wave is a longitudinal wave that travels in the form of compressions (high-pressure regions) and rarefactions (low-pressure regions). The red color shows the pressure peaks (compressions), whereas the blue color shows the troughs (rarefactions) of the acoustic wave during propagation.

![Simulation snapshot for acoustic pressure](image)

**Figure 10.** Simulation snapshot for acoustic pressure of the propagating acoustic wave for all the simulated geometries, showing the reflection and transmission phenomena at the interaction points with internal defects. The red color shows the pressure peaks (compressions), whereas the blue color shows the troughs (rarefactions) of the acoustic wave during propagation.

Figure 11 displays the results of a one-way test performed between the acoustic responses of the three samples shown in the Figure 6a–c. The bars show the average acoustic response intervals for the three geometries compared. The blue bar shows the
comparison interval for the mean of the geometry without any crack, which does not overlap with the comparison intervals for the second, i.e., a geometry with an internal crack at the horizontal axis, and a third group which represents the geometry with an internal crack at the vertical axis. Similarly, Figure 11 also shows that the comparison intervals for the other groups’ means do not overlap with each other, showing that all three groups have means which are significantly different from each other. As the group means from all three geometries are statistically proven to be significantly different, it means that they can be mutually distinguished; hence, the defect and its type may be detected using the simulated PAI technique. Figure 12 shows the box plot for the same three samples, showing the notches for median comparisons. The red center line represents the median of the group data, and the blue box shows the data between the 25th and 75th percentiles. The black end lines are the maximum and minimum whiskers, and the red crosses are the outliers. Figure 13 lists the ANOVA table results showing a small $p$-value of less than 0.0001, showing a statistically significant difference between the comparison datasets.

![Figure 11](image1.png)

**Figure 11.** One-way ANOVA test to compare the means of acoustic responses from a sample with no crack, a crack on the horizontal axis, and crack on the vertical axis to prove statistically that both crack presence and location are distinguishable using a photoacoustic imaging technique. The blue color bar shows that Group 1 is selected.

![Figure 12](image2.png)

**Figure 12.** Box plot showing medians for the one-way ANOVA test to compare the means of acoustic responses from a sample with no crack, a crack on the horizontal axis, and crack on the vertical axis to prove statistically that both crack presence and location are distinguishable using a photoacoustic imaging technique.
Table 8. ANOVA table with groupwise comparison details.

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
<th>Lower Limit</th>
<th>A-B</th>
<th>Upper Limit</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-6.90 × 10^{-5}</td>
<td>-5.55 × 10^{-5}</td>
<td>-4.20 × 10^{-5}</td>
<td>9.56 × 10^{-10}</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>-3.78 × 10^{-5}</td>
<td>-2.43 × 10^{-5}</td>
<td>-1.09 × 10^{-5}</td>
<td>6.84 × 10^{-5}</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1.77 × 10^{-5}</td>
<td>3.12 × 10^{-5}</td>
<td>4.47 × 10^{-5}</td>
<td>1.73 × 10^{-7}</td>
</tr>
</tbody>
</table>

To demonstrate the effect of the location of a crack of the same size, the geometries shown in Figure 6b-d are also mutually evaluated using a one-way ANOVA test. The results of this comparison are shown in Figures 14 and 15. The comparison intervals for both of these groups do not overlap with each other, and the ANOVA table in Figure 15 with a small p-value (p < 0.0001) proves that the means of these two groups have statistically significant differences.

Figure 14. Comparison intervals for the means of the two groups with a crack of the same size but placed at different locations horizontally.
Figure 15. ANOVA table results in a small p-value (p-value < 0.0001), showing statistically significant differences between the groups being compared with cracks of the same size but at different locations horizontally.

To compare two samples with cracks of the same size but at different locations vertically, the geometries shown in Figure 6e,f are compared using a one-way ANOVA test. Figure 16 presents the comparison intervals for both of these groups, which do not overlap with each other. Also, the ANOVA table in Figure 17 gives a small p-value (p < 0.0001), which is statistical proof that the means of these two groups have significant differences.

Figure 16. ANOVA test showing two mutually nonoverlapping intervals for the groups being compared with cracks of the same size but placed at different locations vertically.

Figure 17. ANOVA table giving a small p-value (p-value <0.001), showing significant differences among the compared groups with cracks of the same size but at different locations vertically.

Finally, the geometries shown in Figure 6c,e are compared to evaluate the situation when the internal cracks in two samples are at the same location but are different in size. Figures 18 and 19 give the comparison of intervals and the ANOVA table results for this comparison. Again, the mutually exclusive intervals of both groups and the small p-value from the ANOVA table provide enough proof that these two groups have differences which are statistically significant.
A comparative analysis of the various combinations of CNN models for feature extraction.

The power of machine learning enables us to devise effective schemes to generate accurate results. The authors note that PAI is feasible for the detection of internal steel cracks. In the future, the authors plan to conduct actual PAI experimentation on steel samples with internal defects for the identification, classification, size, and location estimation of these internal defects.

**Author Contributions:** Conceptualization, M.Y.J.; data curation, J.Y.L.; formal analysis, A.A. and J.H.K.; funding acquisition, M.Y.J.; investigation, M.Y.J.; methodology, A.A.; project administration, J.Y.L., J.H.K. and M.Y.J.; resources, J.Y.L., J.H.K. and M.Y.J.; software, A.A.; supervision, M.Y.J.; validation, A.A. and J.H.K.; writing—original draft, A.A.; writing—review and editing, A.A. and M.Y.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) [grant number: 2022R1A2B5B01002377].
Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy issues.

Conflicts of Interest: Authors J.Y.L. and J.H.K. were employed by ESPN Medic, ESPN Medic Cooperation, Busan, Republic of Korea. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflict of interest.

References

2. Schneider, J. Introduction to Safety and Reliability of Structures; Iabse: Zurich, Switzerland, 2006; Volume 5.
4. Alzuhiri, M.; Li, Z.; Rao, A.; Li, J.; Fairchild, P.; Tan, X.; Deng, Y. IMU-assisted robotic structured light sensing with featureless registration under uncertainties for pipeline inspection. NDT E Int. 2023, 139, 102936. [CrossRef]
6. Gr


